Modelling the distributional learning of verb argument structure

Daoxin Li
daoxinli@sas.upenn.edu
Department of Linguistics, University of Pennsylvania

Introduction

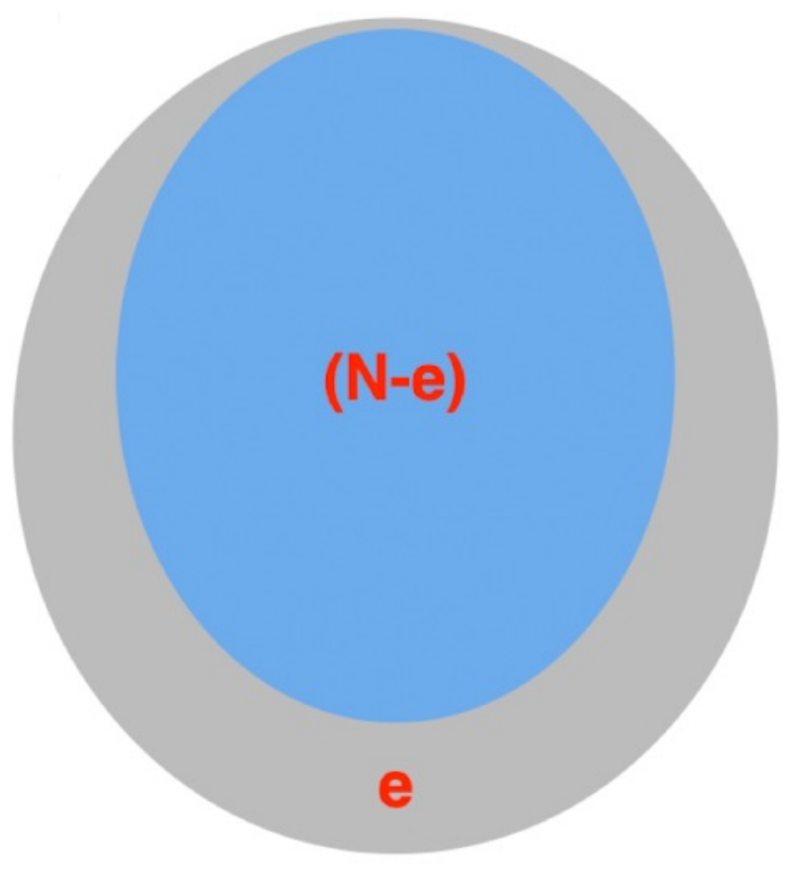
- Linguists agree there are systematic mappings between the syntax and semantics of a verb[1-3] (e.g., (1-3)) and that children know these mapping rules from a young age[4-5] .

(1) Causation – transitive (e.g., *open, break*).
(2) Transfer – dative construction (e.g., *give, send*).
(3) Motion – PP (e.g., *put, move*).

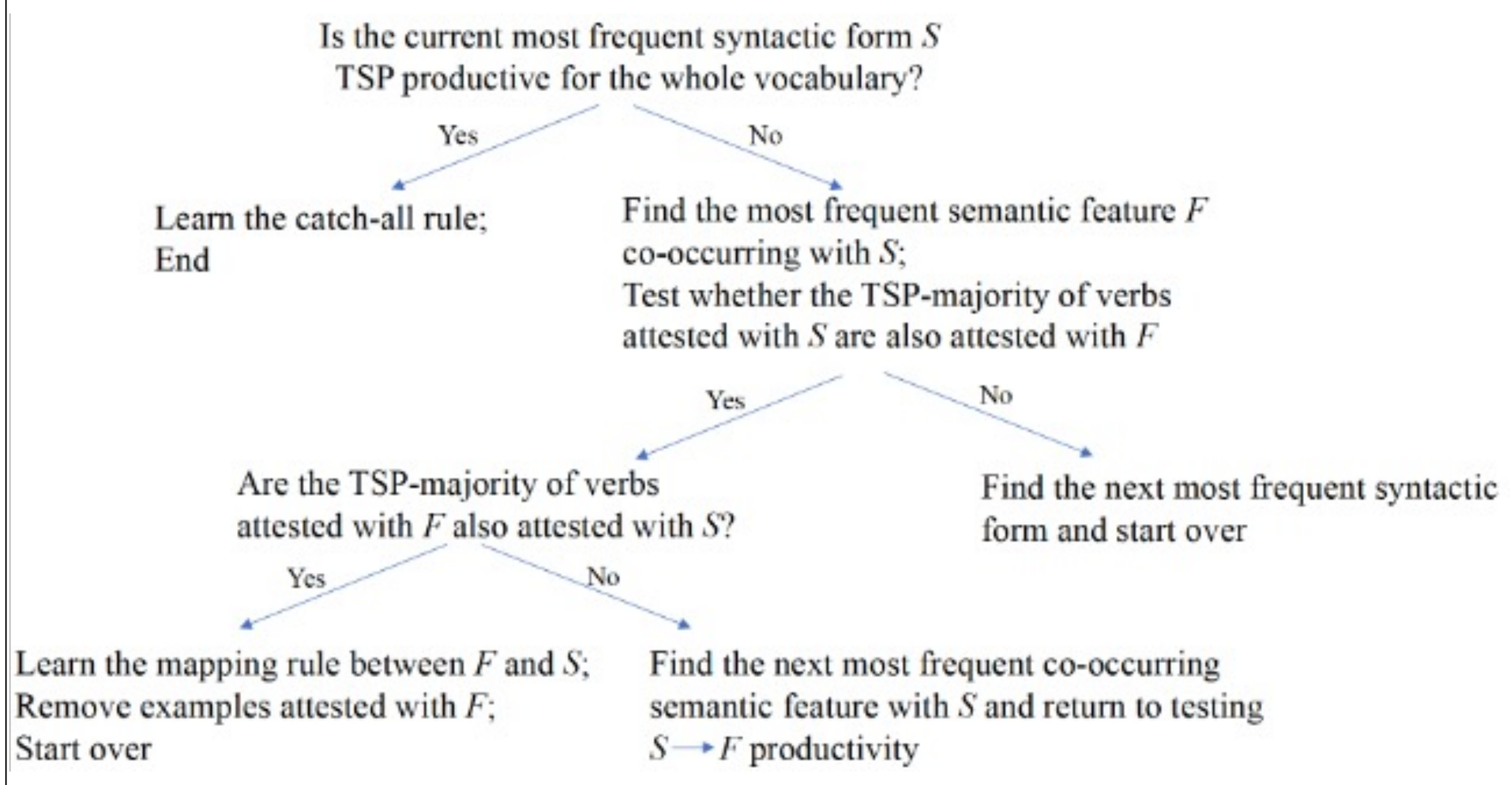
- Where does this knowledge come from?
    - Unlikely entirely universal or innate given the considerable variabilities across languages and idiosyncrasies within[6-8].
- This work: A computational model that automatically learns productive rules between syntax and semantics.
- We show the rules are learnable from child-directed speech without assuming any prior syntax-semantics associations.

Model Description

- Based on the Tolerance/Sufficiency Principle (TSP): A generalization *R* defined over *N* items is productive iff the number of items attested to follow *R* exceeds *N-N/lnN*[9].



- Model input: (verb, syntactic frame, semantic features) e.g., (“open”, “V NP”, (“act”, “causation”))
- Major procedures of the model:



Data

- Source: Input to Alex (1;4-3;5) from Providence corpus[10]
- Vocabulary: 60 most frequent action verbs in early child English[11-12]
- Extracted caregivers’ sentences containing these verbs – 1752 sentences
- Manually coded the syntactic frame and the semantic features from videos (*act, causation, motion, transfer, change-of-state, creation, communication,* all assumed identifiable to young learners[e.g., 2])
- To model the real-world challenge, we did not exclude sentences where the accompanying event in the video did not match the sentence (N=302, ~20%)

Results

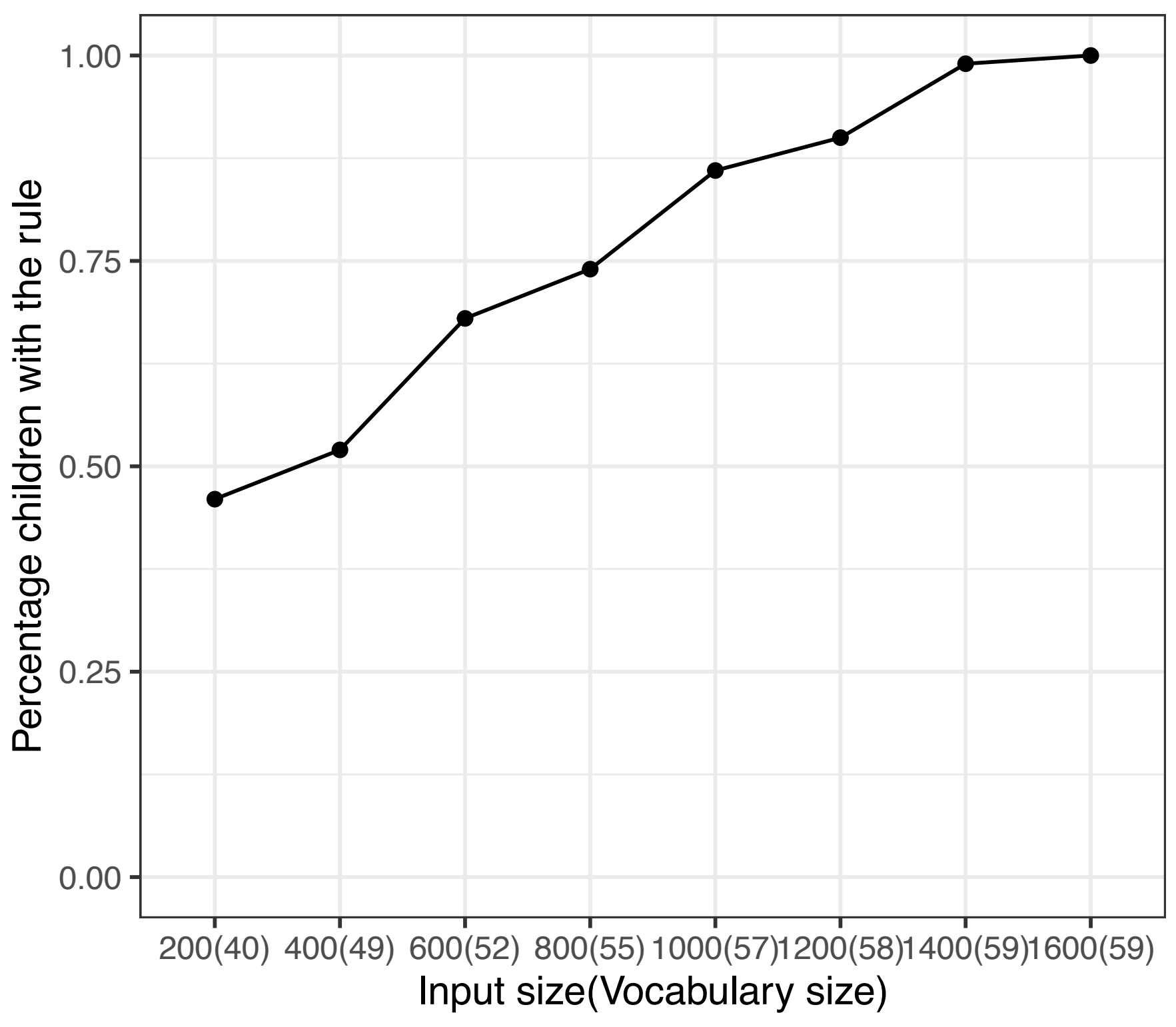
- Mapping rules learned when trained on all data:

(4) Causative – transitive (5) Act – intransitive
(6) Change of state – intransitive
(7) Creation – double object

- Additional rules learned when trained on clean data (non-matching examples excluded):

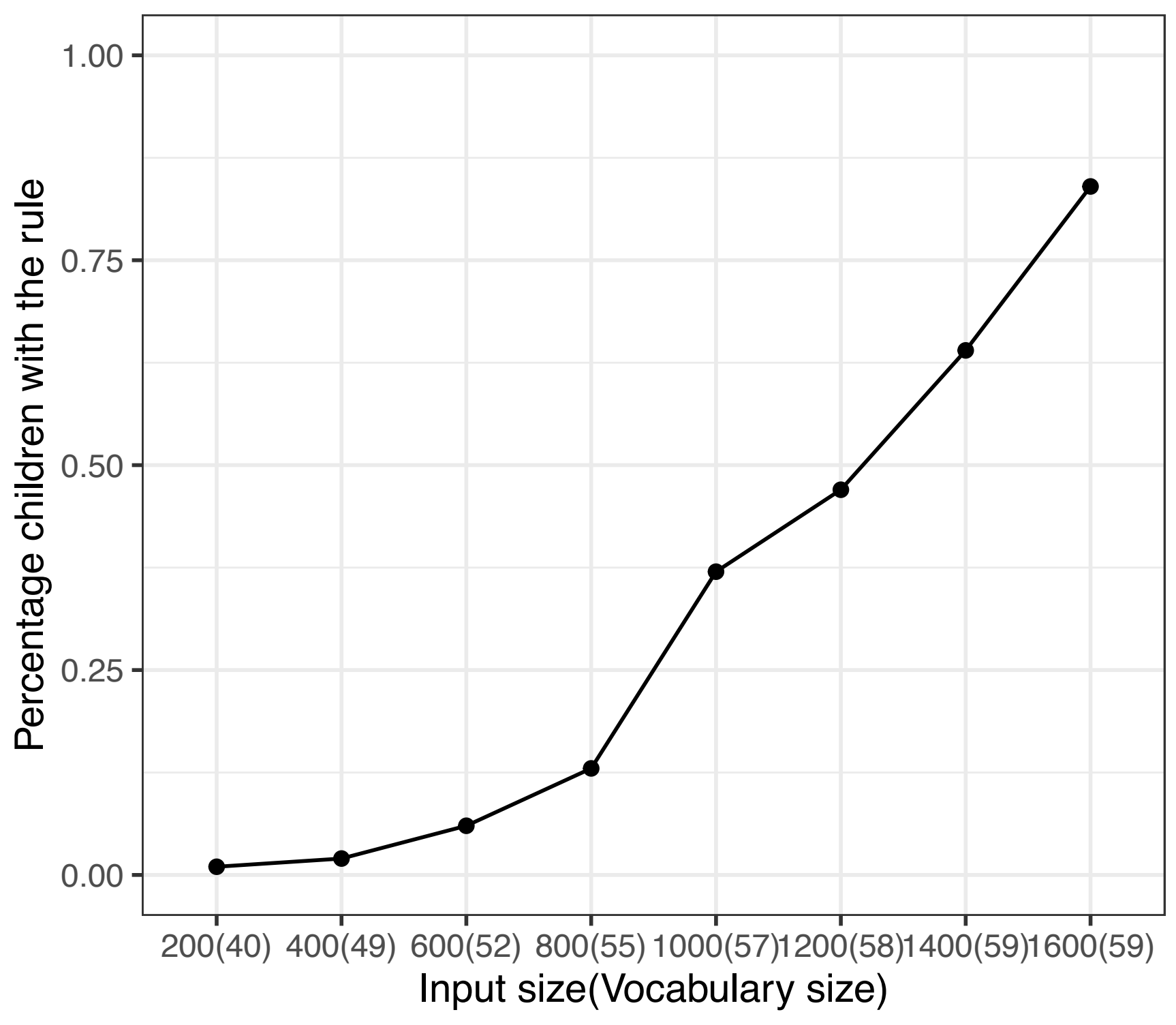
(8) Motion – PP (9) Transfer – dative

- Rule (4), which is acquire by children before 2;0, is indeed learned early and robustly by the model despite modest vocabulary size and input. Modeling 100 children with different input:



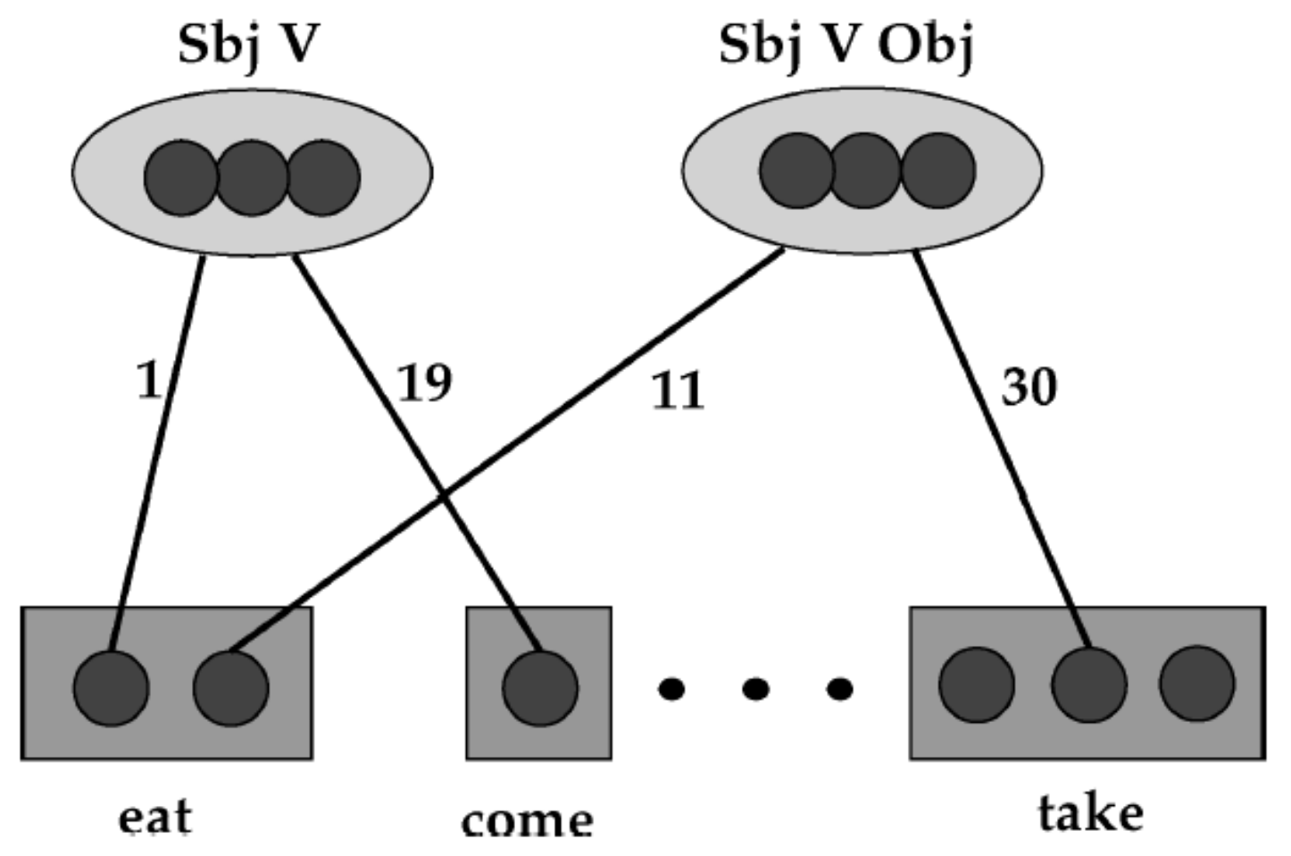
- The model also captures well-documented causative overgeneralizations (10): intransitive & change of state –> transitive

(10) a. Kendall fall that toy. (Kendall, 2;3)
b. I’m gonna ... disappear something... (E, 3;7).
c. He’s gonna die you, David. (Hilary, 4+) [13]



Model Comparisons

- Comparisons against a Bayesian model[14]: Learning probabilistic associations between syntactic frames and semantic features by grouping input pairs into constructions based on unsupervised Bayesian clustering



- A fundamental difference: The Bayesian model relies on token frequency, our model only uses type frequency

Model comparisons

- We trained the Bayesian model with Alex’ input data and examined the acquired knowledge as in [14].
- Semantic features of highest probabilities in different syntactic frames:

Syntactic frame	All data	Clean data
V NP	act (2.9×10^{-7}), causation (1.7×10^{-7}), communication (1.6×10^{-7})	act (3.3×10^{-7}), causation (2.5×10^{-7}), communication (2.0×10^{-7})
V	act (2.8×10^{-7}), causation (1.2×10^{-7}), change of state (7.3×10^{-8})	act (3.3×10^{-7}), causation (1.4×10^{-7}), change of state (8.8×10^{-8})
V NP NP	act (1.0×10^{-7}), causation (5.5×10^{-8}), transfer (4.7×10^{-8})	act (1.1×10^{-7}), transfer (9.5×10^{-8}), causation (5.8×10^{-8})
V NP to NP	act (6.0×10^{-8}), causation (5.4×10^{-8}), transfer (5.0×10^{-8})	transfer (1.5×10^{-7}), act (8.2×10^{-8}), causation (8.2×10^{-8})
V NP PP	act (5.8×10^{-8}), causation (5.8×10^{-8}), caused motion (5.7×10^{-8})	act (9.0×10^{-8}), causation (8.9×10^{-8}), caused motion (8.8×10^{-8})
V PP	act (1.7×10^{-7}), motion (9.2×10^{-8}), causation (6.0×10^{-8})	act (1.5×10^{-7}), motion (1.4×10^{-7}), causation (7.0×10^{-8})

- Syntactic bootstrapping test: Compare probabilities of semantic features given a syntactic frame

Test pair	Probability (all data)	Probability (clean data)
‘V NP’ – ‘act & causation’ (matched)	5.9×10^{-8}	7.1×10^{-8}
‘V NP’ – ‘act’ (unmatched)	1.9×10^{-9}	2.3×10^{-9}
‘V’ – ‘act’ (matched)	5.9×10^{-8}	7.1×10^{-8}
‘V’ – ‘act & causation’ (unmatched)	5.9×10^{-8}	7.1×10^{-8}

- Production test: Find syntactic frames of highest probabilities given the semantic features

Semantic features	All data	Clean data
act	V (5.9×10^{-8})	V (7.1×10^{-8})
act & causation	V (5.9×10^{-8}), V NP (5.9×10^{-8}), V NP to NP (1.0×10^{-8})	V NP (7.1×10^{-8}), V (7.1×10^{-8})

- Problem: High token frequency of optional transitive verbs leads to a strong association between causation meaning and intransitive frame, which is not a regular rule in English (Not a problem in our model because type frequency is low).
- Account for causative overgeneration: Unaccusative verbs are used in the transitive frame when there is a causative agent given the acquired association between causation and transitivity; will retreat with more input, since knowledge of individual words will have a stronger influence as token frequency increases.
    - Problems: (1) Children also overgeneralize unaccusative verbs without a causative agent (e.g., *die, disappear)*; (2) It predicts more frequent verbs to retreat earlier, which is not true: Ross from MacWhinney corpus overgeneralizes all these words around ages 3-4[15]:

Verb	Frequency
disappear	152
stay	2,662
fall	2,819
go	55,689

Conclusion

- Rules of verb argument structure are learnable from realistic input data without universal, innate linking knowledge.
- Our threshold-based model acquires knowledge that is more accurate and more consistent with human behavior than the Bayesian model.
- Future work should apply the model to larger corpora and different languages.
- The model can also be applied to the acquisition of other generalizations.

References. [1] Gruber, J.S. 1965. *Studies in lexical relations.* [2] Jackendoff, R. 1978. *Linguistic theory and psychological reality*. [3] Ladusaw, W. & D. Dowty. 1988. *Syntax and semantics.* [4] Pinker, S. 1989. *Learnability and cognition: The acquisition of argument structure.* [5] Yuan, S. & C. Fisher. 2009. *Psychological Science.* [6] Fisher, C., H. Gleitman & L.R. Gleitman. 1991. *Cognitive Psychology*. [7] Bavin, E.L. & S. Stoll (eds.). 2013. *The acquisition of ergativity.* [8] Bowerman, M. & P. Brown (eds.). 2008. *Crosslinguistic perspectives on argument structure: Implications for learnability.* [9] Yang, C. 2016. *The price of linguistic productivity How children learn to break rules of language.* [10] Demuth, K., J. Culbertson & J. Alter. 2006. *Language and Speech*. [11] Carlson, M., M. Sonderegger & M. Bane. 2014. *Journal of Memory and Language.* [12] Rowe, M.L. & Goldin-Meadow, S. 2009. *Science*. [13] Bowerman, M. 1982. *Language acquisition: The state of the art.* [14] Alishali, A. & S. Stevenson. 2008. *Cognitive Science.* [15] Irani, A. *Learning from positive evidence: The acquisition of verb argument structure.*

Acknowledgements. Thank you to Charles Yang, Kathryn Schuler, Julie Legate, John Trueswell, Marlyse Baptista, and participants of a research seminar at University of Pennsylvania for helpful comments and discussion.