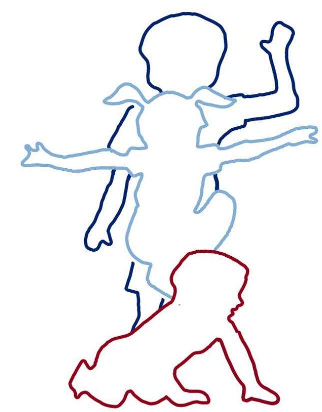
Acquiring recursive structures through distributional learning

Daoxin Li & Kathryn Schuler
The Child Language Lab, University of Pennsylvania
CUNY 2021 Annual Conference, March 4, 2021



Recursion

Recursion: The infinite self-embedding of a particular type of linguistic element or grammatical structure.

The ability for recursion is considered to be the core of the language faculty and universally available (e.g. Hauser, Chomsky, & Fitch, 2002; Nevins, Pesetsky, & Rodrigues, 2009; Partee & Rooth, 1983; Pinker, 1994).

Recursive structures: A learning problem

Languages differ regarding the depth, structure, and syntactic domains of recursive structures (Pérez-Leroux et al., 2018).

(1) English: the man*'s* neighbor*'s* book
(2) German: *das Mann*s* Nachbar*s* Buch (Weiss, 2008)

Recursive structures: A learning problem

Even within a single language, some structures allow infinite self-embedding while others are more restricted.

(3) a. the man'*s* neighbor'*s* computer

b. ?the computer *of* the neighbor

c. ?*the computer *of* the neighbor *of* the man

(e.g. Biber, Geoffrey, Leech, Conrad, & Finegan, 1999; Levi, 1978)

How do children learn which structures allow free recursive embedding and which structures are restricted?

How to learn freely recursive structures

Given the cross- and within-linguistic differences, the recursive structures have to be learned from language specific experience.

What kind of experience is useful and how do learners make use of it?

Observe multiple embedding in the input?

Can the attestation of multiple-level embedding in the input lead to the acquisition of recursive structures? (e.g. Roeper, 2011)

But

- Children acquire recursive structures even though evidence for deep embedding is rarely attested in young children's input (e.g. Giblin et al., 2019; Li et al., 2020).

- A logical problem: no N-level embedding entails N+1 level embedding.

The distributional learning proposal (Grohe et al., 2020; Li et al., 2020)

Recursion as structural interchangeability: X_1's X_2 is recursive if X_1 and X_2 positions are interchangeable. e.g. <u>*cat*</u>*'s tail, kid's* <u>*cat*</u>

Learning interchangeability as a productive generalization: Generalize if a sufficiently large proportion of words attested in the X_1 position in the input are also attested in the X_2 position in the input.

Paucity of deep embedding from input and the logical problem will no longer be problematic.

The distributional learning proposal (Grohe et al., 2020; Li et al., 2020)

cat’s …	…’s cat
mom’s …	…’s mom
car’s …	…’s car
toy’s …	…’s toy
room’s …	…’s room
owner’s …	…’s owner
ball’s …	…’s ball
boy’s…	
game’s …	

The distributional learning proposal (Grohe et al., 2020; Li et al., 2020)

Corpus studies: reliable distributional information in the input

- *det-adj1-adj2-noun* in English and German: sufficient evidence that adjectives can appear in both *adj1* and *adj2* positions - prenominal adjectives can be used recursively (Grohe et al., 2020).

Language	English	German
N in A_1 or A_2	49	38
N in A_1 & A_2	46	31
TSP threshold	36	28
Productive?	Yes	Yes

The distributional learning proposal (Grohe et al., 2020; Li et al., 2020)

Corpus studies: reliable distributional information in the input

- Possessive structures in English, German, and Mandarin: sufficient evidence that nouns can appear in both possessor and possessee positions of the recursive structures (Li et al., 2020).

Language	English		German		Mandarin	
Structure	X_1's X_2	X_2 *of* X_1	X_1's X_2	X_2 *von* X_1	X_1 *de* X_2	X_1 X_2
Recursivity	Yes	No	No	Yes	Yes	No
N in X_2	59	43	34	40	41	27
N in X_1 & X_2	46	28	5	34	35	15
TSP threshold	45	32	24	29	30	19
Productive?	Yes	No	No	Yes	Yes	No

Do learners indeed utilize the distributional information as predicted by the distributional learning proposal?

Experiment

Participants
- 50 native English-speaking adults on Prolific

Input
- X_1-*ka*-X_2 artificial language strings, with no referential world

Conditions

Condition	Words attested in X_1	Words attested in X_2	Prediction: recursive?
productive	12	10	yes
unproductive	12	6	no

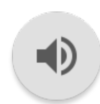
Experiment - exposure

10 out of 12 and 6 out of 12 are consistent with several metrics of productivity:

- Majority of Forms (e.g. Bybee, 1995): productivity threshold = 7
- The Tolerance/Sufficiency Principle (Yang, 2016): productivity threshold = 8
- Word-Form Rule (Aronoff, 1976; Baayen & Lieber, 1991): productivity index = 0.83, 0.50

Zipfian distribution (Zipf, 1949), 44 string exposure corpus, 2 repetition.

Experiment - exposure



Experiment - test

Is this string from the language you have just heard?

Experiment - test

Sample test strings in Unproductive condition (*sane*, *tesa* and *tana* are never attested in X_2 position during exposure)
Word attested in the position; word unattested in the position

Type	One-level	Two-level
attested	*waso-ka-mito*	*sane-ka-kewa-ka-nogi*
unattested	*nogi-ka-sane*	*waso-ka-tesa-ka-tana*
ungrammatical	*ka-bila-kosi*	*ka-waso-kosi-sito-ka*

Experiment - prediction

Participants from the Productive condition are predicted to rate unattested strings higher than participants from the Unproductive condition at both one and two embedding levels.

Condition	Words attested in X1	Words attested in X2	Prediction: recursive?
productive	12	10	yes
unproductive	12	6	no

Experiment - analysis

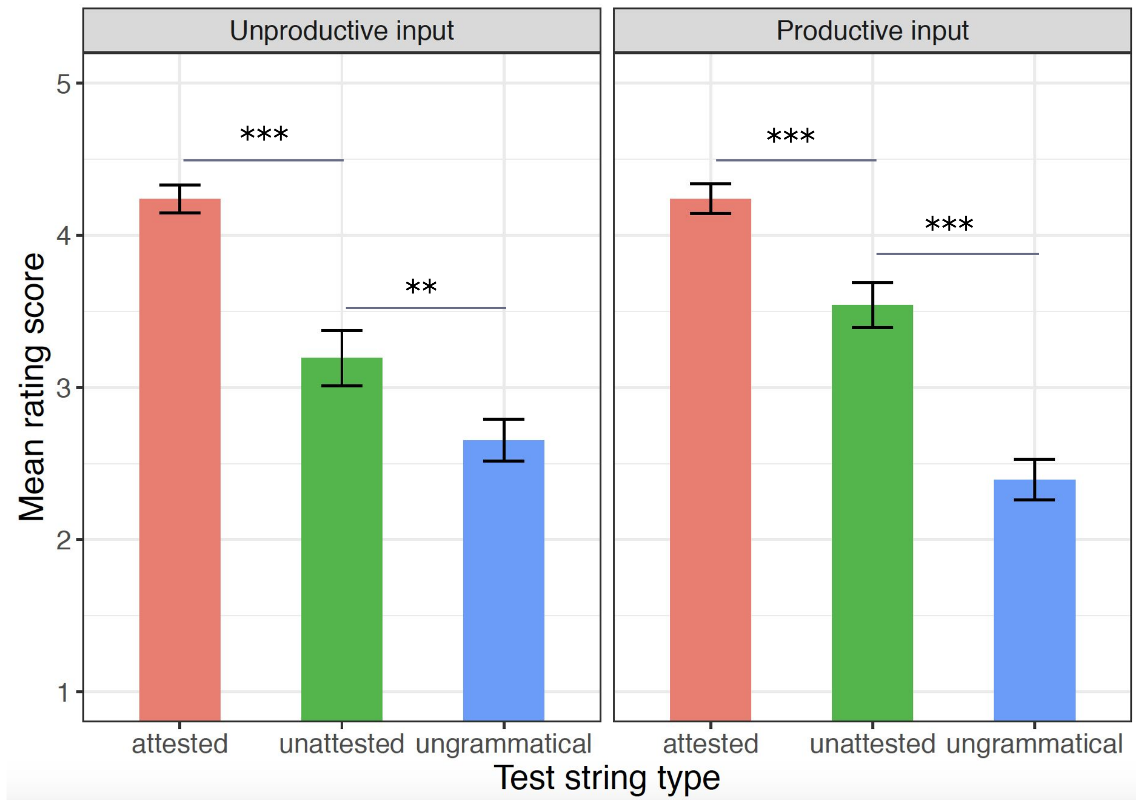
Ordinal regression:

- DV: rating score as an ordered factor from 1 to 5
- Fixed effects: test string Type (attested, unattested, or ungrammatical) and Condition (Unproductive, Productive)
- Random effects: by-participant random intercepts and random slopes for Type

Experiment - results

One-level strings

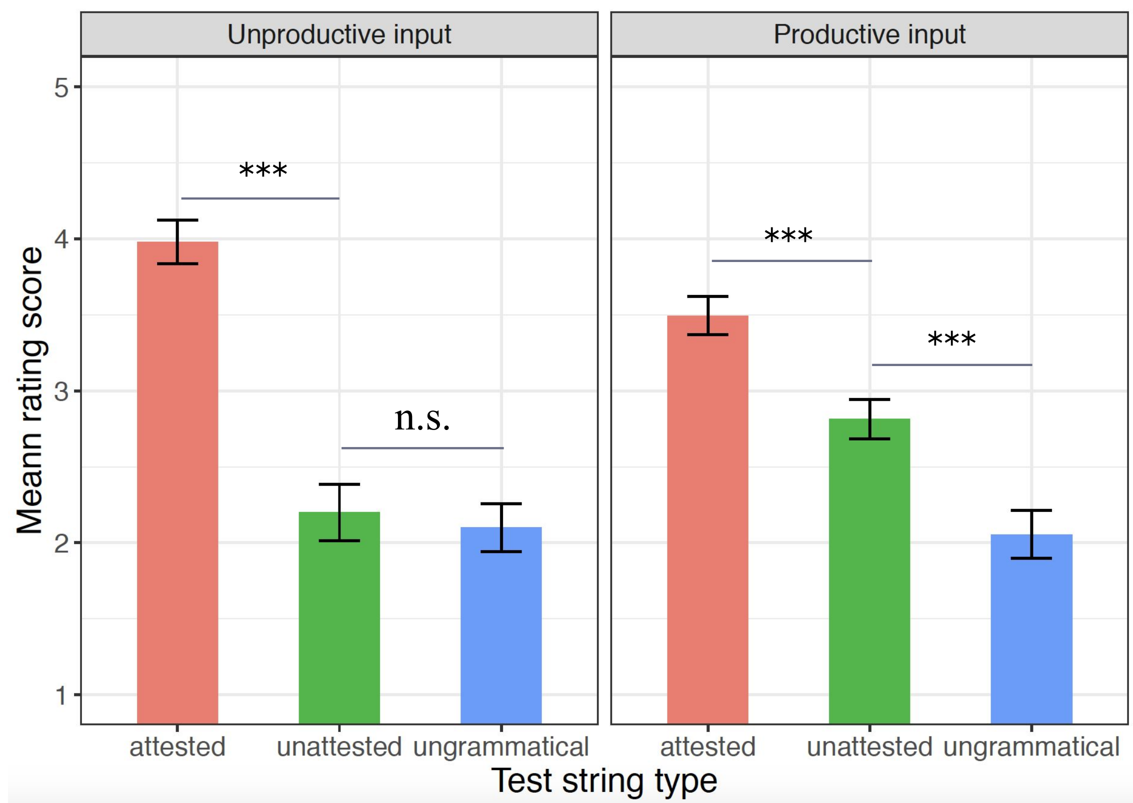
- No main effect of Condition (p=0.90)
- Significant main effect of Type (p<0.001)
- Significant interaction between Type and Condition (p=0.01)



Experiment - results

Two-level strings

- No main effect of Condition (p=0.81)
- Significant main effect of Type (p<0.001)
- Significant interaction between Type and Condition (p<0.001)



Conclusion

- Participants in our study learned the recursivity of a structure distributionally from language-specific level-one experience: a structure is recursive if the two positions are productively interchangeable.
- Recursion can be viewed as structural interchangeability, which is learnable as a productive generalization.

Discussion

- We do not argue the *ability* of recursion is acquired through distributional learning (e.g. Hauser, Chomsky, & Fitch, 2002), but rather: how do learners know in which specific domains the ability of recursion can be freely applied?
- We are focused on the role of purely distributional learning; we do not deny the role of other factors (e.g. semantics, phonology) in the acquisition of recursive structures.

Discussion

Why are unattested recursive strings in the Productive condition rated lower than attested strings and one-level strings?

- Complex structure, short duration.
- Structures with deeper embedding are rated lower even in natural languages (e.g. Christianson & MacDonald, 2009).

Discussion

Did our participants learn a hierarchical structure?

- Maybe, maybe not.
- But if they apply this sort of distributional learning to linear strings, they are also likely to apply it to hierarchical structures (e.g. Thompson & Newport, 2007).
- We can construct our language to be explicitly hierarchical and test learners' interpretation (e.g. Takahashi & Lidz).

Future directions

- Can speakers learn two structures in the same experiment, one freely recursive, the other restricted?
- Can this distributional learning be applied to explicitly hierarchical structures?
- How do learners coordinate different sources of evidence?
- At what age is this distributional learning available? (Aslin, 2017; Gervain, Macagno, Cogoi, Pena, & Mehler, 2008; Teinonen, Fellman, Naatanen, Alku & Huotilainen, 2009)

Thanks

To the Language and Cognition Lab at Penn for helpful comments.

Questions

Aux slides

Experiment - word distribution during exposure

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