Distributional learning of recursive structures

Daoxin Li & Kathryn Schuler
Department of Linguistics, University of Pennsylvania
daoxinli@sas.upenn.edu, kschuler@sas.upenn.edu

Introduction

- The ability for recursion is a crucial part of the language faculty (e.g. Hauser, Chomsky, & Fitch, 2002).
- But languages differ regarding the depth, structure, and syntactic domains of recursive structures (Pérez-Leroux et al., 2018).
  (1) English:
  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)
  (2) German:
  a. das Buch von dem Nachbarn von dem Mann
  'the book of the neighbor of the man'
  b. Marias/Vaters/*Manns Buch
  'Maria's/father's/*man's book'
  c. *Peters Nachbars Buch
  '*Peter's neighbor's book' (Weiss, 2008)
- How do children learn which structures allow free recursive embedding and which structures are restricted?

The Distributional Learning Proposal (Grohe et al., 2021; Li et al., 2021)

- Child cannot learn recursive structures by observing multi-level embedding in the input:
  — Evidence for deep embedding is rarely attested in young children's input (e.g. Giblin et al., 2019).
  — A logical problem: no N-level embedding entails N+1 level embedding, never mind infinite embedding.
- Proposal: Recursion as structural substitutability
  X1's X2 is recursive if words that appear in X_1/X_2 can also be used *in* X_2/X_1.
- Proposal: Productivity and generalization
  Leaners acquire the generalization that structural substitutability holds for all words if a sufficiently large proportion of words attested in one position in the input are also attested in the other position in the input.
- Corpus studies have supported the proposal: There is sufficient evidence for different kinds of recursive structures in the input (Grohe et al., 2021; Li et al., 2021; Yang, 2021).
- But do learners indeed utilize the distributional information as predicted by the proposal?

Methods

- Participants: 50 native English-speaking adults on Prolific
- Exposure: *X1-ka-X2* artificial language strings, Zipfian distribution, 44 string exposure corpus, 2 repetition, no referential world (e.g. 'kewa-ka-nogi')

Methods

- Conditions:

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

The distribution is consistent with several metrics of productivity (e.g. Aronoff, 1976; Bybee, 1995; Yang, 2016).

- Test: On a scale of 1 to 5, is this string from the language you have just heard?

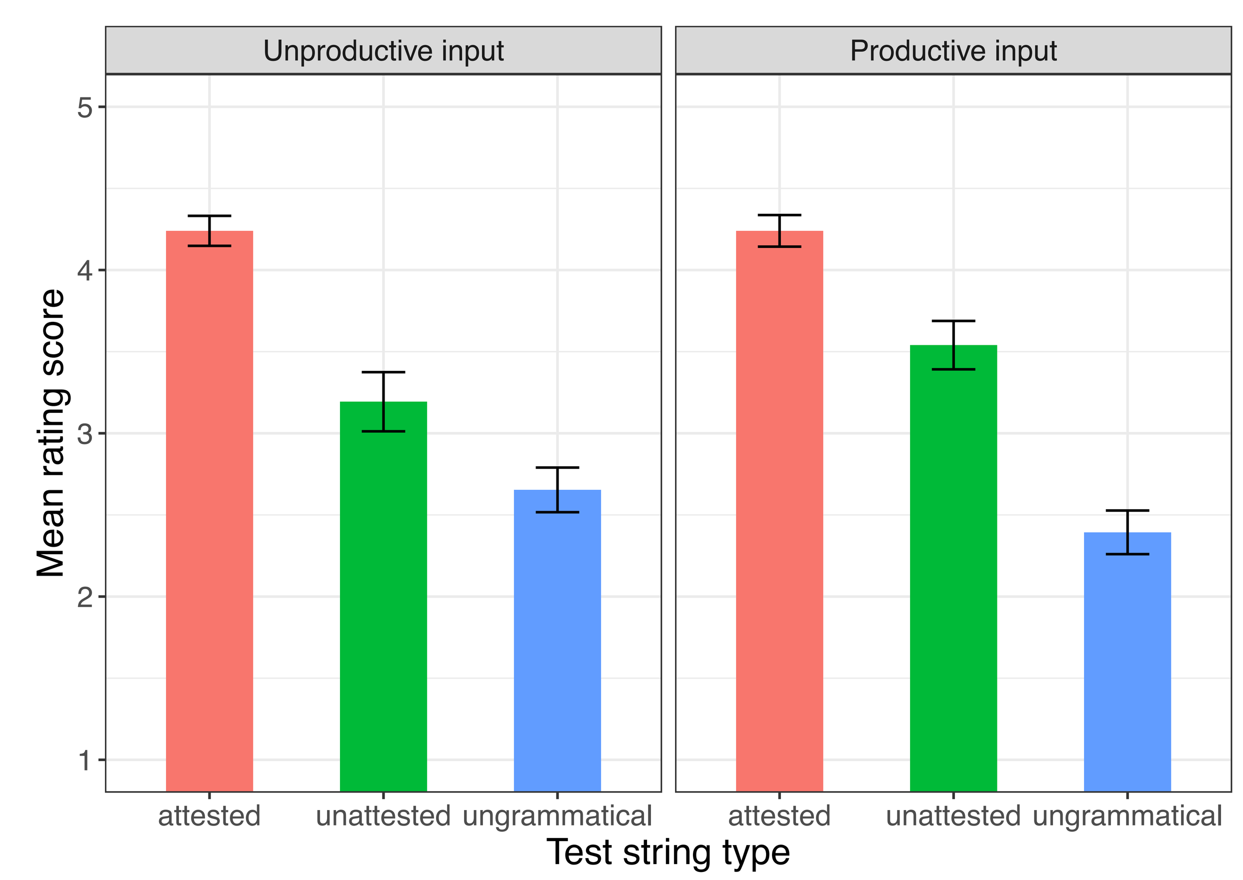
Sample test strings in Unproductive condition (*sane, tesa* and *tana* are never attested in *X2* position during exposure)
Word attested in the position; word unattested in the position

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*

- Prediction: Participants from the Productive condition will rate unattested strings higher than participants from the Unproductive condition at both one and two embedding levels.

Results

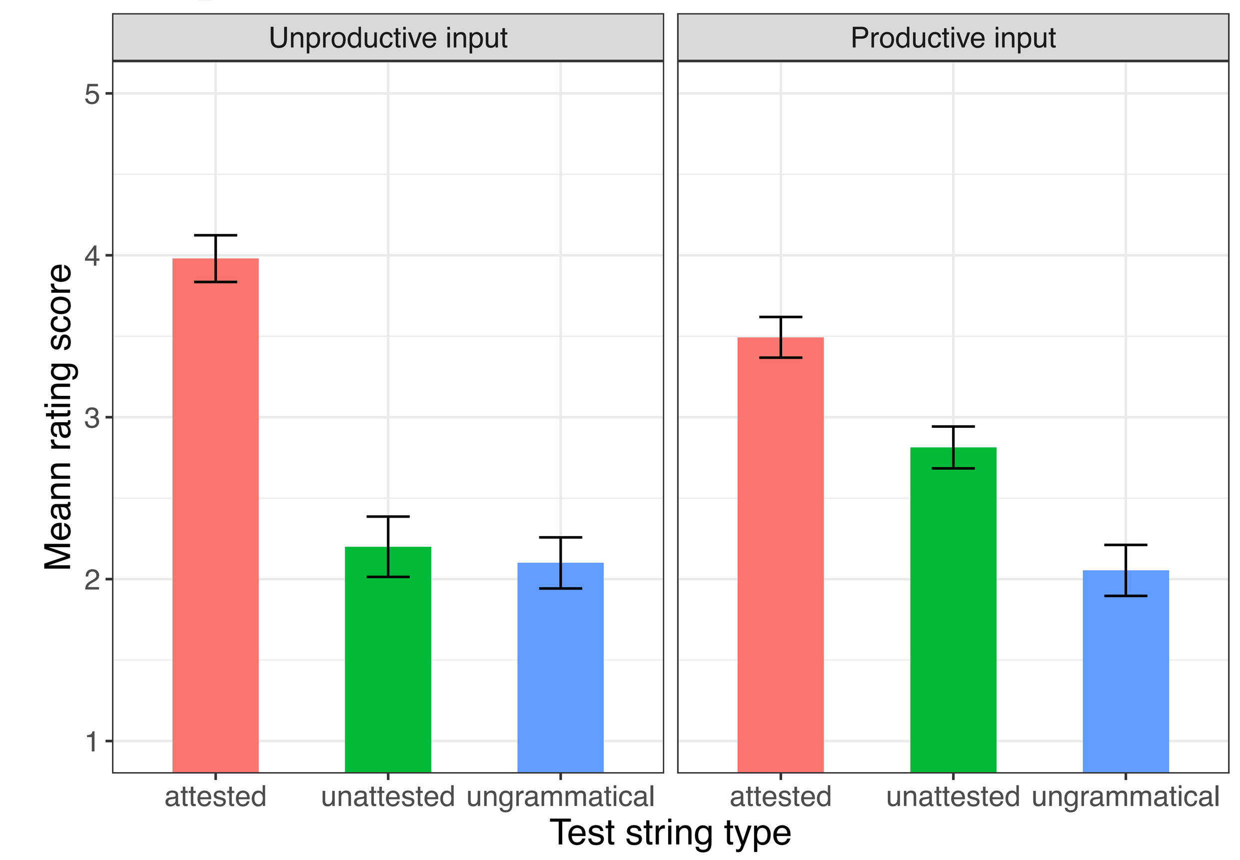
- Ordinal regression:
  DV: rating score as an ordered factor from 1 to 5
  Fixed effects: test string Type (attested, unattested, or ungrammatical) and Condition (Unproductive, Productive)
  Random effects: by-participant random intercepts and random slopes for Type
- 1-level strings:



No main effect of Condition ($p = 0.90$)
Significant main effect of Type ($p < 0.001$)
Significant interaction between Type and Condition ($p = 0.01$)
Unattested strings: marginally lower in Unproductive condition than in Productive condition ($p = 0.09$)

Results

- 2-level strings:



No main effect of Condition ($p = 0.81$)
Significant main effect of Type ($p < 0.001$)
Significant interaction between Type and Condition ($p < 0.001$)
Unattested strings: lower in Unproductive condition than in Productive condition ($p < 0.01$)

Conclusion & Future Directions

- 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.
- 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? (e.g. Thompson & Newport, 2007)
  — 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)

Selected References

Grohe, L., Schulz, P., & Yang, C. (2021). How to learn recursive rules: Productivity of prenominal adjective stacking in English and German. Paper presented at GALANA-9. **Hauser**, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*. **Li**, D., Grohe, L., Schulz, P., & Yang, C. (2021). The distributional learning of recursive structures. In *Proceedings of BUCLD-45*. **Pérez-Leroux**, A. T., Peterson, T., Castilla-Earls, A., Béjar, S., Massam, D., & Roberge, Y. (2018). The acquisition of recursive modification in NPs. *Language*. **Yang,** C. (2021). Productivity, recursion and the discovery procedure. Talk at Recursion across Languages workshop.