The distributional learning of recursive structures

Daoxin Li[1], Lydia Grohe[2], Petra Schulz[2], Charles Yang[1]
[1]University of Pennsylvania, [2]Goethe University Frankfurt
The 34th CUNY Conference on Human Sentence Processing
March 5, 2021

Recursion vs. recursive structures

- The ability to form recursive structures is almost surely innately available to all human language users (e.g. Hauser et al. 2002).

- But languages do differ with regard to the domain of recursion, which must be learned on the basis of language specific experience (e.g. Pérez-Leroux et al. 2018).

Within- and cross-linguistic differences in recursive structures

(1) English: nouns can be infinitely stacked in the *s*-possessive, but not in the postnominal *of*-possessive:

a. the man’s neighbor’s book

b. ?*the book of the neighbor

c. *the book of the neighbor of the man

Within- and cross-linguistic differences in recursive structures

(2) German: the *von*-possessive ‘of’ can embed freely (a), but the *s*-possessive is restricted already at level one (b) and cannot embed infinitely (c):

a. das Buch von dem Nachbarn von dem Mann

the book of the neighbor of the man

‘the book of the neighbor of the man’

b. Marias/Vaters/*Manns Buch

Maria’s/father’s/*man’s book

‘Maria’s/father’s/*man’s book’

c. *Peters Nachbars Buch

Peter’s neighbor’s book

‘*Peter’s neighbor’s book’

Within- and cross-linguistic differences in recursive structures

(3) Mandarin Chinese: nouns embed freely when the possessive marker *de* is present, but not when it is omitted:

a. na ren de linju de shu

that man GEN neighbor GEN book

'that man's neighbor's book'

b. *na linju shu

that neighbor book

'*that neighbor's book'

c. *na ren shu

that man book

'*that man's book'

How should recursive structures be formulated? And how are they acquired by children?

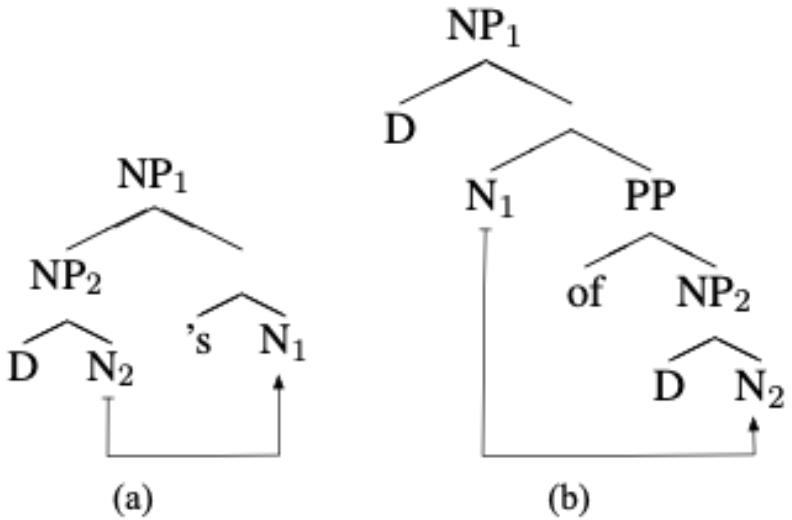
The mere attestation of multi-level embedding cannot be sufficient to support recursion

- Multi-level embedding, like all complex linguistic structures, would be vanishingly rare in the input data.

- There is no principled reason why the presence of *N*-level embedding would ensure even (*N*+1)-level embedding, never mind infinite embedding.

Proposal: Recursion as structural commutativity

- A structure is recursive if nouns that appear in N_1/N_2 can also be used in N_2/N_1
- Non-head Noun $\mapsto$ Head Noun



Proposal: Recursion as structural commutativity

- Learning recursion: the acquisition of lexical items for which structural commutativity holds.

e.g. *the **mother**'s car*, *the boy's **mother**,*

→ *s*-possessive is recursive for ***mother***

→ *mother's mother, the 40-year-old mother's 80-year-old mother's 120-year-old mother…*

- Children form generalizations over attested nouns such as ***mother*** above

Proposal: Productivity and generalization

- How to learn an infinite grammar from a finite sample of data?

- The Tolerance/Sufficiency Principle (TSP):

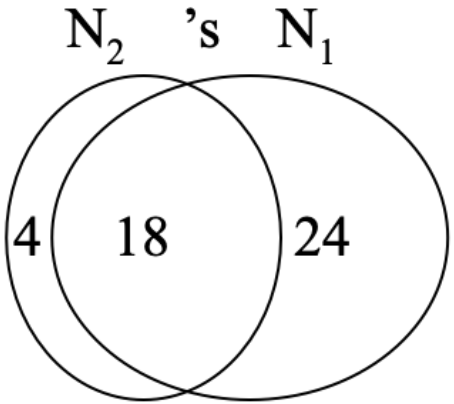
  A rule that holds for *(N-e)* items is productive iff $e \leq \theta_N = N/\ln N$, where N pertains to the child learner's modest, and likely high-frequency, vocabulary (Yang 2016).

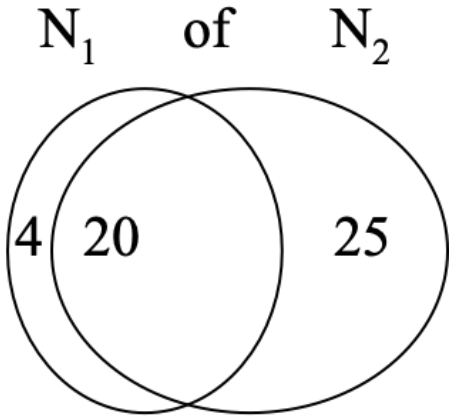
Corpus study

Language	Corpora	Input words	Structures
English	All English corpora on CHILDES	12.6 million	N_2's N_1
			N_1 of N_2
German	5 German corpora on CHILDES	3.5 million	N_1 von N_2
			N_2's N_1
Mandarin	All Mandarin Corpora on CHILDES	1.7 million	N_2 de N_1
			N_2 N_1

- Extract the possessive structures where the head noun in N_1 and/or N_2 position is known to young children (Carlson et al. 2014, Hao et al 2008, Szagun et al. 2006).

Results: Structural commutativity in English





- The s-possessive ($N_2 \mapsto N_1$): 22 N_2 nouns, 18 appear in N_1 (TSP requires 15)
  - Recursive for possessor
- The *of*-possessive ($N_1 \mapsto N_2$): 24 N_1 nouns, 20 appear in N_2 (TSP requires 17)
  - Recursive for possessum

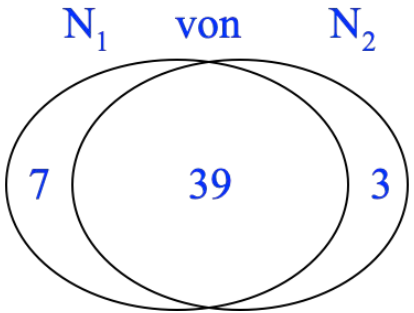
Results: Semantic conditions for English possessives

Construction	N	Counts	20 Most Frequent Nouns
s-possessive (N_2's N_1)	N_1	42	name, head, hair, nose, mouth, room, hat, house, car, bed, hand, chair, food, cup, mommy, juice, water, truck, daddy, school
	N_2	22	baby, daddy, boy, mommy, dog, girl, man, cat, bear, fish, truck, train, cup, name, door, day, way, hat, color, car
of-possessive (N_1 of N_2)	N_1	24	*piece*, top, *bit*, picture, name, *cup*, time, color, day, head, door, *box*, way, hair, thing, mouth, book, school, room, man
	N_2	45	cheese, cake, head, book, train, house, water, milk, box, baby, hair, car, juice, food, school, fish, hat, day, dog, man

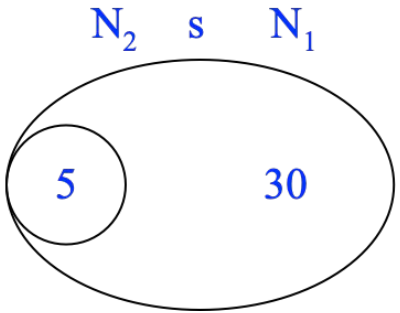
Possessor in 's-possessive: internal (kin, body parts, attributes) and external possession (ownership)

Possessum in *of*-possessive: measure words (*piece, bit, cup, box*) and internal possession

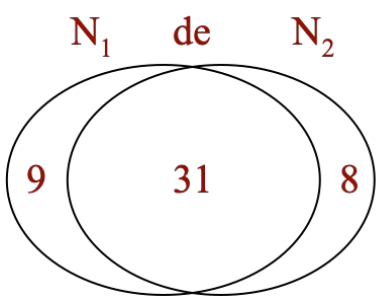
Results: German and Mandarin



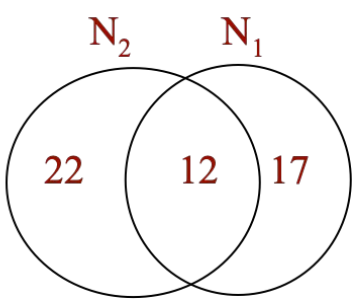
$N_1 \leftrightarrow N_2$



Only kinship terms



$N_1 \leftrightarrow N_2$



Kinship terms, measure words

Conclusion

- Recursion derives from the *commutativity* of two structural positions.
- Therefore, recursion can be acquired as a productive generalization from level-1 input data for specific syntactic domains.

Thanks

To Lila Gleitman, Norbert Hornstein, Julie Legate, Joan Maling, Ana Pérez-Leroux, Yves Roberge, Tom Roeper, Jill de Villiers, Helmut Weiß, and the participants of a research seminar at the University of Pennsylvania for discussion.

Questions