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

Children's lexicon and syntactic abilities grow in tandem, resulting in a tight correlation between vocabulary size and grammatical complexity (Bates et al., 1994; Brinchmann et al., 2018; Frank et al., 2019). This relationship is consistent with the hypothesis that children's early grammatical abilities are well-described by lexicalized syntactic models, where syntactic representations are structured around lexical items encoding both function and form (e.g., Pollard & Sag, 1994; Goldberg, 2003). Some such models are strongly lexicalized, including no representational abstractions; others provide the capacity for learning abstractions.

We compare two such models, the Chunk-based Learner model (CBL) of McCauley and Christiansen (2019; MC2019) and a Long short-term memory (LSTM) recurrent neural network model. The CBL memorizes frequently-occurring chunks and does not learn any representational abstraction, similar to other pure-memory models (e.g., Perruchet & Vinter, 1999; Servan-Schrieber & Anderson, 1990). In contrast, the LSTM makes use of nested hidden layers to learn abstract representations that can predict sequential dependencies between words across a range of dependency lengths. Our goal is to determine the degree to which the emergent abstractions provided by this model help in understanding children's production behavior (Linzen et al., 2016). We hypothesized that the CBL would perform worse than the LSTM at predicting longer child utterances. While shorter utterances can be memorized by both model classes, longer utterances likely require some intermediate structure and might not be as amenable to chunking strategies.

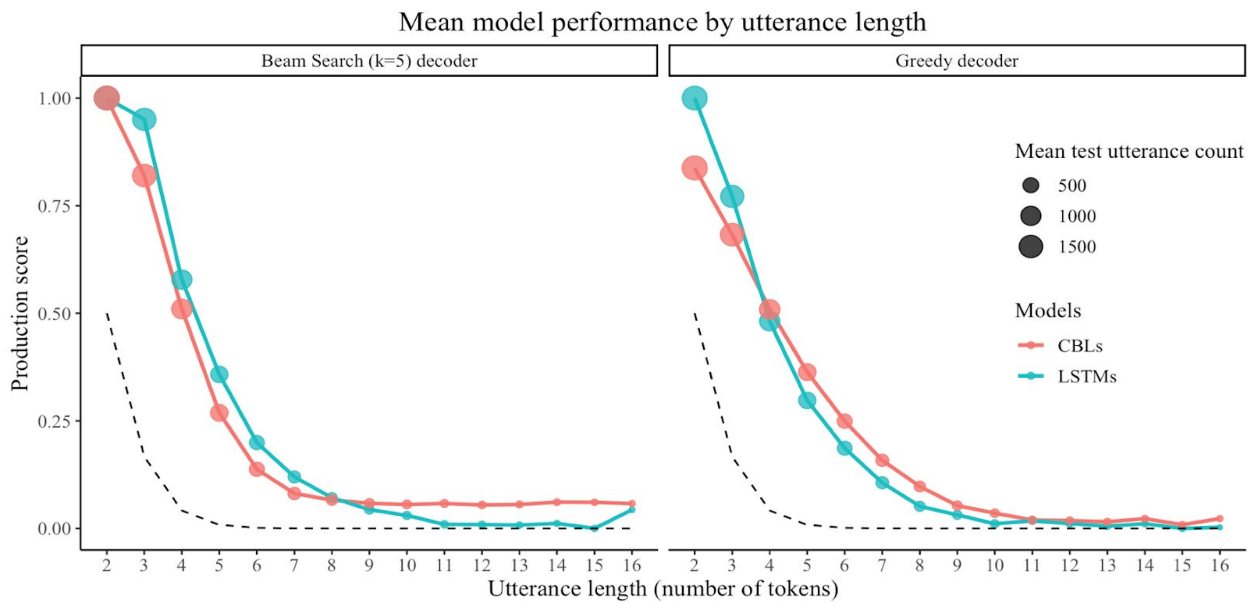
We trained models separately on 40 CHILDES (MacWhinney, 2000) English transcripts with 100% of adult data and 60% of child data, while 40% of child data was held out for evaluation (children's age distribution: Figure 1). We evaluated the models' ability to produce correct linear orderings for unordered sets of lexical items – chunks for CBL and vector representations of words for LSTM (same "production" task as MC2019). For a given child utterance, models produced rankings for different linear orderings of words. Production score was defined as the proportion of utterances in which the correct ordering was either (i) the best ordering found by a "greedy" decoder or (ii) in the top 5 orderings found by a beam search decoder.

The LSTM had better accuracy overall ("greedy" – LSTM: .62, 95%CI[.58-.66]; CBL: .58 [.55-.61] and beam search – LSTM: .69 [.65-.73]; CBL: .64 [.60-.68]). Even though the task favors memory-based models, since the number of items to reorder is smaller when using chunks and therefore the combinatorial possibilities less numerous, the LSTM performed better than the CBL for utterances of seven words or less with the beam search decoder (Figure 2), supporting our hypothesis that abstractions learned by LSTMs help model child production behavior. However, both models' overall performance was largely driven by the numerical predominance of short utterances, since absolute performance dropped substantially as utterances grew more complex, indicating that neither model appropriately captured the linear structures necessary for reliable production sequencing. These results suggest that models learning more structured grammatical representations may be necessary to describe children's syntactic acquisition.

499 words

Figure 1: A density plot of the age range of the 40 English speaking children in the transcripts used from CHILDES. Mean ~33 months.

Figure 2: The mean performance of all 40 LSTMs and all 40 CBLs on the production task by child utterance length using either the Beam Search encoder or the "greedy" encoder as estimation procedures. The dotted line represents chance performance.



References:

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J.S., Reilly, J. and Hartung, J., 1994. Developmental and stylistic variation in the composition of early vocabulary. *Journal of child language*, *21*(1), pp. 85-123.

Brinchmann, E.I., Braeken, J. and Lyster, S.A.H., 2019. Is there a direct relation between the development of vocabulary and grammar?. *Developmental science*, *22*(1), e12709.

Frank, M., Braginsky, M., Marchman, V., and Yurovsky, D. 2019. *Variability and Consistency in Early Language Learning: The Wordbank Project.* [Manuscript]. https://langcog.github.io/wordbank-book/

Goldberg, A.E., 2003. Constructions: A new theoretical approach to language. *Trends in cognitive sciences*, *7*(5), pp. 219-224.

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.

Perruchet, P., and Vinter, A. 1998. PARSER: A model for word segmentation. *Journal of Memory and Language, 39,* pp. 246–263.

Pollard, C. and Sag, I.A., 1994. *Head-driven phrase structure grammar*. University of Chicago Press.

Servan-Schreiber, E., & Anderson, J. R. 1990. Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* pp. 592–608.