Supplementary materials: Modeling the influence of language input statistics on children's speech production

Abstract

Keywords: statistical learning, language learning, abstraction, developmental trajectory, age-invariance, CHILDES, children

Supplementary materials: Modeling the influence of language input statistics on children's speech production

Results using original CBL method

Our implementation of the CBL model diverges slightly from the original, so below we present results using McCauley and Christiansen's original (McCauley & Christiansen, 2011) reconstruction task method. In the analyses reported in the main text, we did not attempt to reconstruct utterances that contained previously unseen words. However, McCauley & Christiansen (2011)] handled these cases differently: they built a new chunk for each unknown word. This chunk with the unknown word was then assigned a BTP equal to zero with respect to any other chunk in the utterance it originated from. Consequently, in the analyses reported below, in which we use their method, we no longer provide analyses of the number of utterances containing unknown words.

We analyzed the effect of child age on the model's reconstruction abilities for the child utterances with a mixed-effects model, including age as a fixed effect and a by-child random intercept with random slopes of age.

First, we used the binary (1: reconstructed correctly, 0: not reconstructed correctly) measure from McCauley & Christiansen (2011) and McCauley & Christiansen (2014)]. The model's average percentage of correctly reconstructed utterances across children and age points was similar for the locally and cumulatively sampled speech (local: mean = 61.3 %, range across children = 51.6%–69.6%; cumulative: mean = 59.4%, range across children = 50.8%–68.4%). The number of correctly reconstructed utterances decreased with age, regardless of the sampling methods (local: $b = -0.939, SE = 0.174, p < 0.001$; cumulative: $b = -0.848, SE = 0.138, p < 0.001$; see Figure 1).

Second, we used our corrected, length-and-repetition-controlled reconstruction score.

The model's average reconstruction score across children and age points was similar for the locally and cumulatively sampled speech (local: mean = 0.12, SE = 0.01; cumulative: mean = 0.08, SE = 0.01). As in the main text, we centered age in the model so that we could investigate whether reconstruction was greater than chance level at the average age in our sample (2;6 years). Using both sampling methods, we found a significant intercept (local sampling: $b = 0.130, SE = 0.016, t = 7.911$; cumulative sampling: $b = 0.0789, SE = 0.012, t = 6.426$), and the model's reconstruction score did not change significantly over age (local sampling: $b = 0.029, SE = 0.016, t = 1.854$; cumulative sampling: $b = 0.031, SE = 0.013, t = 2.333$); see Figure 2). These results show that the model performed at above-chance levels, and indicates age-invariance with the corrected reconstruction score.

Importantly, these results are highly similar to those from our implementation of the CBL model in the main text, which does not attempt to reconstruct utterances with previously unseen words. These findings suggest that the CBL is not significantly impacted in its ability to reconstruct children's utterances in the first four years, regardless of the minor algorithmic differences in how new words are treated between the original and current CBL models.

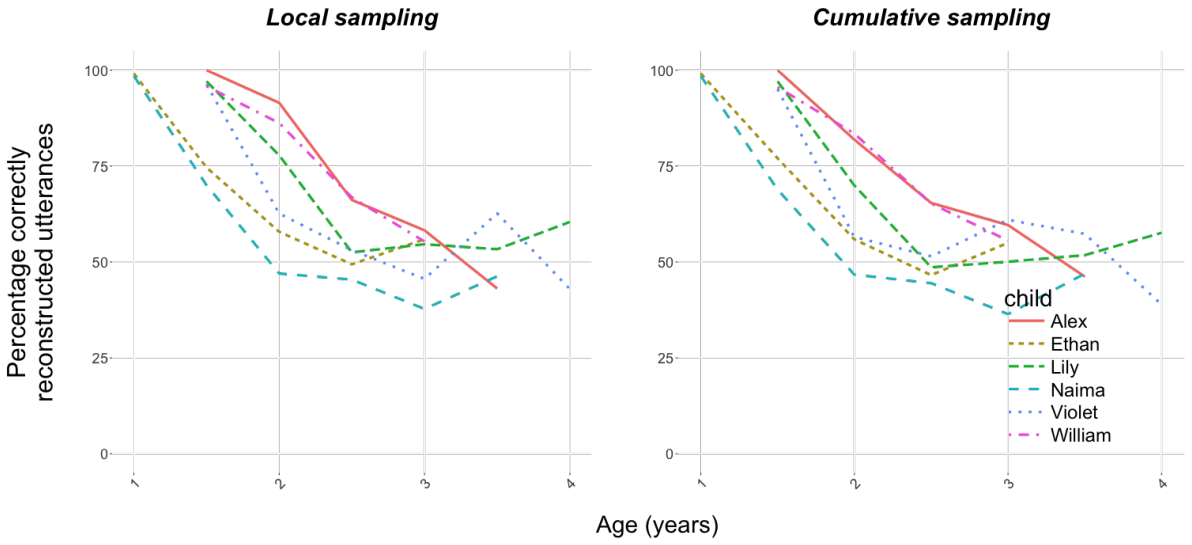

Figure 1. Uncorrected reconstruction scores across the analyzed age range for local (left) and cumulative (right) sampling while using McCauley and Christiansen's method for handling new words.

References

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 1619–1624.

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon*, *9*(3), 419–436.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The CAPPUCCINO model. *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society*, 1619–1624.

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon*, *9*(3), 419–436.

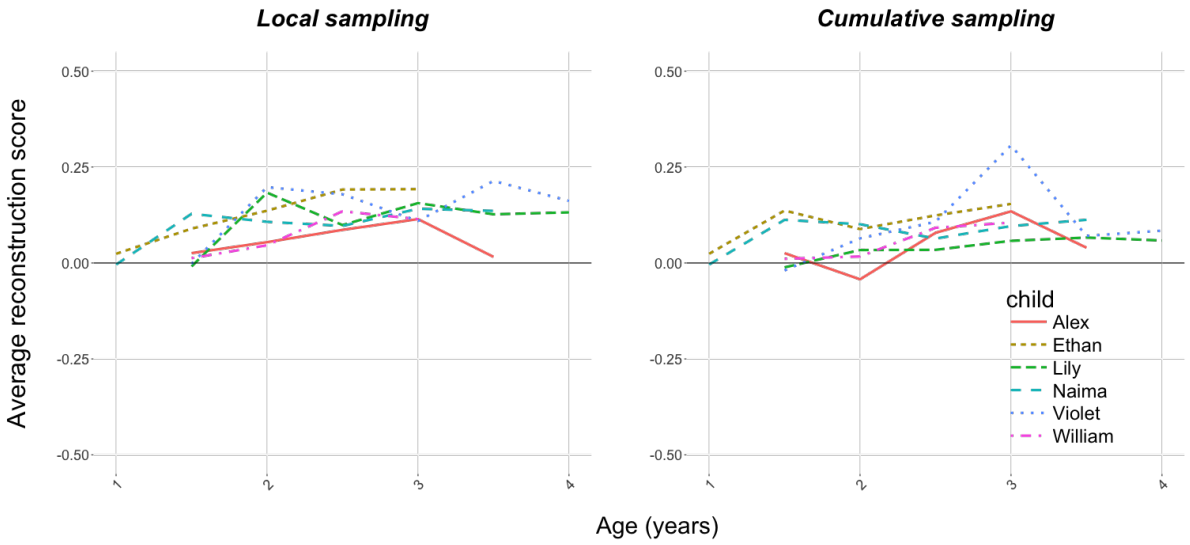

Figure 2. Corrected reconstruction scores across the analyzed age range for local (left) and cumulative (right) sampling while using McCauley and Christiansen's method for handling new words.