Comparing memory-based and neural network models of early syntactic development

Eva Portelance, George Kachergis, and Michael C. Frank, *Stanford University*

PROBLEM

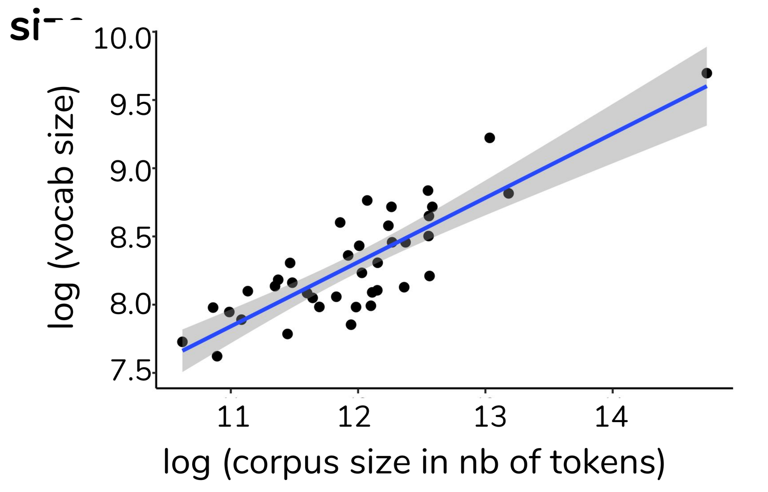
How syntactically productive are children's early utterances?

- We compare models of early syntactic development, the CBL of McCauley and Christiansen (2019) and an LSTM recurrent neural network model to determine which one better reproduces the syntactic production behavior of children.
- HYPOTHESIS: CBL will perform worse than the LSTM at predicting longer child utterances. Shorter utterances can be memorized by both models; Longer utterances likely require learning some intermediate structure.

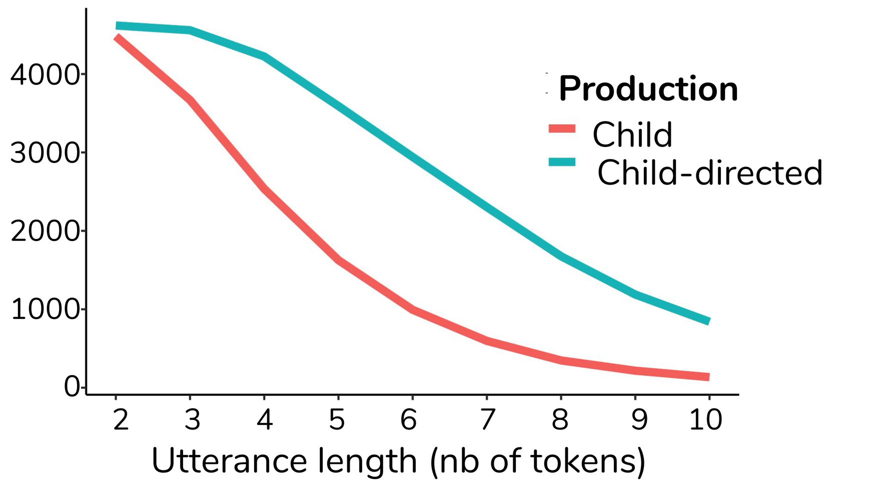
DATA

- 39 English CHILDES corpora
  - At least 1:20 child/caregiver
  - At least 20 000 words
- Train: 100% child directed utterances + 60% child utterances
- Test: 40% held out child utterances

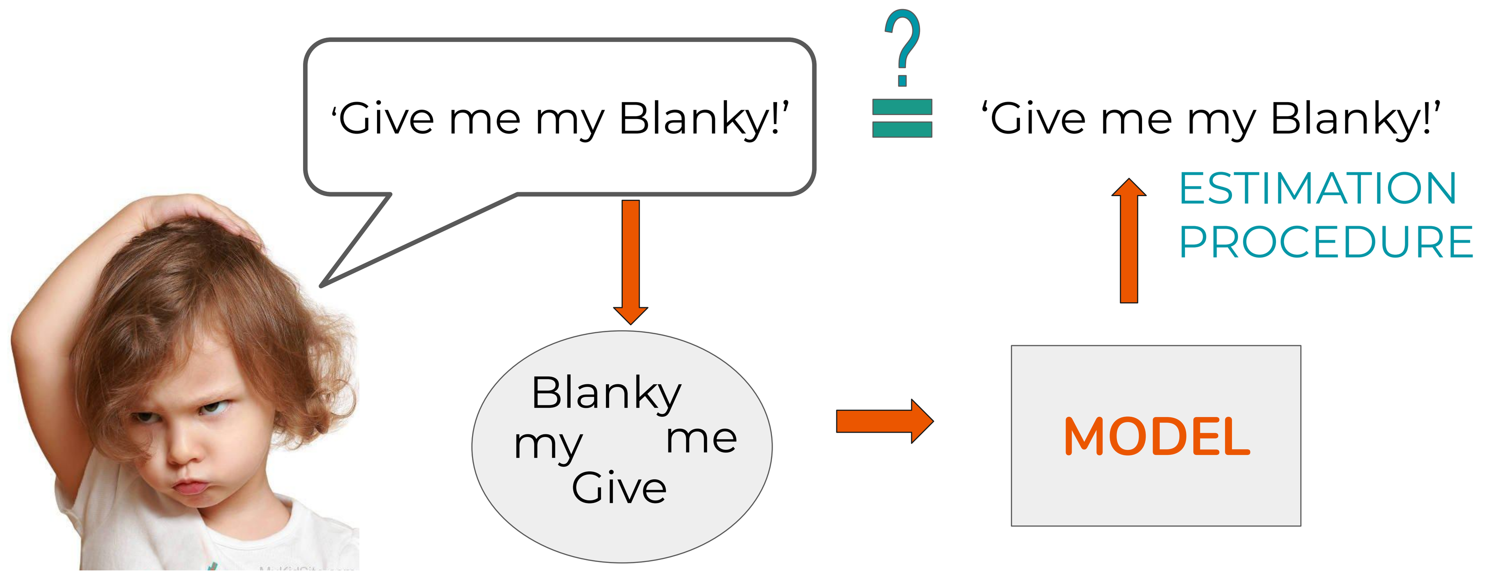
Vocabulary size in relation to corpus size



Average counts for each utterance length



PRODUCTION TASK



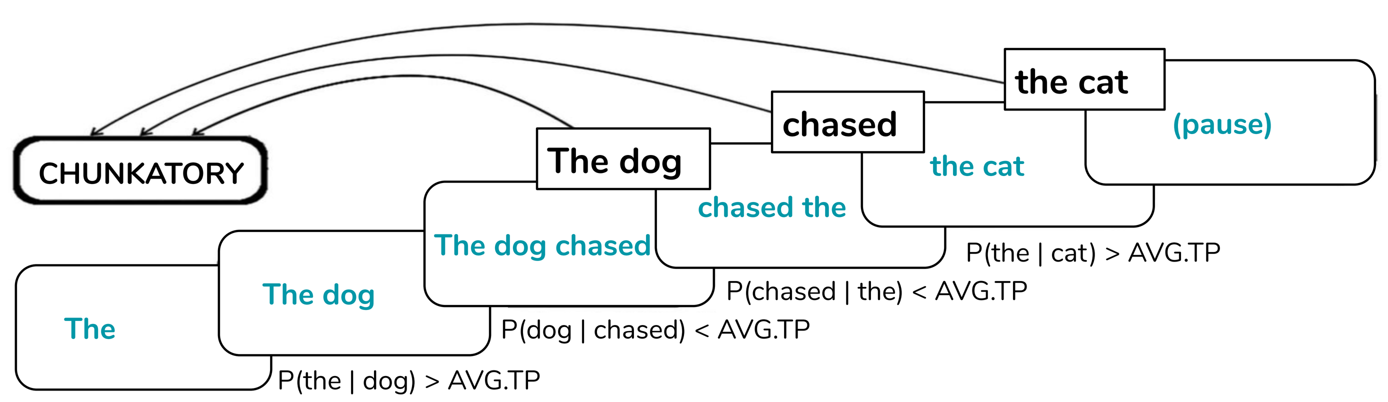
- Originally from McCauley & Christiansen (2019)
- Can models reproduce child utterances?
- 1 trained model/1 child (each CHILDES corpus)
- Models must reorder the tokens of a child utterance.
- For CBL, tokens are chunks (multi word); For LSTM, tokens are words.
- ESTIMATION PROCEDURES:
  - **Greedy decoder** - Always return the most probable next token (returns 1 solution).
  - **Beam Search decoder** - At each state, keep track of the k=5 most probable beams for the next state (returns 5 solutions).

Production Score = proportion of correctly reordered utterances from test set

MODELS

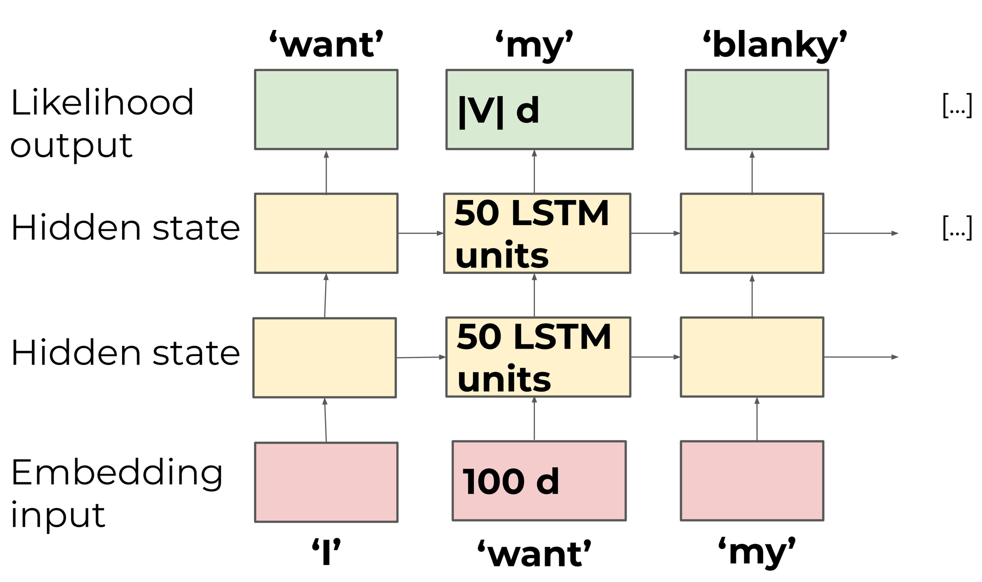
Chunk Based Learner (CBL) - McCauley & Christiansen (2019)

- Learns multiword chunks using backward transition probabilities — no other abstraction
- E.g. Chunk processing for 'The dog chased the cat.'



Long Short Term Memory Recurrent Neural Network (LSTM)

- Straightforward NLP model used for language modelling
- Able to learn longer dependencies (Linzen et al. 2016) and representational abstractions



RESULTS

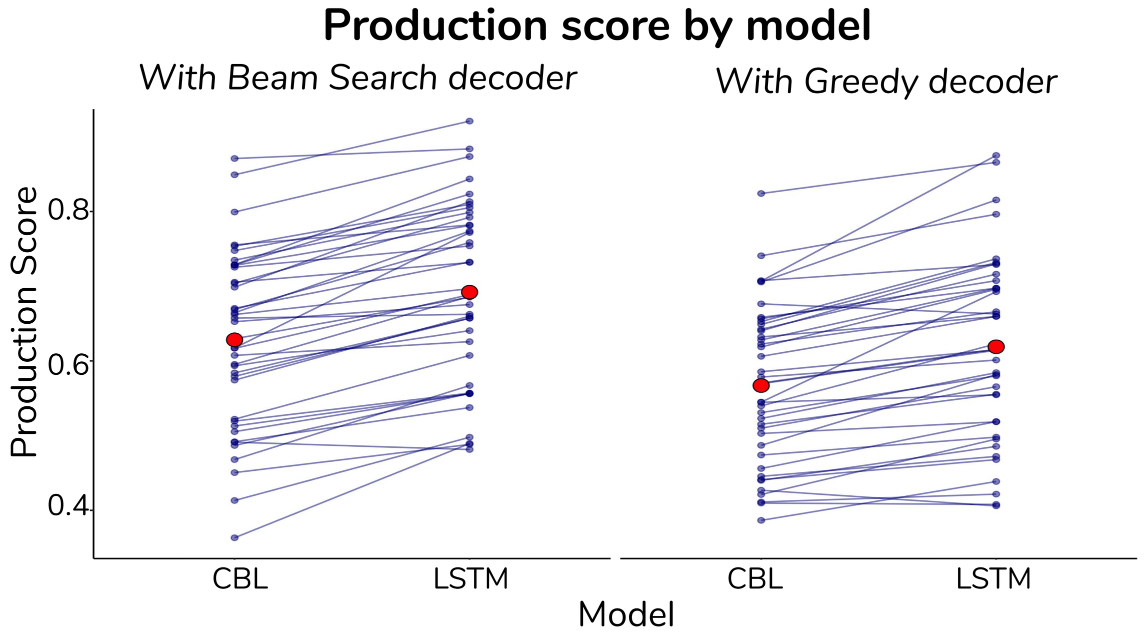
Mean production score using greedy decoder

- CBL: .57, 95%CI[.53-.60]
- LSTM: .62, 95%CI[.58-.66]

Mean production score using beam search decoder

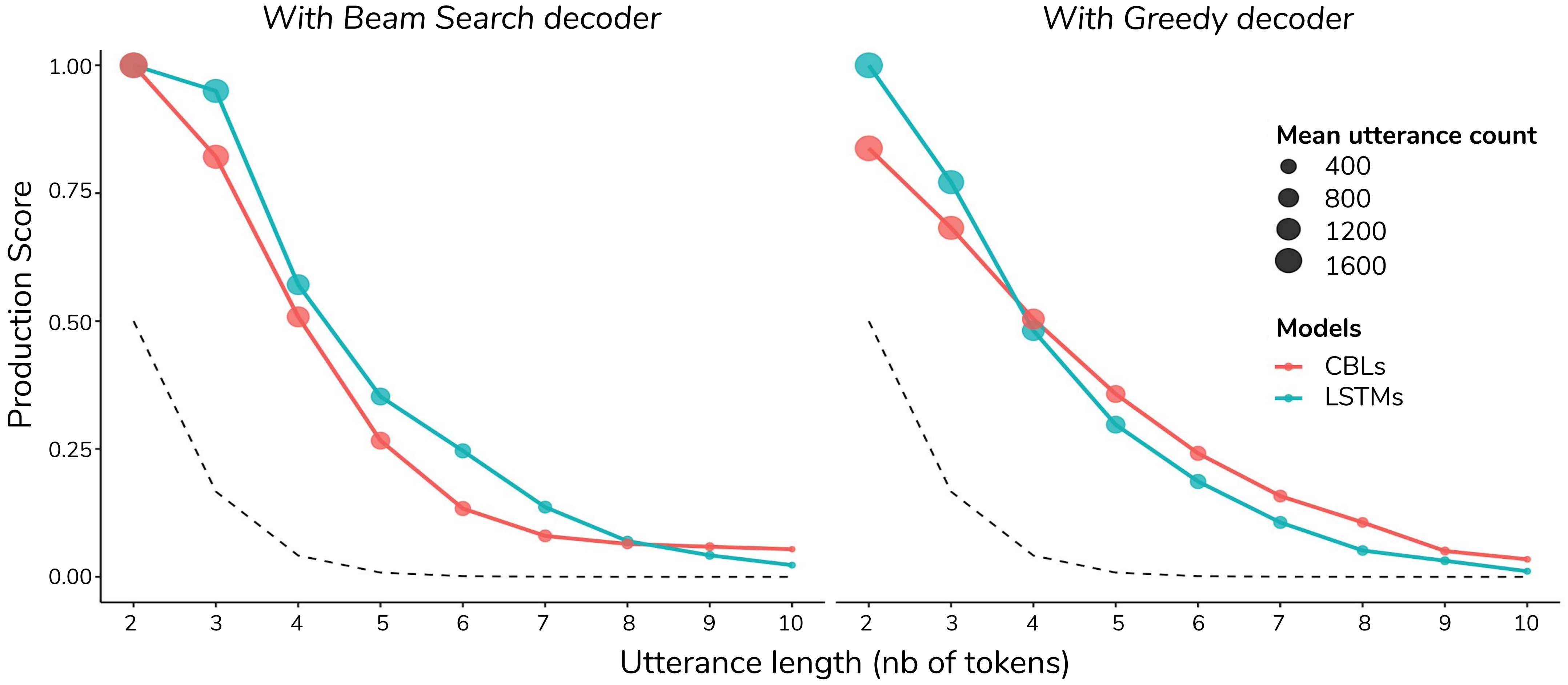
- CBL: .63, 95%CI[.59-.66]
- LSTM: .69, 95%CI[.65-.73]

Production score by model



- **BUT:** Performance is largely driven by shorter utterances (2- 4 words). Performance drops as utterances grow more complex.

Average model performance by utterance length



CONCLUSIONS

- LSTM has better performance overall, supporting our hypothesis that abstractions learned by LSTMs better model child production behavior.
- **BUT:** Neither model was able to reliably predict longer child utterances, suggesting that models learning more structured grammatical representations are more suited to describe children's syntactic acquisition.

REFERENCES

Linzen, T., Dupoux, E. and Goldberg, Y., 2016. Assessing the ability of LSTMs to learn syntax-sensitive dependencies. *Transactions of the Association for Computational Linguistics*, *4*, pp. 521-535.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. Third Edition.* Mahwah, NJ: Lawrence Erlbaum Associates.

McCauley, S.M. and Christiansen, M.H., 2019. Language learning as language use: A cross-linguistic model of child language development. *Psychological review*, *126*(1), pp. 1-51.

ACKNOWLEDGEMENTS

Thank you to the members of the Language and Cognition Lab and ALPS Lab for their helpful comments throughout!

This research was supported by the Social Sciences and Humanities Research Council of Canada.

Social Sciences and Humanities Research Council of Canada

Conseil de recherches en sciences humaines du Canada

Canada