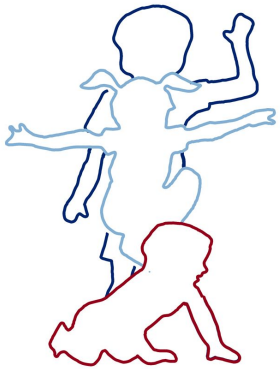
Acquiring recursive structures through distributional learning

Daoxin Li & Kathryn Schuler
The Child Language Lab, University of Pennsylvania
LSA 2022 Annual Meeting, January 7, 2022



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. Berwick & Chomsky, 2017; Hauser, Chomsky, & Fitch, 2002; Nevins, Pesetsky, & Rodrigues, 2009; Partee & Rooth, 1983; Pinker, 1994; Yang, 2013).

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 Manns Nachbars 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)

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).

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.

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.

Recursion of infinite depth must be learnable from level-one evidence!

The distributional learning proposal (Li et al., 2021)

Recursion as structural substitutability: *X_1's X_2* is recursive if *X_1* and *X_2* positions are substitutable. e.g. *cat's tail, kid's cat*

The distributional learning proposal (Li et al., 2021)

Recursion as structural substitutability: *X_1's X_2* is recursive if *X_1* and *X_2* positions are substitutable. e.g. *cat's tail, kid's cat*

Learning substitutability as a productive generalization: Generalize if a sufficiently large proportion of words attested in one position in the input are also attested in the other position in the input.

The distributional learning proposal (Li et al., 2021)

Recursion as structural substitutability: *X_1's X_2* is recursive if *X_1* and *X_2* positions are substitutable. e.g. *cat's tail, kid's cat*

Learning substitutability as a productive generalization: Generalize if a sufficiently large proportion of words attested in one position in the input are also attested in the other position in the input.

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

The distributional learning proposal (Li et al., 2021)

mom's …	…'s mom
daddy's …	…'s daddy
baby's …	…'s baby
cat's …	…'s cat
neighbor's …	…'s color
	…'s game
	…'s room
	…'s shape

$X_1 \longrightarrow X_2$

The distributional learning proposal (Li et al., 2021)

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., 2021).

Language	English	German
N in *A1* or *A2*	49	38
N in *A1* & *A2*	46	31
TSP threshold	36	28
Productive?	Yes	Yes

The distributional learning proposal (Li et al., 2021)

- 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., 2021).

Language	English		German	Mandarin	
Structure	X_1 's X_2	X_2 of X_1	X_2 von X_1	X_1 de X_2	X_1 X_2
$X_1 \rightarrow X_2$ Recursivity	Yes	No	Yes	Yes	No
N in X_1	22	45	42	40	34
N in X_1 & X_2	18	20	39	31	22
TSP threshold	15	34	31	30	25
$X_1 \rightarrow X_2$ Productive?	Yes	No	Yes	Yes	No

The distributional learning proposal (Li et al., 2021)

- 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., 2021).

Language	English		German	Mandarin	
Structure	X_1's X_2	X_2 of X_1	X_2 von X_1	X_1 de X_2	X_1 X_2
$X_1 \rightarrow X_2$ Recursivity	Yes	No	Yes	Yes	No
N in X_1	22	45	42	40	34
N in X_1 & X_2	18	20	39	31	22
TSP threshold	15	34	31	30	25
$X_1 \rightarrow X_2$ Productive?	Yes	No	Yes	Yes	No

The distributional learning proposal (Li et al., 2021)

- 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., 2021).

Language	English		German	Mandarin	
Structure	X_1's X_2	X_2 of X_1	X_2 von X_1	X_1 de X_2	X_1 X_2
$X_1 \rightarrow X_2$ Recursivity	Yes	No	Yes	Yes	No
N in X_1	22	45	42	40	34
N in X_1 & X_2	18	20	39	31	22
TSP threshold	15	34	31	30	25
$X_1 \rightarrow X_2$ Productive?	Yes	No	Yes	Yes	No

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

Experiment

Participants

- 48 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: $X_1 \rightarrow X_2$ 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:

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

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

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

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

Some words are more frequent than others, 44 string exposure corpus, 2 repetition.

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

Experiment - analysis

For each participant:

- Learning index = attested - ungrammatical
- Generalization index = unattested - ungrammatical

Experiment - analysis

For each participant:

- Learning index = attested - ungrammatical
- Generalization index = unattested - ungrammatical

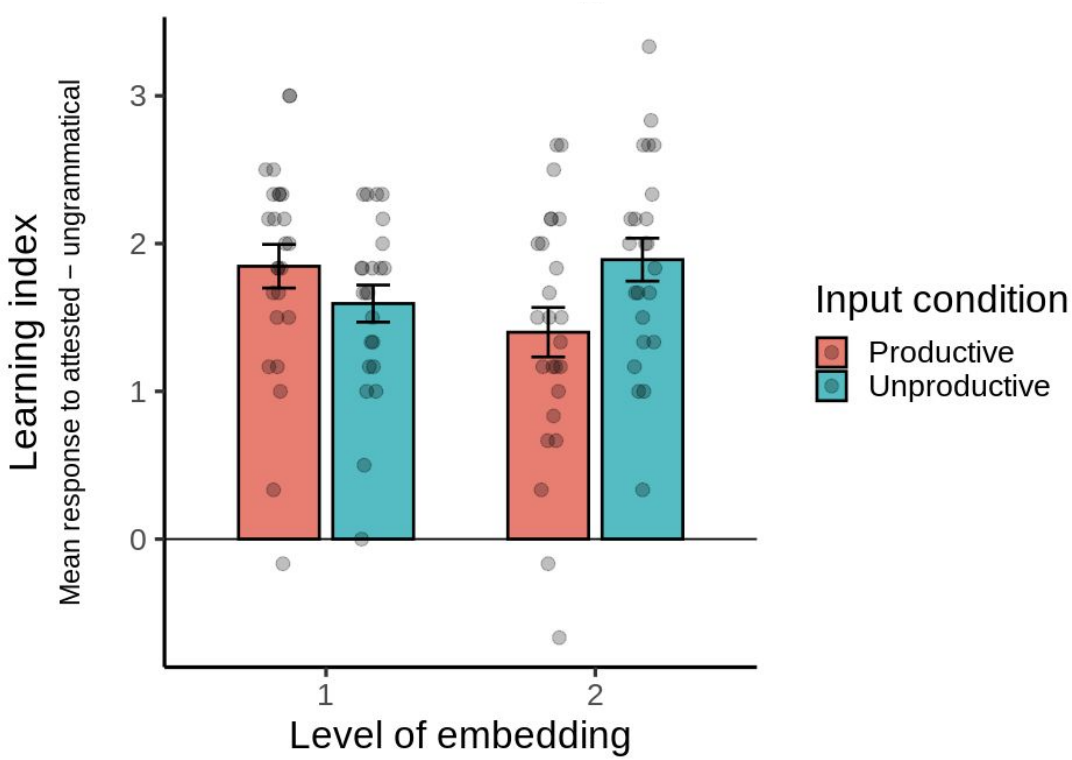
Mixed effects regression:

- DV: indices
- Fixed effects: Condition (productive, unproductive) and Level (1, 2)
- Random effects: by-participant random intercepts

Experiment - results

Learning index

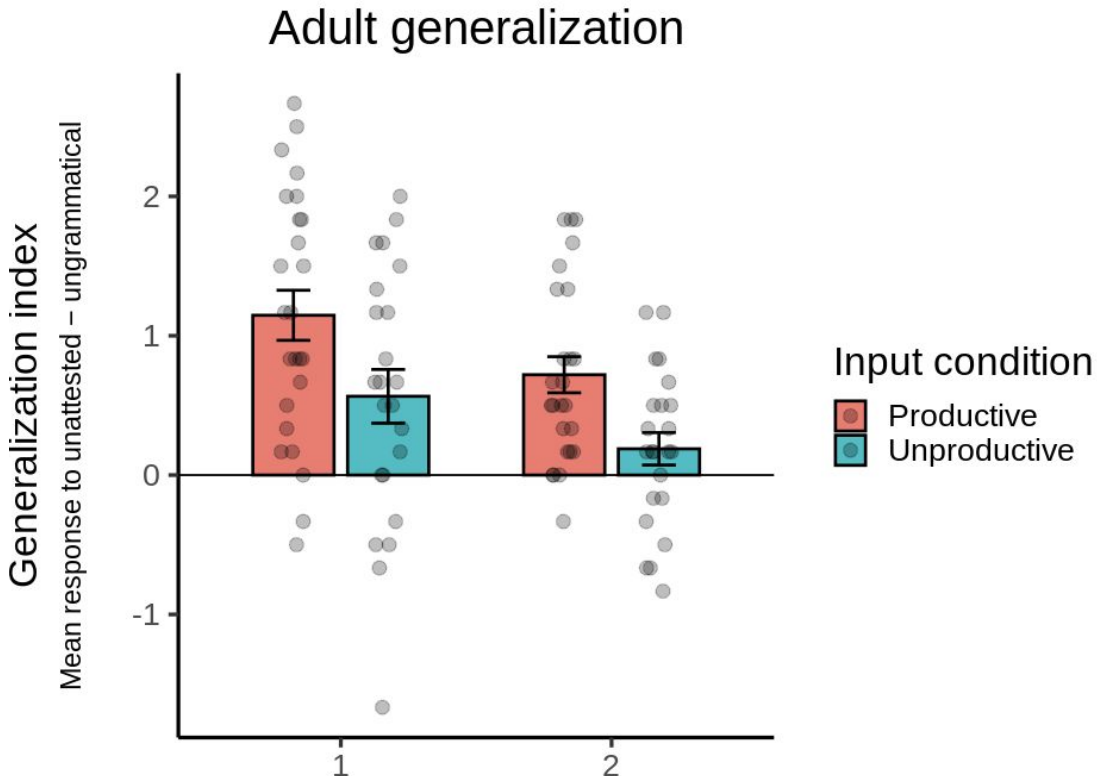
- No main effect of Condition (p=0.48)
- No main effect of Level (p=0.48)
- Significant interaction between Condition and Level (p=0.002)



Experiment - results

Generalization index

- Main effect of Condition (p=0.002)
- Main effect of Level (p=0.006)
- No significant interaction between Condition and Level (p=0.86)



Experiment - summary

- In both conditions, participants generalize a bit less for level-2 strings (but not unexpected).

- As predicted, participants generalize significantly more in the Productive condition than in the Unproductive condition at both levels of embedding.

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 substitutable.

- Recursion can be viewed as structural substitutability, 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?

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 did participants generalize less for 2-level strings?

- Processing factors.
- Structures with deeper embedding are rated lower even in natural languages (e.g. Christianson & MacDonald, 2009).

Discussion

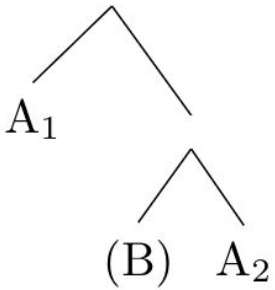
Did our participants learn a hierarchical structure?

- 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).

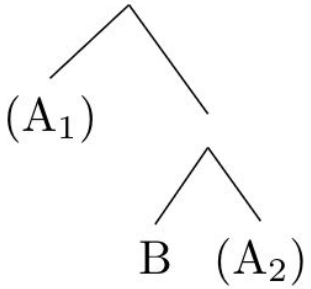
Discussion

Work in progress:

- Using distributional information to indicate the two different hierarchical structures:



A-head
e.g. ‘think she knows’



B-head
e.g. ‘dogs chase cats’

Discussion

Work in progress:

- The distributional learning proposal (Li et al., 2021) assumes children know which is the head.

- Prediction: With *A1-B-A2* input where *A1* and *A2* are substitutable, participants will learn only the A-head structure can be used recursively.

Discussion

At what age is this distributional learning available?

- It is suggested that distributional learning is available from birth (Aslin, 2017; Gervain, Macagno, Cogoi, Pena, & Mehler, 2008; Teinonen, Fellman, Naatanen, Alku & Huotilainen, 2009).
- Children experiment in progress on Lookit.

Discussion

Exposure phase

Discussion

Test phase

Future directions

- Can speakers learn two structures in the same experiment, one freely recursive, the other restricted?
- How do learners coordinate different sources of evidence?

Thanks

To Charles Yang and 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

Trees for English possessives