1   Modeling the influence of language input statistics on children's speech production

Abstract

We trained a computational model (the Chunk Based Learner; CBL) on a longitudinal corpus of child-caregiver interactions to test whether one proposed statistical learning mechanism—backward transitional probability (BTP)—is able to predict children's speech productions with stable accuracy throughout the first few years of development. We predicted that the model less accurately generates children's speech productions as they grow older because children gradually begin to generate speech using abstracted forms rather than specific "chunks" from their speech environment. To test this idea, we trained the model on both recently encountered and cumulative speech input from a longitudinal child language corpus. We then assessed whether the model could accurately reconstruct children's speech. Controlling for utterance length and the presence of duplicate chunks, we found no evidence that the CBL becomes less accurate in its ability to reconstruct children's speech with age. Our findings suggest that BTP is an age-invariant learning mechanism.

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

Word count: 6056

18   Modeling the influence of language input statistics on children's speech production

19   During the first few years of life children learn the basic building blocks of the

20 language(s) around them. One way they do so is via statistical learning (SL), the process of

21 extracting regularities present in the language environment. Over the past few decades, SL

22 has become a major topic in the field of first language acquisition, ranging in application

23 from speech segmentation (Jusczyk & Aslin, 1995; Saffran, Aslin, & Newport, 1996) and

24 phonotactic learning (Chambers, Onishi, & Fisher, 2003) to producing irregulars (Arnon &

25 Clark, 2011), discovering multi-word structures (Bannard, Lieven, & Tomasello, 2009; Chang,

26 Lieven, & Tomasello, 2006; Frost, Monaghan, & Christiansen, 2019), and much more (see

27 Saffran and Kirkham (2018) for a recent review). By its nature, work in this domain is

28 heavily concerned with at least two major topics: (1) the information available in children's

29 language environments (the "input") from which they can pick up on patterns, and (2) the

30 precise mechanisms by which children convert these "raw" environmental statistics into

31 internalized knowledge about language. A third issue is whether and how children's SL

32 behavior changes as they develop (Shufaniya & Arnon, 2018). The current paper taps into

33 each of these three issues: we train a computational model on a longitudinal corpus of

34 child-caregiver interactions to test whether one proposed SL mechanism—backward

35 transitional probability (BTP)—is able to predict children's speech productions with stable

36 accuracy as they get older.

37 **SL over development**

38   The ability to detect and store patterns in the environment begins in infancy (e.g., S.

39 P. Johnson et al., 2009; Kidd, Junge, Spokes, Morrison, & Cutler, 2018; Saffran et al., 1996;

40 Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), continues into adulthood (e.g.,

41 Conway, Bauernschmidt, Huang, & Pisoni, 2010; Frost & Monaghan, 2016; Saffran, Johnson,

42 Aslin, & Newport, 1999), and crosses a range of modalities (Conway & Christiansen, 2005;

43 Emberson, Conway, & Christiansen, 2011; Monroy, Gerson, & Hunnius, 2017). However, it is

44 still a matter of debate whether SL is an age-invariant skill or not (Arciuli & Simpson, 2011;

45 Raviv & Arnon, 2018; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Shufaniya &

46 Arnon, 2018). Recent work that investigates SL abilities in 5–12-year-old children suggests

47 that, while both visual and auditory SL improve with age for non-linguistic stimuli,

48 performance stays the same across childhood for linguistic stimuli (Raviv & Arnon, 2018;

49 Shufaniya & Arnon, 2018). From this finding, the authors conclude that SL for language

50 might be age-invariant. On the other hand, infant SL abilities do appear to shift within the

51 first year, both for linguistic (Kidd et al., 2018) and non-linguistic (S. P. Johnson et al.,

52 2009) stimuli. For example, while 11-month-olds can detect and generalize over regularities

53 in a sequence, 8-month-olds are only capable of detecting the regularities, and neither group

54 succeeds yet at learning visual non-adjacent dependencies (S. P. Johnson et al. (2009); see

55 also Bulf, Johnson, and Valenza (2011) and Slone and Johnson (2015)).

56   These changes in SL **ability** during infancy and early childhood may relate to changes

57 in other fundamental cognitive skills. For example, SL-relevant brain regions, such as the

58 pre-frontal cortex, continue maturing through childhood (Casey, Giedd, & Thomas, 2000;

59 Diamond, 2002; Rodríguez-Fornells, Cunillera, Mestres-Missé, & Diego-Balaguer, 2009;

60 Uylings, 2006), which may change how children attend to the linguistic information around

61 them as they get older. Similarly, infants' long-term memory continuously improves between

62 ages 0;2 and 1;6 (Bauer, 2005; Wojcik, 2013). Therefore, the manner in which they store

63 linguistic regularities in long-term memory may also shift during this period. Relatedly,

64 working memory and speed of processing change continuously throughout early childhood

65 (Gathercole, Pickering, Ambridge, & Wearing, 2004; Kail, 1991), implying that there could

66 be a developmental change in the rate and scale at which children can process chunks of

67 information from the unfolding speech signal.

68   Continued exposure to linguistic input itself can also be an impetus for change in SL

69 **ability**—a view supported by multiple, theoretically distinct, approaches to early syntactic

70 learning. For example, Yang (2016) proposes that children gather detailed, exemplar-based

71 statistical evidence until it is more cognitively efficient for them to make a categorical

72 abstract generalization. He proposes that, at that point, the learner instantiates a rule to

73 account for patterns in the data. Usage-based theories of early language development instead

74 propose that children first learn small concrete linguistic sequences from their input that are

75 made up of specific words or word combinations (e.g., "dog" and "I wanna"; or multi-word

76 combinations, "where's the . . . "; Tomasello (2008)). Then, over time, children are proposed

77 to form abstract schemas centered on lexical items (see also Bannard et al. (2009) and

78 Chang et al. (2006)). This representational shift, from probabilistic and lexical to abstract

79 and syntactic, is used to account for how children can eventually create utterances that they

80 have never heard before. Crucially, the representational shift implies a change in the way

81 children apply the original SL mechanism(s) to incoming linguistic information (see also

82 Lany & Gómez, 2008).

83 Change in SL **ability** following further linguistic experience is also predicted in models

84 that do not assume abstraction. In chunk-based models of language learning (Arnon,

85 McCauley, & Christiansen, 2017; Christiansen & Arnon, 2017; Christiansen & Chater, 2016;

86 Misyak, Goldstein, & Christiansen, 2012; StClair, Monaghan, & Christiansen, 2010), children

87 use statistical dependencies in the language input (e.g., between words or syllables) to store

88 chunks of co-occurring forms. Dependencies between the chunks themselves can also be

89 tracked **with** continued exposure and chunk storage **(see, e.g., Jost & Christiansen,**

90 **2016)**. **In this case, the development of a detailed chunk inventory can**

91 **gradually change overt SL performance.** Fundamentally, however, **this apparent**

92 **change in SL still comes through the use of the original** underlying mechanisms

93 (Misyak et al., 2012)**;** there is no **qualitative change in how the system processes**

94 **data**, and the mechanisms for processing, storing, and deploying information stay the same.

95 We investigated the possibility of developmental change in SL using computational

96 modeling, which enables us to define and test the goodness-of-fit for any given learning

97 mechanism on a dataset of natural speech. We chose to use a longitudinal child language

98 dataset, in which the same children were tracked across the developmental period of interest

99 for early speech production (1;0–4;0). By choosing data in this age range, we could test

100 whether use of a learning mechanism changed for each child across the studied developmental

101 time points. We tested for developmental change in the use of a single proposed statistical

102 learning mechanism: backward transitional probability (McCauley & Christiansen, 2011;

103 Onnis & Thiessen, 2013; Pelucchi, Hay, & Saffran, 2009; Perruchet & Desaulty, 2008).

BTP and the Chunk-Based Learner

105 Our model is based on McCauley and Christiansen's (2011, 2014a) Chunk-Based

106 Learner (CBL) model, which uses one measure—backward transitional probability **(BTP;**

107 **Perruchet & Desaulty, 2008)**—to detect statistical dependencies in the speech stream.

108 **Backward transitional probability is one of multiple approaches for dividing**

109 **streams of continuous speech into meaningful units; other approaches include,**

110 **for example, forward transitional probability and memory-based chunking**

111 **(Aslin, Saffran, & Newport, 1998; Cleeremans & Elman, 1993; French,**

112 **Addyman, & Mareschal, 2011; see Frost & Monaghan, in press for an review;**

113 **Mareschal & French, 2017; Onnis & Thiessen, 2013; Pelucchi et al., 2009;**

114 **Perruchet & Desaulty, 2008; Perruchet & Vinter, 1998; Saffran et al., 1996).**

115 BTP for a given pair of words is defined as the occurrence probability of the previous word

116 (w_{-1}) given the current word (w_0). It can be estimated for each word in a sentence in order

117 to reveal peaks and dips in transitional likelihood, which reflect places where words are likely

118 (peaks) or unlikely (dips) to co-occur.

119 The CBL model divides utterances into chunks, splitting the utterances whenever the

did you | look at | the doggy
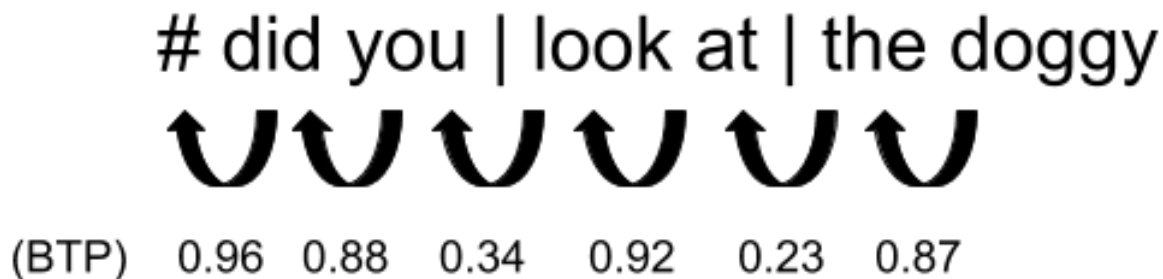
(BTP) 0.96 0.88 0.34 0.92 0.23 0.87

Figure 1. Example of a sentence with BTP between consecutive words. Chunks are split at points of low BTP (indicated by the vertical lines). "#" denotes a start-of-sentence marker.

BTP between two words drops below the running average BTP. In the example in Figure 1, the CBL might decide to split the sentence ("did you look at the doggy") into three chunks "did you", "look at", and "the doggy", and store all three in its memory. As it sees more sentences, it would continue to add new chunks and track how often they co-occurred.

The CBL was developed to model children's early speech production and comprehension without appealing to abstract grammatical categories. Specifically, it was designed as an implementation of the hypothesis that children detect and store multi-word chunks using BTP, and also use the stored chunks to parse speech and produce new utterances (see also Arnon and Snider (2010) and Bannard and Matthews (2008)). The model's ability to simulate learning can be measured by first training it on what children hear and then having the model reproduce what children say from the chunks that it learned.

We chose to build on the CBL model because it has successfully accounted for production data in multiple corpora, including **child language** datasets. For example: (a) it parsed text better than a shallow parser in three different languages (English, German and French) when using individual words rather than word classes, (b) it was able to recreate up to 60% of child utterance productions in 13 different languages, and (c) it closely replicated data from an artificial grammar learning study (McCauley & Christiansen, 2011; Saffran,

137 2002). The model has also been able to replicate experimental data on children's multi-word

138 utterance repetitions (Bannard & Matthews, 2008), over-regularization of irregular plural

139 nouns (Arnon & Clark, 2011), and L2-learner speech (see also McCauley & Christiansen,

140 2014b, 2017). In sum, the CBL model appears to robustly predict the word-chunk patterns

141 in children's speech when given information about what they hear in their input. **We**

142 **extend this work by testing how the model performs with longitudinal data; it**

143 **is not yet known how well it functions as a predictor of what children can say**

144 **as they become more linguistically sophisticated.**

Testing for change with age

146   Following McCauley and colleagues (2011, 2014a, 2019) we tested the CBL model's

147 ability to learn language by checking how well it can reconstruct children's utterances from

148 the chunks discovered in their caregivers' speech. As we are interested in developmental

149 change over the first three years of speech production, we analyzed the model's

150 reconstruction ability with two measures:

151 • "Uncorrected": The binary (success/fail) reconstruction score originally used by

152   McCauley and colleagues (2011, 2014a, 2019).

153 • "Corrected": A length-and-repetition-controlled reconstruction score that accounts for

154   the fact that longer utterances have more opportunities for error reconstruction, and

155   for the fact that some child utterances contain repetitions of **chunks**(s), making

156   multiple reconstructions match the original utterance.

157   If BTP is an age-invariant mechanism, it should apply equally well to shorter

158 utterances and longer utterances; the latter of which are more often produced as children get

159 older. We therefore tested for age invariance both with the original binary ("uncorrected")

160 reconstruction score and a new ("corrected") score we proposed to account for utterance

¹⁶¹ length and word repetitions. If we find age-invariance, even while controlling for utterance

¹⁶² length and word repetitions, it would strongly suggest that the mechanism is stable over the

¹⁶³ first three years of speech production and not simply influenced by other factors, e.g.,

¹⁶⁴ utterance length. If not, it would suggest that use of the mechanism, in fact, changes with

¹⁶⁵ age (Bannard et al., 2009; Tomasello, 2005; Yang, 2016).

Predictions

¹⁶⁷ **With these** previous findings as a starting point, we investigated whether the CBL

¹⁶⁸ could account for child speech production with equal precision over the first four years of life.

¹⁶⁹ **Taking for granted that children *eventually* develop abstract representations**

¹⁷⁰ **(Tomasello, 2008; as in, e.g., Yang, 2016), w**e predicted that:

- ¹⁷¹ The CBL would less accurately generate children's speech productions as they grew

  ¹⁷² older; given the assumption that children gradually learn to abstract over the specific

  ¹⁷³ "chunks" they encounter (Bannard et al., 2009; Tomasello, 2005; Yang, 2016) and,

  ¹⁷⁴ therefore, their speech should less often directly mirror their linguistic input at later

  ¹⁷⁵ ages. This finding would indicate that the immediate influence of children's language

  ¹⁷⁶ input statistics on their speech production decreases across development.

- ¹⁷⁷ Children will be more likely to use words that are not documented in the caregiver

  ¹⁷⁸ speech as they get older. These words could originate from other sources, such as peer

  ¹⁷⁹ speech or non-recorded caregiver speech (Hoff, 2010; Hoff-Ginsberg & Krueger, 1991;

  ¹⁸⁰ Mannle, Barton, & Tomasello, 1992; B. C. Roy, Frank, & Roy, 2009).

- ¹⁸¹ Younger children's utterances would be reconstructed well on the basis of recently

  ¹⁸² heard speech alone, whereas older children's utterances would be best constructed

  ¹⁸³ when considering a longer period of their historical input. Our reasoning was that

  ¹⁸⁴ older children's increased memory capacity (Bauer, 2005; Gathercole et al., 2004;

  ¹⁸⁵ Wojcik, 2013) allows them to draw on older input more easily in producing speech. If

so, the findings would suggest that memory plays a critical role in the use of the same learning mechanism with age.

In sum, we expected to find that the CBL's ability to reconstruct children's speech decreases in-line with a concomitant increase in children's linguistic sophistication; an effect driven by children's use of more abstracted representations, words from other speech sources, and their increased ability to use historical input.

Methods

Model

The CBL model (McCauley & Christiansen, 2011) is an incremental, online computational model of language acquisition, that explores the possibility that infants and children parse their input into (multi-word) chunks during the process of acquiring language.

The model takes transcribed speech as input and divides the transcribed utterances into multi-word chunks. Each utterance begins with a start cue (denoted "#"). The exact placement of a chunk boundary within an utterance is determined by two factors: (1) the backward transitional probability (BTP) between consecutive words in the utterance, and (2) the inventory of already-discovered chunks. All newly discovered chunks are saved into the inventory, alongside the BTPs associated with each chunk. **The model tracks and stores:** the discovered chunks, the BTPs between words, and the BTPs between discovered chunks. For example, the model might parse the utterances "I see the doggy" and "did you look at the doggy?" into five different chunks, namely "I", "see", "the doggy", "did you", and "look at" based on the BTPs between these words compared to the average BTP found in the corpus so far.

Child utterance reconstruction task

Once the model has been trained on adult utterances, and thereby has discovered chunks in the adults' speech, we can test whether it closely matches the linguistic structures produced by the children in the same caregiver-child corpus. Following McCauley and Christiansen (2011), we use a child utterance reconstruction task to test whether the chunk statistics present in the adults' utterances are also present in the child's utterances. The model reconstructs the child utterances from the chunks and their related BTPs from the adult's utterances at the same age point. This reconstruction **process, which is slightly different from McCauley and Christiansen's (2011) process,** is done in two steps (see Figure 2). First, a child utterance is converted into an unordered bag-of-chunks containing chunks discovered in the adults' speech, in line with the bag-of-words approach proposed in Chang, Lieven, and Tomasello (2008). Whenever the model encounters a word in the child utterance that is not present in the adult-based chunk inventory, it stops processing that utterance.[1] For instance, in the toy example in Figure 2, the child utterance "look at the doggy" would be **broken down into a** set of **known** chunks which were discovered in the adults' speech **(e.g., "look at" and "the doggy", as in the step 2 speech bubble). I**f the utterance were "look at the doggy there", and the model had **not already added a** chunk for the word "there" **during training, then the word is unknown to the model and the utterance cannot be reconstructed; therefore the utterance would be rejected from further processing. However, in the case that the utterance *can* be broken down into known chunks, the model then** tries to reconstruct **the utterance by shuffling the chunks detected and reordering them based on their known transitional probailities:** the model begins with the utterance start cue and then follows that initial cue **with the chunk that has the**

---

[1]McCauley and Christiansen (2011) handle these cases differently**. Our CBL implementation is identical to theirs up to this point. Therefore we also** provide sentence reconstruction scores using their original method in the Supplementary Materials.

232 **highest transitional probability following the start cue, which is followed by the**

233 **remaining chunk that has the highest transitional probability following the**

234 **previous chunk, and again and again, until the set of chunks for that utterance**

235 **is exhausted..** So, the set of chunks "look at" and "the doggy" would be ordered depending

236 on which chunk had the highest BTP with respect to the utterance start cue ("look at"),

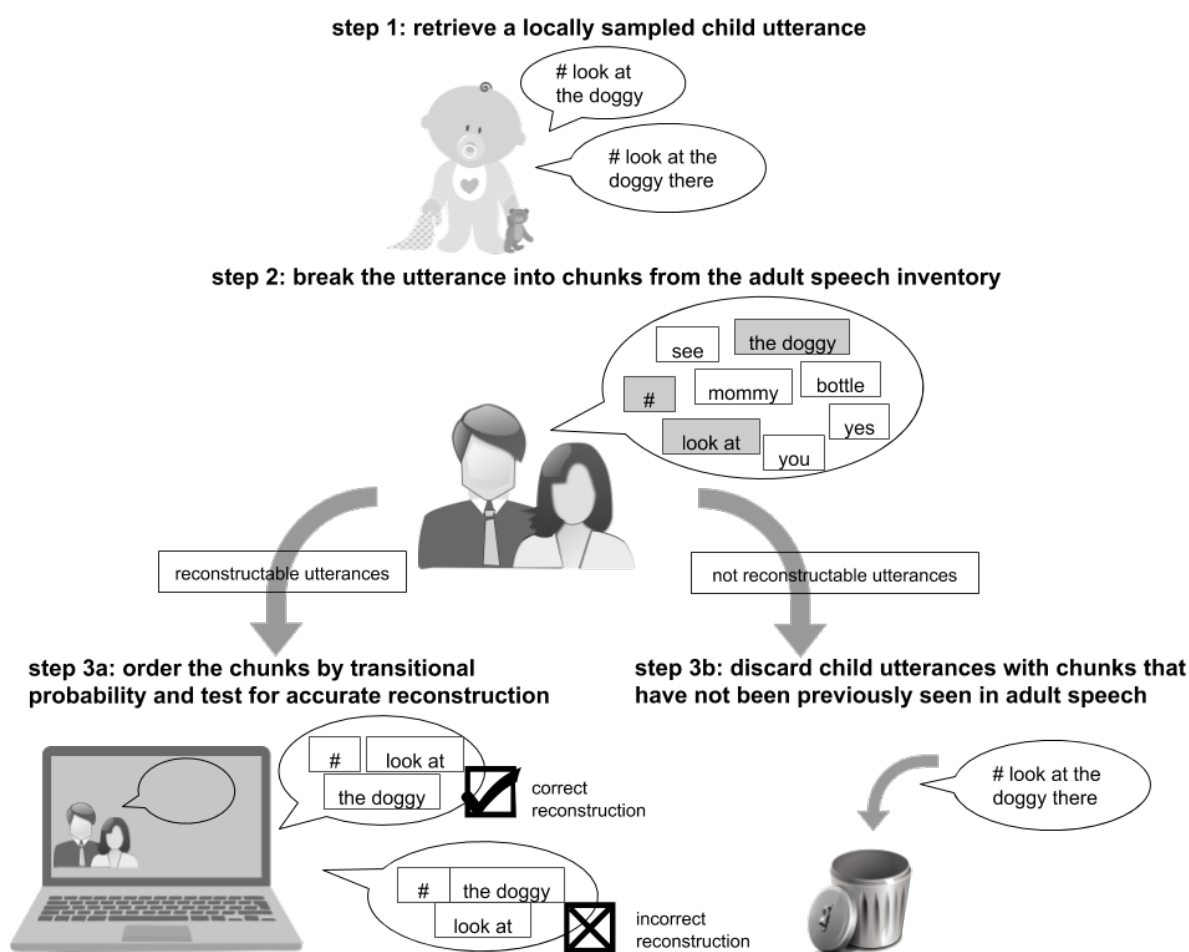237 followed by the chunk with the highest BTP with respect to "look at" ("the doggy").

Figure 2. Example of a reconstruction attempt of two child utterances by a toy model. The model is able to reconstruct the first utterance, but it cannot do so with the second utterance, which contains a word ("there") that it has not previously seen.

Materials and Procedure

As input to the model we used transcripts of 1–2-hour recordings of at-home interaction between six North American children and their caregivers who were recorded approximately every two weeks between ages 1;0 and 4;0 (the Providence corpus; Demuth, Culbertson, and Alter (2006)). We pre-processed the transcripts, which were formatted using CHAT conventions (MacWhinney, 2000), such that the input to the model only contained plain text orthographic transcriptions of what was said.[2] We split the transcripts into two separate files, one with all the caregivers' utterances and one with all the child's utterances. Our pre-processing also added a "#" prefix to the start of each utterance.

The transcripts were sampled at approximate 6-month intervals between ages 1;0 and 4;0. We used two different sampling methods: a local data sampling method and a cumulative data sampling method. With the local data sampling method we selected data within a two-month interval around each age point. For example, for age point 1;6 we selected transcripts in which the child was between 1;5.0 and 1;6.31 years of age. This method led to ~800–4000 caregiver utterances at each age point. By design, the local sampling method focuses the model's training solely on *recent* linguistic input so that, when it tries to reconstruct children's utterances, the result is a test of how closely their current speech environment can account for what they say. **We sample *around* target age points and not *up-until* target age points because, while the Providence corpus is relatively densely sampled, recording sessions weren't frequent enough to guarantee a representative picture of each child's input in the month preceding each of the target age points. For this reason, we decided that training the model on input proximal to the tested age was a better method for getting a**

---

[2]All punctuation marks, grammatical notes, omitted word annotations, shortenings, and assimilations were removed from the utterances, such that only the text representing the spoken words of the utterance remained.

broad, but age-specific model of adult speech for each child at each age point.

In contrast, the cumulative sampling method focuses the model's training on all previously heard linguistic input so that, when it tries to reconstruct children's utterances, the result is a test of how closely their current *and* previous speech environments can account for what they say. For the cumulative sample we selected data for each age point by taking all the available transcripts up to that age point. For example, for age 1;6 we selected all transcripts in which the child was 1;6 or younger. This method led to ~800–60,000 caregiver utterances across the different age points, with the number of caregiver utterances increasing (i.e., accumulating) with child age. As a consequence, the cumulative sample always contained more caregiver utterances than the local sample, except at age 1;0, the first sampled age point.

While we used two different sampling methods for training the model on adult data, all child utterances used for the reconstruction task were retrieved using the local sampling method for that particular age point. In other words, we only reconstructed the child utterances local to each tested age, regardless of the training strategy.

Analysis

We modeled two primary scores related to utterance reconstruction: the uncorrected (binary: success/fail) reconstruction score used by McCauley and colleagues (2011, 2014a, 2019) and the corrected reconstruction score we introduce in the current paper. The uncorrected reconstruction score (1: success, 0: fail) was computed for all child utterances that could be decomposed into previously seen chunks (see step 3a in Figure 2). The corrected reconstruction score (defined below) was computed for the same set of utterances. We additionally included a third analysis: the likelihood that an utterance contains previously unseen words which, by our version of the CBL, cannot be reconstructed (see step

285 3b in Figure Figure 2).

286   We used mixed-effects regression to analyze the effect of child age on both of the

287 reconstruction scores and whether utterances contained previously unseen words. All

288 mixed-effects models included child age as a fixed effect and by-child random intercepts with

289 random slopes of child age. The mixed-effects model of utterances with previously unseen

290 words also included the number of words in the utterance, as explained below. By default,

291 child age was modeled in years (1–4) so that the intercept reflects a developmental trajectory

292 beginning at age 0. **However, for the model of corrected reconstruction accuracy,**

293 **we wanted to know if the CBL's accuracy score would exceed chance**

294 **performance on average, i.e., at the average age in our longitudinal dataset**

295 **(age 2;6). To test this, we centered age on zero in the statistical (ages 1;0, 1;6,**

296 **2;0, 2;6, 3;0, 3;6, and 4;0 are re-coded as -1;5, -1, -0.5, 0, 0.5, 1, and 1.5) such**

297 **that the default model output from the lme4 statistical package (Bates,**

298 **Mächler, Bolker, & Walker, 2015) would reflect the estimated difference from**

299 **chance at the middle point of our age range (i.e., age 2;6).**

300   All analyses were conducted using the `lme4` package (Bates et al., 2015) and all figures

301 were generated with the `ggplot2` package in `R` (R Core Team, 2014; Wickham, 2009). All

302 code used to create the model and analyze its output is available at

303 https://osf.io/ca8ts/?view_only=daaa1bcc71654842b0d70efe785a26b9. Before turning to the

304 main results we briefly describe the corrected reconstruction score and the analysis of

305 previously unseen words in more detail.

306 **Corrected reconstruction accuracy**

307   The corrected, length-and-repetition-controlled reconstruction score is a function of

308 three factors: (a) whether the model successfully reconstructed the child utterance or not,

309 (b) the number of chunks used to reconstruct the utterance, and (c) the number of duplicate

310 chunks involved in the reconstruction. By taking the number of chunks into account, this

311 reconstruction score compensates for the fact that successful reconstruction is less likely for

312 longer utterances. When an utterance contains duplicate chunks, the exact ordering of those

313 duplicate chunks does not influence the correctness of the reconstruction. For example, if the

314 utterance "I wanna, I wanna" is decomposed into the two chunks "I wanna" and "I wanna",

315 it does not matter which of the two "I wanna" chunks is placed first when calculating the

316 reconstruction accuracy of the utterance. Thus, utterances containing duplicate chunks are

317 more likely to be reconstructed by chance alone than utterances with the same number of

318 chunks with no duplicates. **Note that here we detecting duplicate *chunks* in the**

319 **utterance rather than duplicate *words*. At this post-training stage, the model**

320 **is only able to parse the utterance into chunks; that is the relevant unit over**

321 **which duplication may affect reconstruction accuracy.**

322 An utterance that is decomposed into N unique chunks can be reconstructed in $N!$

323 different orders. Hence, the probability of obtaining the correct order of N unique chunks

324 merely by chance equals $1/N!$. When we take into account that chunks can be repeated

325 within an utterance, chance level equals $(n_1!n_2!\ldots n_k!)/N!$ where N is the total number of

326 chunks in the utterance, and $n_1, \ldots, n_k$ are the number of times a chunk is repeated for each

327 of the k unique chunks found in the utterance.

328 When probability of reconstruction was lower, we scored a correctly reconstructed

329 utterance higher. We assigned a score of $-\log(chance)$ for each correct reconstruction and

330 $\log(1 - chance)$ for each incorrectly reconstructed utterance. In layman's terms, this means

331 that successfully reconstructed utterances were scored positively, but were weighed relative

332 to the number of chunks and the number of repetitions they had, such that reconstructions

333 of long utterances were given higher scores than reconstructions of short utterances. Along

334 the same lines, incorrectly reconstructed utterances were scored negatively and were also

weighed relative to the number of chunks they had, such that incorrect reconstructions of long utterances were given higher (i.e., less negative) scores than incorrect reconstructions of short utterances.

To illustrate the corrected scoring method, let's compare two three-chunk utterances, one of which contains a duplicate chunk: "I wanna I wanna see" (chunks: "I wanna", "I wanna", "see") and "I wanna see that" (chunks: "I wanna", "see", "that"). For the first utterance, chance level equals $(2! \times 1!)/(3!)$: The numerator is determined by the number of times each unique chunk is used, so because "I wanna" occurs two times and "see" occurs once, that is $2! \times 1!$. The denominator is determined by the factorial of total number of chunks (here: $3! = 3 \times 2 \times 1$). The resulting chance level is then $2/6$. For the second utterance, chance level equals $(1! \times 1! \times 1!)/(3!)$: The numerator is equal to $1! \times 1! \times 1!$ here because all chunks occur only once in the utterance. The denominator is the same as for the first utterance as the total number of chunks in the utterance is the same. Here, the resulting chance level is $1/6$. If the utterances are reconstructed correctly, the score is computed by $-\log(chance)$. So, the first utterance would get a positive score of $-\log(chance) = -\log(2/6) \approx 1.098$ and the second utterance would get a higher positive score of $-\log(chance) = -\log(1/6) \approx 1.791$ for increased reconstruction difficulty. If the utterances are reconstructed incorrectly, the score is computed by $\log(1 - chance)$. Thus, the first utterance would get a negative score of $\log(1 - chance) = \log(1 - (2/6)) \approx -0.405$ and the second utterance would get a less negative score of $\log(1 - chance) = \log(1 - (1/6)) \approx -0.182$.

Previously unseen words

Our third analysis focused on whether or not each child utterance contained words that were not previously seen, that is, not present in the trained adult-speech chunk inventory. Each utterance was coded for unseen words with a binary value (1: at least one unseen word,

360   0: no unseen words). **Longer utterances are more likely to contain unseen words:**

361   **an utterance with N words has a probability of $1-((1-p)^N)$ of containing at least**

362   **one unseen word. We are interested in the likelihood of utterance with unseen**

363   **words beyond these effects of sentence length. We therefore include utterance**

364   **length as a control predictor in the regression.**

365   **Results**

366   **Uncorrected reconstruction accuracy**

367   The uncorrected score of accurate utterance reconstruction (McCauley & Christiansen,

368   2011, 2014a) showed that model's average percentage of correctly reconstructed utterances

369   across children and age points was similar for the locally and cumulatively sampled speech

370   (local: mean = 65.4%, range across children = 59.9%–70.3%; cumulative: mean = 59.9%,

371   range across children = 53.1%–68.2%). This is similar to, or slightly higher than, results

372   reported by McCauley and Christiansen (2011) who found an average percentage of correctly

373   reconstructed utterances of 59.8% over 13 typologically different languages with a mean age

374   range of 1;8–3;6 years. Additionally, McCauley and Christiansen (2019) reported an average

375   reconstruction percentage of 55.3% for 160 single-child corpora of 29 typologically different

376   languages, including a performance of 58.5% for 43 English single-child corpora with a mean

377   age range of 1;11–3;10.

378   In our statistical models of the uncorrected reconstruction accuracy[3], we first analyzed

379   the CBL model's performance when it was trained on locally sampled caregiver speech. The

380   number of correctly reconstructed utterances decreased with age

381   $(b = -0.805, SE = 0.180, p < 0.001)$; over time the BTP statistics present in the caregivers'

382   speech were less reflected in the child's own speech (Figure 3, left panel).

---

[3]**accuracy ~ age + (age|child), family = binomial(link = "logit").**

We then tested the model's performance when it was trained with a cumulative sample of caregiver speech, rather than just a local sample. As before, the number of correctly reconstructed utterances decreased with child age ($b = -0.821, SE = 0.146, p < 0.001$; Figure 3, right panel). These results indicate age-variance for the SL mechanism; its utility for modeling children's utterances changes with age.

Importantly, however, the length of the child utterances varied quite a lot (range = 1–44 words long; mean = 2.8, median = 2), and some of them contained repetitions of chunks (e.g., "I wanna, I wanna"), both of which influence the baseline likelihood of accurate reconstruction. Utterances from older children tended to contain more words than utterances from younger children (Figure 4, left panel). As a consequence, on average, utterances from older children are systematically less likely to be correctly reconstructed by chance, contributing to the decrease in the CBL's overall performance with age. Additionally, the percentage of child utterances that contained duplicate chunks decreased over time (Figure 4, right panel). Utterances with duplicate chunks had a higher baseline probability of being accurately reconstructed by the model. So again, on average, utterances from older children were systematically more difficult, contributing to the age-related decrease in uncorrected reconstruction scores.

Corrected reconstruction accuracy

Next, we used our corrected reconstruction score to assess the model's reconstruction accuracy while controlling for utterance length and the use of duplicate chunks. The score weighs whether each utterance was accurately reconstructed against its chance level of reconstruction, depending on the total number of chunks and number of duplicate chunks it contains. The model's average reconstruction score across children and age points was similar for the locally and cumulatively sampled speech (local: mean = 0.10, SE = 0.01; cumulative: mean = 0.06, SE = 0.01). Note again that one aim of this analysis was to test
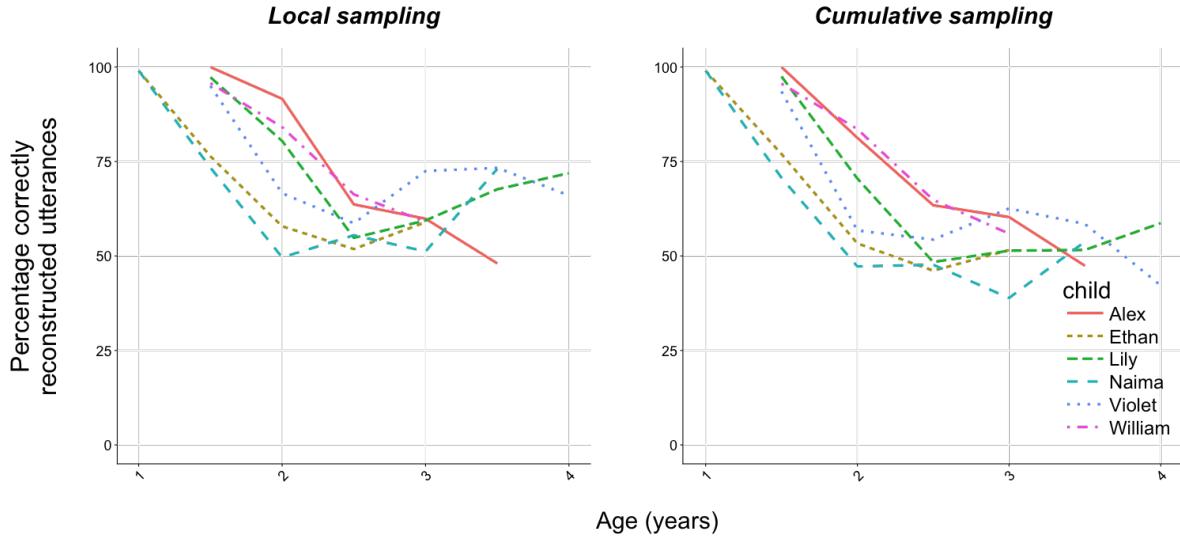
Figure 3. Percentage of correctly reconstructed utterances across the age range, using local (left) and cumulative (right) sampling.

⁴⁰⁸ whether the corrected reconstruction score was above chance—here represented by a score of

⁴⁰⁹ zero—so in the statistical models we centered child age on zero so that the estimation would

⁴¹⁰ reflect the difference from zero for the average age in our sample (2;6).[4]

⁴¹¹   Again, we first analyzed the model's performance when it was trained on locally

⁴¹² sampled caregiver speech. We found a significant positive intercept

⁴¹³ $(b = 0.11, SE = 0.02, t = 5.064)$ and no significant change across age

⁴¹⁴ $(b = 0.030, SE = 0.018, t = 1.681)$; the BTP statistics from the caregivers' speech were

⁴¹⁵ consistently reflected in the child's own speech (Figure 5, left panel).

⁴¹⁶   As before, we created a parallel set of analyses to test the model's performance when it

⁴¹⁷ was trained with a cumulative sample of caregiver speech. We again found a significant

⁴¹⁸ positive intercept $(b = 0.06, SE = 0.010, t = 6.238)$ and that accuracy did not change

⁴¹⁹ significantly across age $(b = 0.02, SE = 0.013, t = 1.590$; Figure 5, right panel).

⁴²⁰   In sum, contrary to the uncorrected reconstruction accuracy analysis, these corrected
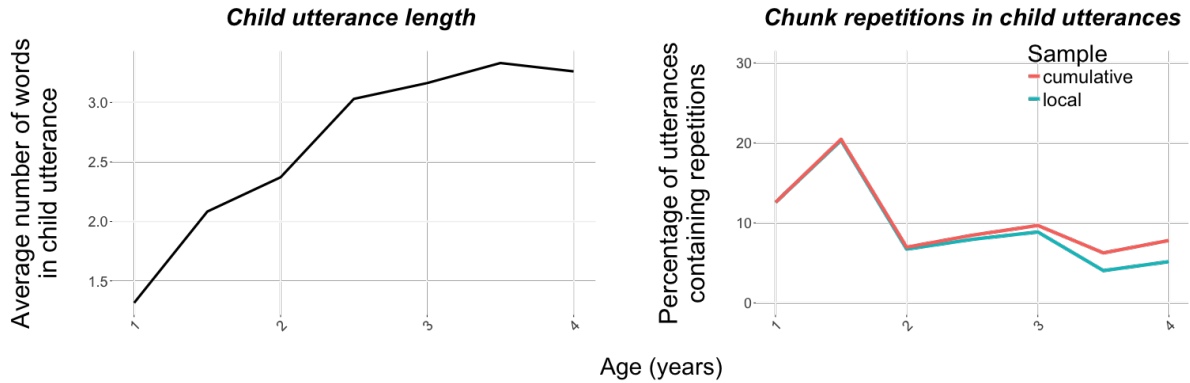
----

[4]**accuracy ~ age + (age|child).**

Figure 4. Children's utterances increased in length (number of words) with age (left) while simultaneously decreasing in the number of duplicate chunks used (right).

reconstruction score results indicate age-invariance for the SL mechanism. In addition, the model performed significantly above chance level in both the local and cumulative sampling contexts.

Children's use of unseen words

Finally, we modeled whether an utterance had a previously unseen word or not (1: at least one unseen word, 0: no unseen words).[5] Using local sampling the number of child utterances that contained one or more previously unseen words increased with age $(b = 0.364, SE = 0.093, p < 0.001$; Figure 6, left panel). Unsurprisingly, longer utterances were also significantly more likely to have previously unseen words in them $(b = 0.170, SE = 0.005, p < 0.001)$. By taking a longer history of linguistic input into account (i.e., by using cumulative sampling), we expected to see a smaller increase in previously unseen words with age. **That is, an unseen chunk in the local sampling could become a seen chunk in the cumulative sampling.** We indeed found that utterances containing previously unseen words became *less likely* with age

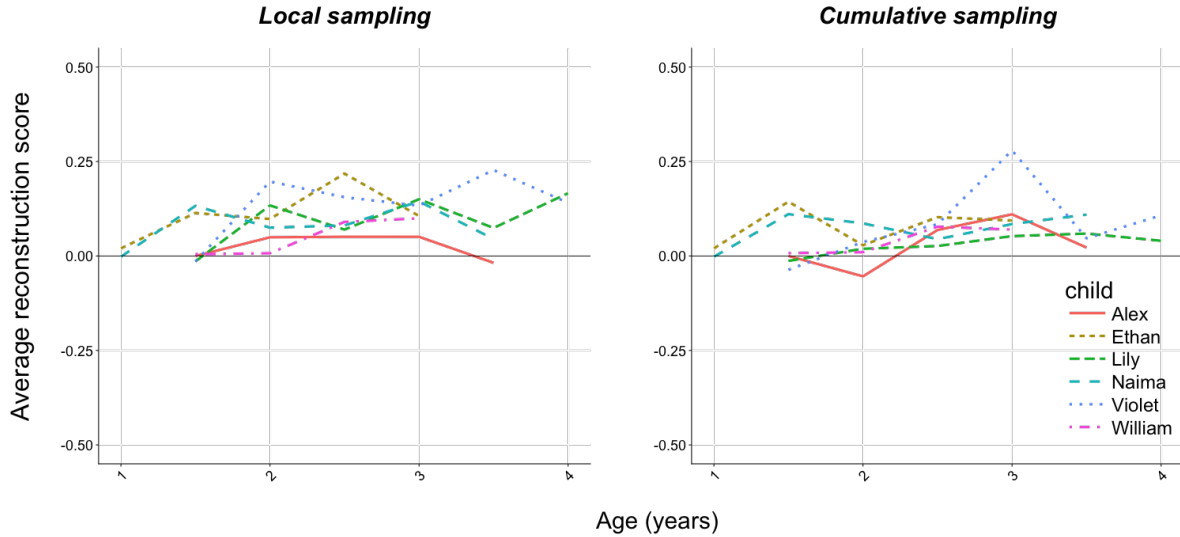[5]**has_unseen_words ~ age + (age|child) + num_words_in_utt, family = binomial(link = "logit")**

Figure 5. Corrected reconstruction scores across the age range, using local (left) and cumulative (right) sampling.

$(b = -0.139, SE = 0.068, p < 0.05$; Figure 6, right panel). Also, as before, previously unseen words were more likely to occur in longer utterances $(b = 0.105, SE = 0.006, p < 0.001)$.

Discussion

Our primary research question (as raised by, e.g., Arciuli & Simpson, 2011; Raviv & Arnon, 2018; Saffran et al., 1997; Shufaniya & Arnon, 2018) was whether the CBL **would change in its ability** to predict children's speech productions throughout development. We tested the model using both the original measure of accuracy as well as a new measure that takes into account utterance length and duplicate chunks in the utterance, which can make accurate reconstruction less likely (length) or more likely (duplicates). Using this corrected measure, we found that there was no significant change in the use of BTP with age. Notably, the CBL was able to construct utterances at above-chance levels despite these changes with age. Overall, **and against our predictions in the Introduction, the current** findings support the view that BTP is an age-invariant learning mechanism for
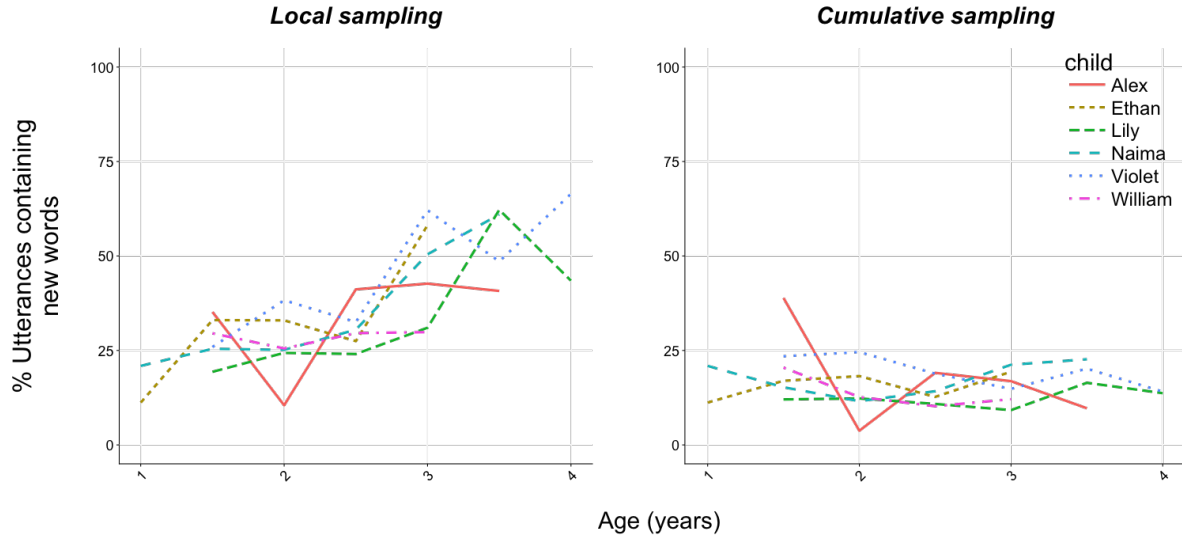
Figure 6. Number of utterances with previously unseen words across the age range, using local (left) and cumulative (right) sampling.

speech production. In fact, the positive, but non-significant coefficients for the effect of age on corrected reconstruction accuracy indicate that, the CBL is, at least, not getting worse at reconstructing children's utterances with age. **Also, the divergence in findings between the corrected and uncorrected accuracy scores illustrates how effects of length and chunk/word duplication can critically shift baseline performance during reconstruction; these features of natural speech should be controlled for in future work.**

Different words at different ages

We also analyzed the number of utterances with previously unseen words in them, arguing that older children's increased memory capacity (Bauer, 2005; Gathercole et al., 2004; Wojcik, 2013) would possibly allow them to draw upon older input more easily in producing speech. Indeed, we found an increase in the number of utterances containing previously unseen words with age in the local sample but a decrease when taking their longer

461 linguistic history into account. The change in word usage we find here could be partly due to

462 a change in linguistic input not captured in the transcripts. The corpus we used is relatively

463 dense: multi-hour at-home recordings made approximately every two weeks for 2–3 years.

464 However, this corpus still only contained a small fraction of what each child heard during the

465 represented periods of time (i.e., 2 hours of ~200 waking hours in a fortnight). Non-recorded

466 caregiver speech may contribute an increasing amount of lexical diversity. Consider, for

467 example, that input from peers containing different lexical items could have increased as

468 children became old enough to independently socialize with other children or attend daycare

469 or preschool (Hoff, 2010; Hoff-Ginsberg & Krueger, 1991; Mannle et al., 1992), which may

470 help to account for the increased presence of words not found in the caregiver's speech. This

471 problem is difficult to address directly since, even with cutting-edge tools and significant

472 supporting resources, it is still nearly impossible to collect and transcribe a child's complete

473 language environment (Casillas & Cristia, under review; B. C. Roy et al., 2009). This effect

474 could instead be simulated in future work by feeding speech from other children or adults

475 into the model to mimic speech from peers and other caregivers. That said, our results

476 showed that the likelihood of previously unseen words actually decreased with age for the

477 cumulative sample, suggesting that the "missing" words *are* present in caregiver speech, just

478 not always in the recently recorded input.

479 Additionally, an improvement in memory capacity with age provides a potential

480 explanation for the current findings. Throughout childhood, including the first few years,

481 SL-relevant cortical regions continue maturing (Casey et al., 2000; Diamond, 2002;

482 Rodríguez-Fornells et al., 2009; Uylings, 2006) with concurrent increases in long-term

483 memory (Bauer, 2005; Wojcik, 2013), working memory, and speed of processing (Gathercole

484 et al., 2004; Kail, 1991). By ages three and four, the children in the current study may have

485 been able to much more reliably draw upon information they were exposed to in the more

486 distant past. If so, we would expect no significant increase in the use of previously unheard

487 words as children get older with the cumulative sampling method—consistent with what we

488 found here (Figure 6, right panel). This pattern of results indicates that children's
489 developing memory could play an important role in the way they use environmental input
490 statistics over age.

Abstraction and complex utterances

492 Our findings are not consistent with a representational shift toward abstraction during
493 the early language learning process. For instance, if children schematized their constructions
494 or switching to rule-based representations (Bannard et al., 2009; Tomasello, 2005; Yang,
495 2016), we would expect a decrease in reconstruction accuracy over time, given that the CBL's
496 reconstructions are limited to the immediate statistics of the child's language environment.
497 In contrast, we saw that the model's ability to reconstruct child utterances from caregivers'
498 speech was age-invariant when taking into account utterance length and chunk duplicates.
499 These results do fall in line with SL theories proposing that the mechanisms for processing,
500 storing, and deploying information stay the constant over age, even though SL behavior on
501 the surface might seen to change over time (e.g., Misyak et al., 2012).

502 As the CBL model only employs a single, simple mechanism for creating and tracking
503 linguistic units, it is impressive that it performs at above-chance levels when accounting for
504 children's speech productions in the first few years. If the mechanism is truly age-invariant,
505 it should be able to handle both young children's speech and adults' speech; here we see that
506 it handles the developing linguistic inventory of children ages 1;0 to 4;0, during which time
507 children's utterances come much more sophisticated and much closer to adult-like form.

508 Going beyond the scope of this paper, a next step would be to explore how the CBL
509 could be modified to augment its performance, particularly on more complex utterances. For
510 example, the CBL model does not include the use of semantics when dividing the caregivers'
511 speech into chunks or when reconstructing the child utterances. However, the meaning of

what both caregivers and child are trying to convey plays a fundamental role in selecting

words from the lexicon and in constructing utterances—they are interacting, and not just

producing speech. The same set of words, ordered in different ways, can have entirely

different meanings (e.g., "the dog bites the man" vs. "the man bites the dog"). Additionally,

the CBL currently works on text-only transcriptions of conversations, but speech prosody

could potentially critically change how children detect chunks. Prosodic structures within an

utterance highlight syntactic structures and help to distinguish between pragmatic

intentions, for example, distinguishing between questions, imperatives, and statements (e.g.,

Bernard & Gervain, 2012; Speer & Ito, 2009). Ideally, the CBL model would also be tested

on a (more) complete corpus of what children hear in the first few years to further

investigate the origins of the "previously unseen" words in children's utterances; though we

appreciate that densely sampled and transcribed collections of audio recordings are

extremely costly to create (Casillas & Cristia, under review; B. C. Roy et al., 2009).

In principle, the "next steps" proposed above—indeed the whole idea of analyzing chunking performance across developmental time—are not limited to the CBL, or even BTP, but rather form a general call for dealing with richer data, regardless of the core underlying mechanism (Aslin et al., 1998; Cleeremans & Elman, 1993; French et al., 2011; Mareschal & French, 2017; Onnis & Thiessen, 2013; Pelucchi et al., 2009; Perruchet & Desaulty, 2008; Perruchet & Vinter, 1998; Saffran et al., 1996). In the current study, we decided to use the CBL because it had previously been successful in reconstructing children's utterances within our target age range (McCauley & Christiansen, 2011, 2014a, 2019) and had not yet been tested for age-invariance. However, given the required memory and comparison (executive function) components of this model, as well as its requirement of discrete (not gradable) chunks, other approaches—particularly those inspired by maturational neurocognitive development (Cleeremans & Elman, 1993; e.g.,

Mareschal & French, 2017; Perruchet & Vinter, 1998)—would be welcome comparisons to the present findings. Notably, while the CBL here performed above chance on average, there is still room to improve in modeling what the children said based on what they heard in the recordings.

Conclusion

In this study, we investigated whether the CBL model—a computational learner using one SL mechanism (BTP)—could account for children's speech production with equal accuracy across ages 1;0 to 4;0 given information about their speech input. **This work extended previous CBL studies by testing the robustness of utterance reconstruction across an age range featuring substantial grammatical development and while also introducing a new controlled accuracy measure for reconstruction.** The model's ability to reconstruct children's utterances remained stable with age when controlling for utterance length and duplicate chunks, both when taking into account recent and cumulative linguistic experience. These findings suggest that this particular mechanism for segmenting and tracking chunks of speech may be age-invariant (Raviv & Arnon, 2018; Shufaniya & Arnon, 2018). A rich and growing literature on SL in development has demonstrated that similar mechanisms can account for much of children's early language behaviors; knowing whether the use of these mechanisms changes as children get older is a crucial piece of this puzzle. To explore this topic further, future work will have to address additional cues to linguistic structure and meaning, the density of data needed to get reliable input estimates, and the interaction of SL with other developing skills that also impact language learning.

Acknowledgements

References

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science*, *14*(3), 464–473.

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth–Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–248.

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, *106*(41), 17284–17289.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Bauer, P. J. (2005). Developments in declarative memory: Decreasing susceptibility to

storage failure over the second year of life. *Psychological Science*, *16*(1), 41–47.

Bernard, C., & Gervain, J. (2012). Prosodic cues to word order: What level of representation? *Frontiers in Psychology*, *3*, 451.

Bulf, H., Johnson, S. P., & Valenza, E. (2011). Visual statistical learning in the newborn infant. *Cognition*, *121*(1), 127–132.

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology*, *54*(1-3), 241–257.

Casillas, M., & Cristia, A. (under review). A step-by-step guide to collecting and analyzing long-format speech environment (LFSE) recordings.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experience. *Cognition*, *87*(2), B69–B77.

Chang, F., Lieven, E., & Tomasello, M. (2006). Using child utterances to evaluate syntax acquisition algorithms. *Proceedings of the 28th Annual Meeting of the Cognitive Science Society*, 154–159.

Chang, F., Lieven, E., & Tomasello, M. (2008). Automatic evaluation of syntactic learners in typologically-different languages. *Cognitive Systems Research*, *9*(3), 198–213.

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in Cognitive Science*, *9*(3), 542–551.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62.

Cleeremans, A., & Elman, J. (1993). *Mechanisms of implicit learning: Connectionist models*

606   *of sequence processing.* MIT press.

607   Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of

608   tactile, visual, and auditory sequences. *Journal of Experimental Psychology:*

609   *Learning, Memory, and Cognition, 31*(1), 24–39.

610   Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical

611   learning in language processing: Word predictability is the key. *Cognition, 114*(3),

612   356–371.

613   Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda

614   licensing in the early acquisition of English. *Language and Speech, 49*(2), 137–173.

615   doi:doi:10.1177/00238309060490020201

616   Diamond, A. (2002). Normal development of prefrontal cortex from birth to young

617   adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R.

618   Knights (Eds.), *Principles of frontal lobe function* (pp. 466–503). New York: Oxford

619   University Press.

620   Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything:

621   Changes in presentation rate have opposite effects on auditory and visual implicit

622   statistical learning. *The Quarterly Journal of Experimental Psychology, 64*(5),

623   1021–1040.

624   French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based

625   connectionist framework for sequence segmentation and chunk extraction.

626   *Psychological Review, 118*(4), 614.

627   Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of

628   non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74.

629   Frost, R. L. A., & Monaghan, P. (in press). Insights from studying statistical learning. In
630   (pp. XX–XX).

631   Frost, R. L. A., Monaghan, P., & Christiansen, M. H. (2019). Mark my words: High
632   frequency words impact early stages of language learning. *Journal of Experimental*
633   *Psychology: Learning, Memory, & Cognition*, Advance online publication.

634   Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of
635   working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2),
636   177–190.

637   Hoff, E. (2010). Context effects on young children's language use: The influence of
638   conversational setting and partner. *First Language*, *30*(3-4), 461–472.

639   Hoff-Ginsberg, E., & Krueger, W. M. (1991). Older siblings as conversational partners.
640   *Merrill-Palmer Quarterly*, *37*(3), 465–481.

641   Johnson, S. P., Fernandes, K. J., Frank, M. C., Kirkham, N., Marcus, G., Rabagliati, H., &
642   Slemmer, J. A. (2009). Abstract rule learning for visual sequences in 8-and
643   11-month-olds. *Infancy*, *14*(1), 2–18.

644   Jost, E., & Christiansen, M. H. (2016). Statistical learning as a domain-general mechanism
645   of entrenchment. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of*
646   *language learning: How we reorganize and adapt linguistic knowledge* (pp. 227–244).
647   Washington D.C.: Mouton de Gruyter.

648   Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in
649   fluent speech. *Cognitive Psychology*, *29*(1), 1–23.

650   Kail, R. (1991). Developmental change in speed of processing during childhood and

adolescence. *Psychological Bulletin*, *109*(3), 490–501.

Kidd, E., Junge, C., Spokes, T., Morrison, L., & Cutler, A. (2018). Individual differences in snfant speech segmentation: Achieving the lexical shift. *Infancy*, *23*(6), 770–794.

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science*, *19*(12), 1247–1252.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Psychology Press.

Mannle, S., Barton, M., & Tomasello, M. (1992). Two-year-olds' conversations with their mothers and preschool-aged siblings. *First Language*, *12*(34), 57–71.

Mareschal, D., & French, R. M. (2017). TRACX2: A connectionist autoencoder using graded chunks to model infant visual statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160057.

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. (2014a). Acquiring formulaic language: A computational model. *The Mental Lexicon*, *9*(3), 419–436.

McCauley, S. M., & Christiansen, M. H. (2014b). Reappraising lexical specificity in children's early syntactic combinations. *Proceedings of the 36th Annual Meeting of the Cognitive Science Society*, 1000–1005.

McCauley, S. M., & Christiansen, M. H. (2017). Computational investigations of multiword

chunks in language learning. *Topics in Cognitive Science*, *9*(3), 637–652.

McCauley, S. M., & Christiansen, M. H. (2019). Language learning as language use: A cross-linguistic model of child language development. *Pyschological Review*, *126*(1), 1–51.

Misyak, J. B., Goldstein, M. H., & Christiansen, M. H. (2012). Statistical-sequential learning in development. *Statistical Learning and Language Acquisition*, 13–54.

Monroy, C. D., Gerson, S. A., & Hunnius, S. (2017). Toddlers' action prediction: Statistical learning of continuous action sequences. *Journal of Experimental Child Psychology*, *157*, 14–28.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268–284.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244–247.

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, *36*(7), 1299–1305.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263.

R Core Team. (2014). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age.

Developmental Science, *21*(4), e12593.

Rodríguez-Fornells, A., Cunillera, T., Mestres-Missé, A., & Diego-Balaguer, R. de. (2009). Neurophysiological mechanisms involved in language learning in adults. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1536), 3711–3735.

Roy, B. C., Frank, M. C., & Roy, D. (2009). Exploring word learning in a high-density longitudinal corpus. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*, 2106–2111.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172–196.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology*, *69*, 181–203.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, *70*(1), 27–52.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105.

Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science*, *42*(8), 3100–3115.

Slone, L. K., & Johnson, S. P. (2015). Infants' statistical learning: 2-and 5-month-olds' segmentation of continuous visual sequences. *Journal of Experimental Child*

717 *Psychology, 133*, 47–56.

718 Speer, S. R., & Ito, K. (2009). Prosody in first language acquisition–Acquiring intonation as
719   a tool to organize information in conversation. *Language and Linguistics Compass,*
720   *3*(1), 90–110.

721 StClair, M. C., Monaghan, P., & Christiansen, M. H. (2010). Learning grammatical
722   categories from distributional cues: Flexible frames for language acquisition.
723   *Cognition, 116*(3), 341–360.

724 Teinonen, T., Fellman, V., Näätänen, R., Alku, P., & Huotilainen, M. (2009). Statistical
725   language learning in neonates revealed by event-related brain potentials. *BMC*
726   *Neuroscience, 10*, 21.

727 Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*
728   (1st ed.). Harvard University Press.

729 Tomasello, M. (2008). Acquiring linguistic constructions. In Damon W, R. Lerner, D. Kuhn,
730   & R. Siegler (Eds.), *Child and adolescent development* (pp. 263–297). New York:
731   Wiley.

732 Uylings, H. B. M. (2006). Development of the human cortex and the concept of "critical" or
733   "sensitive" periods. *Language Learning, 56*(1), 59–90.

734 Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis* (2nd ed.). Springer
735   Publishing Company, Incorporated.

736 Wojcik, E. H. (2013). Remembering new words: Integrating early memory development into
737   word learning. *Frontiers in Psychology, 4*, 151.

738 Yang, C. (2016). *The price of linguistic productivity: How children learn to break the rules of*
739   *language* (1st ed.). MIT Press.