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

Ingeborg Roete[1,2], Stefan L. Frank[2], Paula Fikkert[2], & Marisa Casillas[1]

[1] Max Planck Institute for Psycholinguistics

[2] Radboud University

Abstract

We trained a computational model (the Chunk Based Learner; CBL) on a longitudinal corpus of child-caregiver interactions in English 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 reconstructs 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.

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

Word count: 9891 (7939 excluding references)

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

During the first few years of life children learn the basic building blocks of the language(s) around them. One way they do so is via statistical learning (SL), the process of extracting regularities present in the language environment. Over the past few decades, SL has become a major topic in the field of first language acquisition, ranging in application from speech segmentation (Jusczyk & Aslin, 1995; Saffran, Aslin, & Newport, 1996) and phonotactic learning (Chambers, Onishi, & Fisher, 2003) to producing irregulars (Arnon & Clark, 2011), discovering multi-word structures (Bannard, Lieven, & Tomasello, 2009; Chang, Lieven, & Tomasello, 2006; Frost, Monaghan, & Christiansen, 2019), and much more (see Saffran and Kirkham (2018) for a recent review). By its nature, work in this domain is heavily concerned with a few major topics: (1) the information available in children's language environments (the "input") from which they can pick up on patterns, and (2) the precise mechanisms by which children convert these "raw" environmental statistics into internalized knowledge about language. A third issue is whether and how children's SL behavior changes as they develop (Shufaniya & Arnon, 2018). The current paper taps into each of these three issues: we train a computational model on a longitudinal corpus of child-caregiver interactions to test whether one proposed SL mechanism—backward transitional probability (BTP; Perruchet & Desaulty, 2008)—is able to reconstruct children's speech productions with stable accuracy as they get older.

Statistical learning over development

The ability to detect and store patterns in the environment begins in infancy (e.g., Johnson et al., 2009; Kidd, Junge, Spokes, Morrison, & Cutler, 2018; Saffran et al., 1996; Teinonen, Fellman, Näätänen, Alku, & Huotilainen, 2009), continues into adulthood (e.g., Conway, Bauernschmidt, Huang, & Pisoni, 2010; Frost & Monaghan, 2016; Saffran, Johnson, Aslin, & Newport, 1999), and crosses a range of modalities (Conway &

Christiansen, 2005; Emberson, Conway, & Christiansen, 2011; Monroy, Gerson, & Hunnius, 2017). However, it is still a matter of debate whether SL is an age-invariant skill or not (Arciuli & Simpson, 2011; Raviv & Arnon, 2018; Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; Shufaniya & Arnon, 2018). Recent work that investigates SL abilities in 5–12-year-old children suggests that, while both visual and auditory SL improve with age for non-linguistic stimuli, performance stays the same across childhood for linguistic stimuli (Raviv & Arnon, 2018; Shufaniya & Arnon, 2018). From this finding, the authors conclude that SL for language might be age-invariant. On the other hand, infant SL abilities do appear to shift within the first year, both for linguistic (Kidd et al., 2018) and non-linguistic (Johnson et al., 2009) stimuli. For example, while 11-month-olds can detect and generalize over regularities in a sequence, 8-month-olds are only capable of detecting the regularities, and neither group succeeds yet at learning visual non-adjacent dependencies (Johnson et al., (2009); see also Bulf, Johnson, & Valenza, (2011), and Slone & Johnson, (2015)).

These changes in SL ability during infancy and early childhood may relate to changes in other fundamental cognitive skills. For example, SL-relevant brain regions, such as the prefrontal cortex, continue maturing through childhood (Casey, Giedd, & Thomas, 2000; Diamond, 2002; Rodríguez-Fornells, Cunillera, Mestres-Missé, & Diego-Balaguer, 2009; Uylings, 2006), which may change how children attend to the linguistic information around them as they get older (see also Vlach & Johnson, 2013). Similarly, infants' long-term memory continuously improves between ages 0;2 and 1;6 (Bauer, 2005; Wojcik, 2013). Therefore, the manner in which they store linguistic regularities in long-term memory may also shift during this period. Relatedly, working memory and speed of processing change continuously throughout early childhood (Gathercole, Pickering, Ambridge, & Wearing, 2004; Kail, 1991), implying that there could be a developmental change in the rate and scale at which children can process chunks of information from the unfolding speech signal.

Continued exposure to linguistic input itself can also be an impetus for change in SL ability—a view supported by multiple, theoretically distinct, approaches to early syntactic learning. For example, Yang (2016) proposes that children gather detailed, exemplar-based statistical evidence until it is more cognitively efficient for them to make a categorical abstract generalization. He proposes that, at that point, the learner instantiates a rule to account for patterns in the data. Usage-based theories of early language development alternatively propose that children first learn small concrete linguistic sequences from their input that are made up of specific words or word combinations (e.g., "dog" and "I wanna"; or multi-word combinations, "where's the ..."; Tomasello (2008)). Then, over time, children are proposed to form abstract schemas centered around lexical items (see also Bannard et al. (2009) and Chang et al. (2006)). This representational shift, from probabilistic and lexical to abstract and syntactic, is used to account for how children can eventually create utterances that they have never heard before. Crucially, the representational shift implies a change in the way children apply the original SL mechanism(s) to incoming linguistic information (see also Lany and Gómez (2008)).

Change in SL ability following further linguistic experience is also predicted in models that do not assume abstraction. In chunk-based models of language learning (Arnon, McCauley, & Christiansen, 2017; Christiansen & Arnon, 2017; Christiansen & Chater, 2016; Misyak, Goldstein, & Christiansen, 2012; StClair, Monaghan, & Christiansen, 2010), children use statistical dependencies in the language input (e.g., between words or syllables) to store chunks of co-occurring forms. Dependencies between the chunks themselves can also be tracked with continued exposure and chunk storage (see, e.g., Jost & Christiansen, 2016). In this case, the development of a detailed chunk inventory can gradually change SL performance. Fundamentally, however, this apparent change in SL still comes through the use of the original underlying mechanisms (Misyak et al., 2012); there is no qualitative change in how the system processes data, and the mechanisms for processing, storing, and deploying information stay the same.

Our aim in the present study was to investigate the possibility of developmental change in SL by focusing on a single mechanism that is proposed to be at work over the longer arc of early language development (i.e., in speech segmentation *and* in utterance production and comprehension). Concomitantly, we focused on a developmental language phenomenon that shows gradual change over several years: children's spontaneous utterances. Suiting our needs perfectly, BTP can be applied to the discovery and combination of linguistic chunks to predict patterns in sentence production (McCauley & Christiansen, 2011; Onnis & Thiessen, 2013; Pelucchi, Hay, & Saffran, 2009; Perruchet & Desaulty, 2008). Further, BTP has been proposed as a continuous mechanism throughout development, influencing language processing from infancy to adulthood (Christiansen & Chater, 2016; McCauley & Christiansen, 2019a; Misyak et al., 2012). However, this hypothesis has to our knowledge not yet been tested with longitudinal data. While developmental change in SL could theoretically be tested with many other SL mechanisms and/or developmental language phenomena, the use of BTP and chunking to predict increasing utterance complexity presented a compelling starting place for the present work.

We use a BTP-based computational learner model with a longitudinal collection of natural child-caregiver interaction transcripts to test for developmental change in SL. This computational modeling approach enabled us to define and test the goodness-of-fit of the BTP-based model across the whole period of interest for early speech production (1;0–4;0), and to therefore check whether BTP's performance changed for each child within the studied developmental range. In what follows, we further explain how we chose our model and how we evaluate its results. We then describe the model's accuracy across the tested age range and discuss the implications and limitations of the findings.

Backward transitional probability and the Chunk-Based Learner

The present study uses a model based on McCauley and Christiansen's (2011, 2014a, 2019a) Chunk-Based Learner (CBL), which uses BTP (Perruchet & Desaulty, 2008) to detect statistical dependencies in the speech stream. We chose to focus on the CBL for multiple reasons, as outlined below.

First, as mentioned, we were interested in pursuing a model based on backward transitional probability. BTP is one of multiple approaches for dividing streams of continuous speech into coherent and/or re-combinable units; other approaches include, for example, forward transitional probability (FTP) and memory-based chunking (Aslin, Saffran, & Newport, 1998; Cleeremans & Elman, 1993; French, Addyman, & Mareschal, 2011; Mareschal & French, 2017; Onnis & Thiessen, 2013; Pelucchi et al., 2009; Perruchet & Desaulty, 2008; Perruchet & Vinter, 1998; Saffran et al., 1996). While both BTP and FTP have been shown to effectively enable infants, adults, and simulated learners to segment chunks from continuous speech, direct comparisons between the two for planning and parsing whole spoken utterances suggests an asymmetry in their performance. For example, BTPs outperform FTPs in predicting phonetic word durations in spoken English for both function and content words (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009), in shallow parsing of English, French, and German child-directed speech (McCauley & Christiansen, 2019a), and in reconstructing child-produced sentences in 29 languages (McCauley & Christiansen, 2019a).

Second, among models using BTP, the CBL was of particular interest in the current study because, at the computational level (Marr, 1982), it is designed to be psycholinguistically plausible for utterance processing (see McCauley and Christiansen (2019a) for a review). It uses BTP to incrementally build up an inventory of speech chunks (e.g., “doggy”, “look at”), and stores the chunks and their co-occurrence frequencies such that the accumulated chunk inventory can be used to both parse and produce utterances on the basis of what the model has encountered so far. By only storing shallow information about how chunks combine, its performance in processing multi-chunk utterances also depends exclusively on local relations in the speech signal, mirroring the transitory and sequential nature of spontaneous speech (Christiansen & Chater, 2016). The model can also utilize its BTP-based chunks to engage in *predictive* processing during parsing tasks

(McCauley & Christiansen, 2019a). This “recognition-based prediction” method, together with the central use of multi-word chunks and the parallelism between comprehension and production, renders the CBL impressively consistent with findings from both spontaneous and elicited language processing by adults and children (e.g., Arnon & Snider, 2010; Arnon & Clark, 2011; Diessel & Tomasello, 2000; Ferreira & Patson, 2007; Pickering & Garrod, 2013). Of course, this psycholinguistic plausibility only extends to the computational level of analysis—translations of this model to the algorithmic level will be essential to its long-term utility (Griffiths, Lieder, & Goodman, 2015)—but the CBL’s attention to the incremental, local, and structurally parallel nature of natural language use increased its appeal for the present study.

Third, the CBL has previously succeeded at modeling naturalistic speech production, the task we target in the current paper. For example: (a) as mentioned above, it parsed text better than a shallow parser in three different languages (English, German and French), (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, 2019a; Saffran, 2002). The model has also been able to replicate experimental data on children’s multi-word utterance repetitions (Bannard & Matthews, 2008), over-regularization of irregular plural nouns (Arnon & Clark, 2011), and L2-learner speech (McCauley & Christiansen, 2014b, 2017, 2019a). The model’s performance on utterance production tasks over developmental time is of prime interest as a next theoretical step. Instability in performance over developmental time would hint at significant influences of children’s growing language knowledge, cognitive resources (e.g., working memory, speed of processing), or a combination of the two, on the overt utility of the mechanism.

Basic description of the Chunk-Based Learner. We first briefly describe the CBL model and the performance metrics we use here. BTP for a given pair of words is defined as the occurrence probability of the previous word (w_{-1}) given the current word

(w_0). It can be estimated for each word in a sentence in order to reveal peaks and dips in transitional likelihood, which reflect places where words are likely (peaks) or unlikely (dips) to co-occur. The CBL divides utterances into chunks, splitting the utterances whenever the 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. Once stored in memory, the chunks are not forgotten. 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.

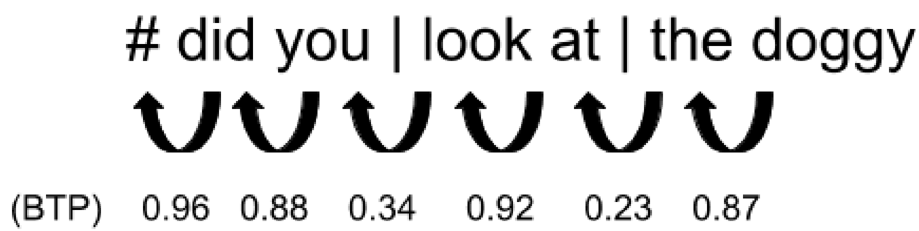

Figure 1. Example of a sentence with BTP between consecutive words. Chunks are split at points of low BTP (indicated by the vertical lines). "#" denotes a start-of-utterance marker.

Testing for change with age. Following McCauley and colleagues (2011, 2014a, 2019b) we tested the CBL model's ability to learn language by checking how well it can reconstruct children's utterances from the chunks discovered in their caregivers' speech. As we are interested in developmental change over the first three years of speech production, we analyzed the model's reconstruction ability with two measures:

- “Uncorrected”: The binary (success/fail) reconstruction score originally used by McCauley and colleagues (2011, 2014a, 2019b).
- “Corrected”: A length-and-repetition-controlled reconstruction score that accounts for the fact that longer utterances have more opportunities for errors in reconstruction, and for the fact that some child utterances contain repetitions of chunks, making multiple reconstructions match the original utterance.

If BTP is an age-invariant mechanism, it should apply equally well across age. However, because children’s utterances get longer as they get older, we would expect age invariance to only hold when we correct for utterance length. We therefore test for age invariance both with the original binary (“uncorrected”) reconstruction score and a new (“corrected”) score we propose to account for utterance length and word repetitions. If we find age-invariance, even while controlling for utterance length and word repetitions, it would strongly suggest that the mechanism is stable over the first three years of speech production and not simply influenced by other factors, e.g., utterance length. Otherwise, it would suggest that the mechanism’s utility for speech production, in fact, changes with age (Bannard et al., 2009; Tomasello, 2005; Yang, 2016).

Predictions

With these previous findings as a starting point, we investigated whether the CBL could reconstruct children’s utterances with equal precision over the first four years of life. Taking for granted that children *eventually* develop abstract representations (as in, e.g., Tomasello, 2008; Yang, 2016), we predicted that:

- The CBL would less accurately generate children’s speech productions as they grew older; given the assumption that children gradually learn to abstract over the specific “chunks” they encounter (Bannard et al., 2009; Tomasello, 2005; Yang, 2016) and, therefore, their speech should less often directly mirror their linguistic input at later ages. This finding would indicate that the immediate influence of children's language input statistics on their speech production decreases across development.

- Children will be more likely to use words that are not documented in the caregiver speech as they get older. These words could originate from other sources, such as peer speech or non-recorded caregiver speech (Hoff, 2010; Hoff-Ginsberg & Krueger, 1991; Mannle, Barton, & Tomasello, 1992; Roy, Frank, & Roy, 2009).
- Younger children's utterances would be reconstructed well on the basis of recently heard speech alone, whereas older children's utterances would be best constructed when considering a longer period of their historical input. Our reasoning was that older children's increased memory capacity (Bauer, 2005; Gathercole et al., 2004; 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 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 placement of chunk boundaries 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 discovered chunks and their co-occurrence counts (enabling it to compute BTPs between chunks on the fly). For example, the model might parse the utterances "I see the puppy" and "did you look at the puppy?" into five different chunks, namely "I", "see", "the puppy", "did you", and "look at" based on the BTPs between these words compared to the average BTP found in the inventory 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. We follow McCauley and Christiansen's (2011) 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 related BTPs) derived from the adult's utterances. Our 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 the set of largest possible chunks that had already been seen 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,

[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.

the child utterance “look at the puppy” would be broken down into a set of known chunks which were discovered in the adults’ speech (e.g., “look at” and “the puppy”, as in the step 2 rounded boxes). If the utterance were “look at the poodle”, and the model had not already added a chunk for the word “poodle” 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 reordering the chunks detected, based on their known BTPs: the model begins with the utterance start cue and then finds the chunk that has the highest BTP to the start cue, following that first chunk with the next one, which will be the remaining chunk with the highest backwards transitional probability to the first chunk, and again and again, until the set of chunks for that utterance is exhausted. So, the set of chunks “look at” and “the puppy” would be ordered depending on the chunk that maximizes the BTP of the start cue (i.e., “look at”), followed by the chunk that maximizes the BTP of “look at” (i.e., “the puppy”).

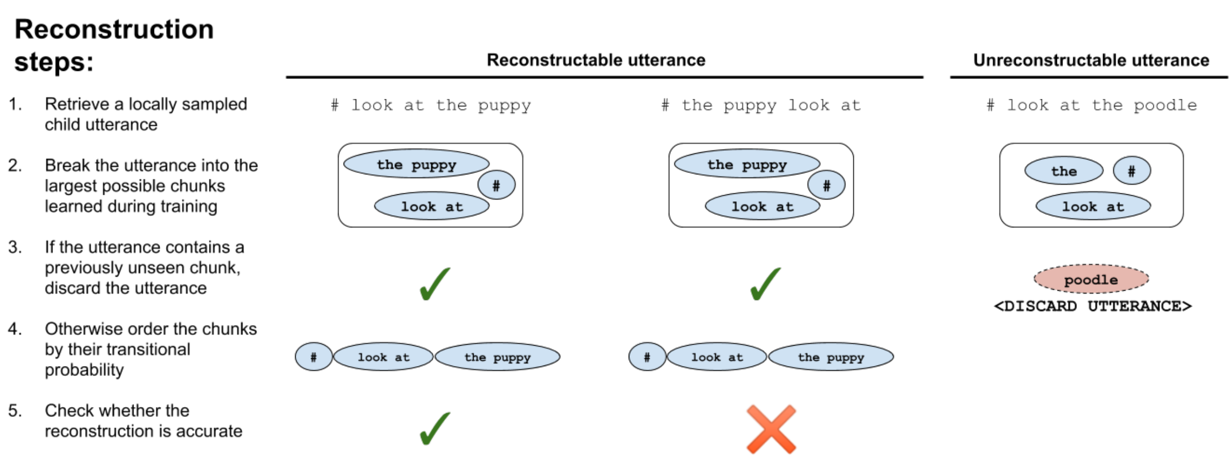

Figure 2. Example of reconstruction attempts for three child utterances. The model tries to reconstruct the first two utterances using BTPs of the chunks it finds, but it cannot do so with the third utterance, which contains a word ("poodle") that had not been previously seen during training.

Materials and Procedure

As input to the model we used transcripts of 1–2-hour recordings of at-home interaction between six North American English-speaking 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 help reconstruct 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

[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.

age was a better method for getting a 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 help reconstruct 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, 2019b) 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 steps 4 and 5 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 a word encountered during the reconstruction task was not seen during training; utterances with unseen words, by our version of the CBL, cannot be reconstructed (see step 3 in Figure 2).

We used mixed-effects regression to analyze the effect of child age on both of the reconstruction scores and also whether a word encountered during the reconstruction task was not encountered during training. All mixed-effects models included child age as a fixed effect and by-child random intercepts with random slopes of child age. By default, child age was modeled in years (1–4) so that the intercept reflects a developmental trajectory beginning at age 0. However, for the model of corrected reconstruction accuracy we had the additional advantage of being able to test whether the CBL performance significantly exceeded the baseline chance of correct reconstruction. We tested this difference at the average age in our longitudinal dataset (2;6) by centering age on zero in the statistical model (ages 1;0–4;0 are re-coded numerically as $-$1.5 to $+$1.5) such that the default model output would reflect the estimated difference from chance at the middle point of our age range.

All analyses were conducted using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015) and all figures of the findings were generated with the `ggplot2` package in `R` (R Core Team, 2014; Wickham, 2009). All code used to create the model and analyze its output is available at https://github.com/marisacasillas/CBL-Roete. Full tables of statistical model output are available in the Supplementary Materials. Before turning to the main results we briefly describe the corrected reconstruction score and the analysis of previously unseen words in more detail.

Corrected reconstruction accuracy

The corrected, length-and-repetition-controlled reconstruction score is a function of three factors: (a) whether the model successfully reconstructed the child utterance or not, (b) the number of chunks used to reconstruct the utterance, and (c) the number of duplicate chunks involved in the reconstruction. By taking the number of chunks into account, this reconstruction score compensates for the fact that successful reconstruction is less likely for longer utterances. When an utterance contains duplicate chunks, the exact ordering of those duplicate chunks does not influence the correctness of the reconstruction. For example, if the utterance “I wanna, I wanna” is decomposed into the two chunks “I wanna” and “I wanna”, it does not matter which of the two “I wanna” chunks is placed first when determining whether reconstruction accuracy was successful. Thus, utterances containing duplicate chunks are more likely to be reconstructed by chance alone than utterances with the same number of chunks but no duplicates. Note that here we are detecting duplicate *chunks* in the utterance rather than duplicate *words*. At this post-training stage, the model is only able to parse the utterance into chunks; that is the relevant unit over which duplication may affect reconstruction accuracy.

An utterance that is decomposed into N unique chunks can be reconstructed in $N!$ different orders. Hence, the baseline probability of obtaining the correct order of N unique chunks equals $1/N!$. When we take into account that chunks can be repeated within an utterance, chance level equals $(n_1!n_2!\ldots n_k!)/N!$, where N is the total number of chunks in the utterance, and $n_1, \ldots, n_k$ are the number of times a chunk is repeated for each of the k unique chunks found in the utterance (Figure 3).

When probability of reconstruction was lower, we scored a correctly reconstructed utterance higher. We assigned a score of $-\log(chance)$ for each correct reconstruction and $\log(1 - chance)$ for each incorrectly reconstructed utterance. In layperson’s terms, this means that successfully reconstructed utterances were scored positively, but were weighed relative to the number of chunks and the number of repetitions they had, such that reconstructions of long utterances were given higher scores than reconstructions of short utterances. Along 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.

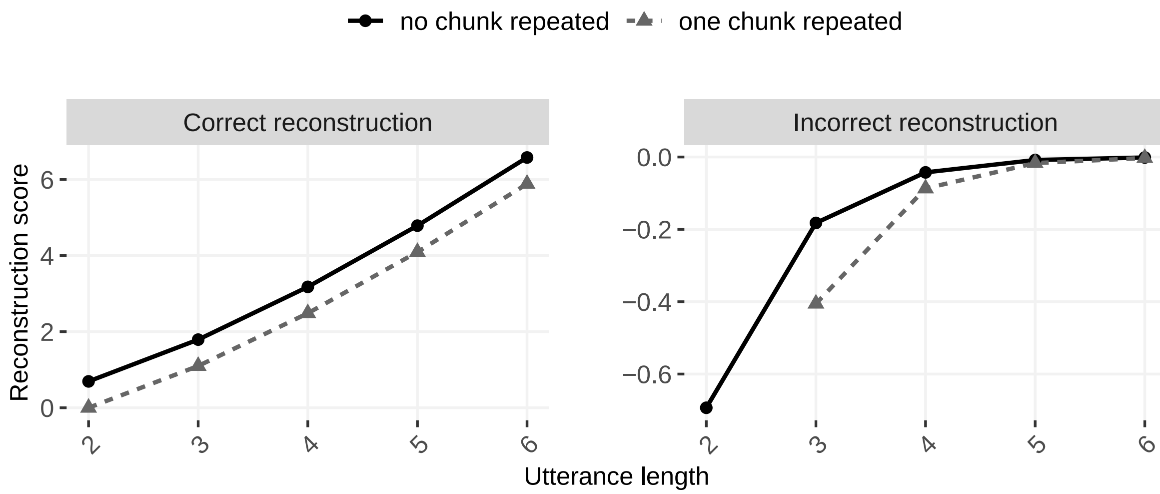

Figure 3. Corrected reconstruction score for correct (left; positive values) and incorrect (right; negative values) reconstructions, as a function of utterance length (2–6 chunks). In this example, either no chunks are repeated (black/solid lines) or one chunk occurs twice in the utterance (gray/dashed lines).

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 the likelihood that words used in the child utterances were seen during training, given child age and sampling type. To prepare for this analysis we marked each word used by each child at each age point as having been seen during training (1) or not (0), given local and cumulative sampling.

Results

Uncorrected reconstruction accuracy

The uncorrected score of accurate utterance reconstruction (McCauley & Christiansen, 2011, 2014a) showed that 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 = 65.4%, range across children = 59.9%–70.3%; cumulative: mean = 59.9%, range across children = 53.1%–68.2%). This is similar to, or slightly higher than, results reported by McCauley and Christiansen (2011) who found an average percentage of correctly reconstructed utterances of 59.8% over 13 typologically different languages with a mean age range of 1;8–3;6 years. Additionally, McCauley and Christiansen (2019b) reported an average reconstruction percentage of 55.3% for 160 single-child corpora of 29 typologically different languages, including a performance of 58.5% for 43 English single-child corpora with a mean age range of 1;11–3;10.

In our statistical models of the uncorrected reconstruction accuracy[3], we first analyzed the CBL model's performance when it was trained on locally sampled caregiver speech. The number of correctly reconstructed utterances decreased with age ($b = -0.805, SE = 0.180, p < 0.001$); over time the BTP statistics present in the caregivers' speech were less reflected in the child's own speech (Figure 4, left panel), as we shall see, this decrease is due to the fact that the score was uncorrected.

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 4, 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 probability of accurate reconstruction. Utterances from older children tended to contain more words (and typically therefore more chunks) than utterances from younger children (Figure 5, 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 5, right panel). Utterances with duplicate chunks have 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.

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

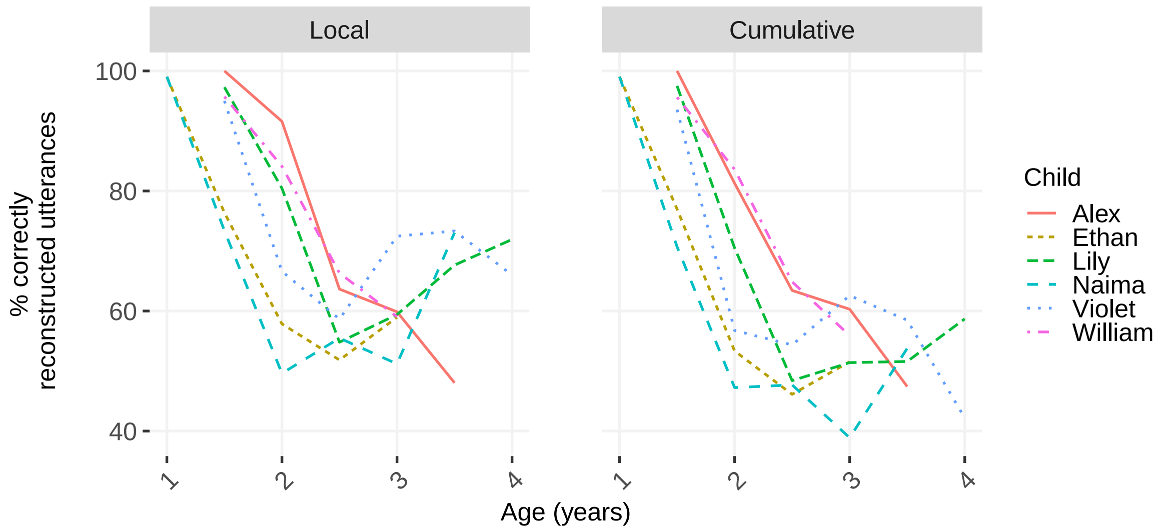

Figure 4. Percentage of correctly reconstructed utterances across the age range, using local (left) and cumulative (right) sampling.

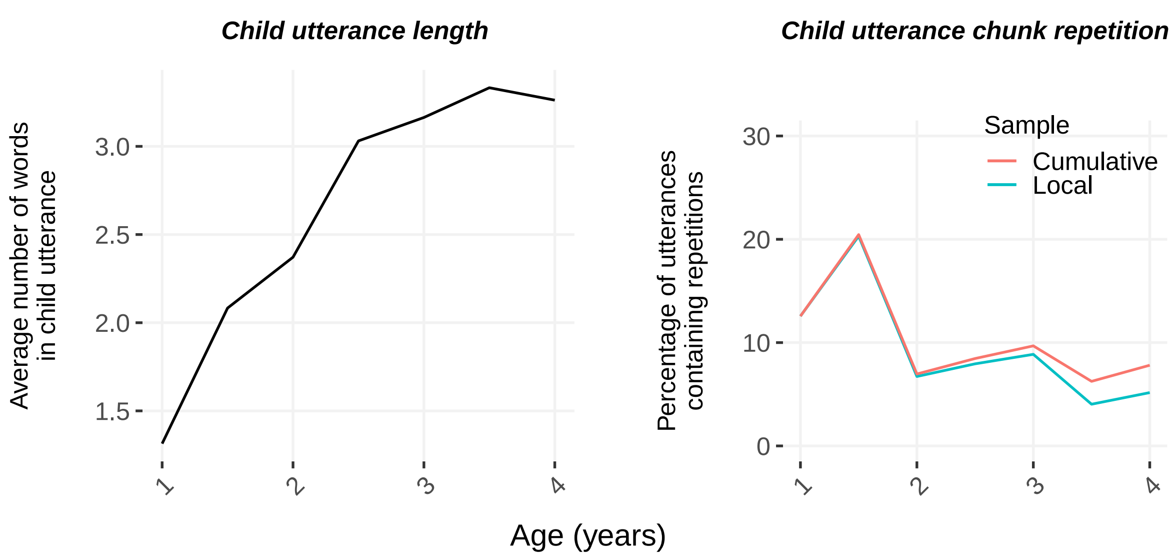

Figure 5. Children's utterances increased in length (number of words) with age (left) while simultaneously decreasing in the number of duplicate chunks used (right).

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. As explained above, the corrected 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 whether the corrected reconstruction score was above chance—here represented by a score of zero—so in these particular statistical models we centered child age on zero so that the estimation would reflect the difference from zero for the average age in our sample (2;6).[4]

Again, we first analyzed the model's performance when it was trained on locally sampled caregiver speech. We found a significant positive intercept ($b = 0.11, SE = 0.02, t = 5.064$) and no significant change across age ($b = 0.030, SE = 0.018, t = 1.681$); the BTP statistics from the caregivers' speech were consistently reflected in the child's own speech (Figure 6, left panel).

As before, we created a parallel set of analyses to test the model's performance when it was trained with a cumulative sample of caregiver speech. We again found a significant positive intercept ($b = 0.06, SE = 0.010, t = 6.238$) and that accuracy did not change significantly across age ($b = 0.02, SE = 0.013, t = 1.590$; Figure 6, right panel).

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

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

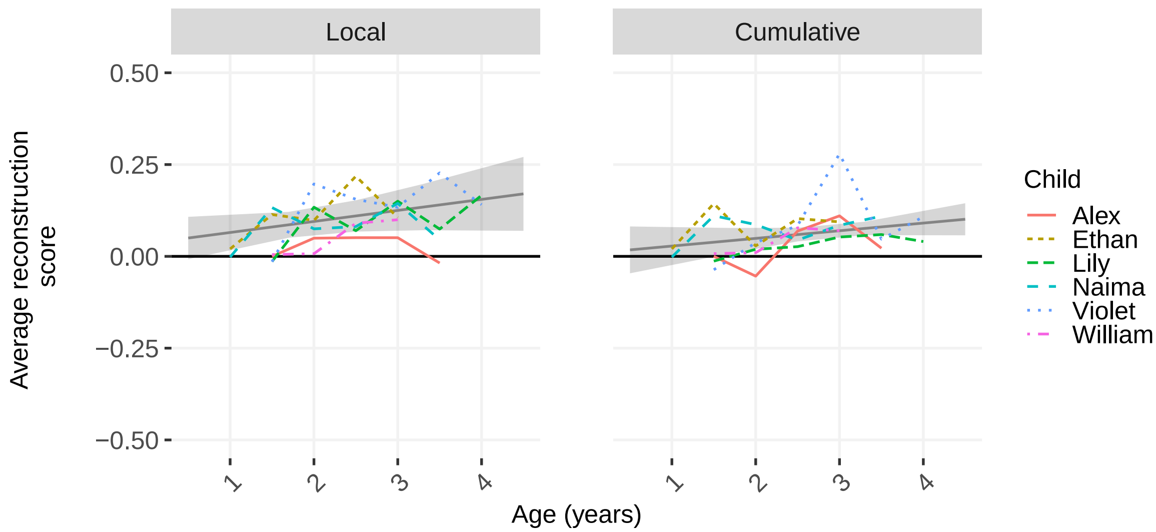

Figure 6. Corrected reconstruction scores across the age range, using local (left) and cumulative (right) sampling. By-child scores are shown in the colored lines, with the mixed-effect model estimates (fit = line and confidence interval = band by age) shown in gray.

Children's use of sentences containing unseen words

Utterances with words that were not encountered and stored as chunks during training were not included in the reconstruction task. We therefore also analyzed whether child age and sampling type influenced the likelihood that a word in the child's speech had already been seen. For this analysis we compared the words used by each child at each age point to the words that *that* child had heard during training (local or cumulative), marking each word as having been seen during training (1) or not (0). For each sampling type, we then modeled the likelihood that a word was previously seen given a fixed effect of child age and random effect child with random slopes of child age.[5] With local sampling, words in the children's utterances were significantly less likely to have been previously seen as children got older ($b = -0.549, SE = 0.11, p < 0.001$; Figure 7, left panel). With cumulative sampling, this effect was neutralized; increasing age was associated with a small

[5] prev_seen ~ age + (age|child), family = binomial(link = “logit”)

and non-significant decrease in the likelihood of previously seen words ($b = -0.022, SE = 0.121, p = 0.857$; Figure 7, right panel). By taking a longer history of linguistic input into account (i.e., by using cumulative sampling), words that were not seen in the local sampling were indeed seen during cumulative sampling.

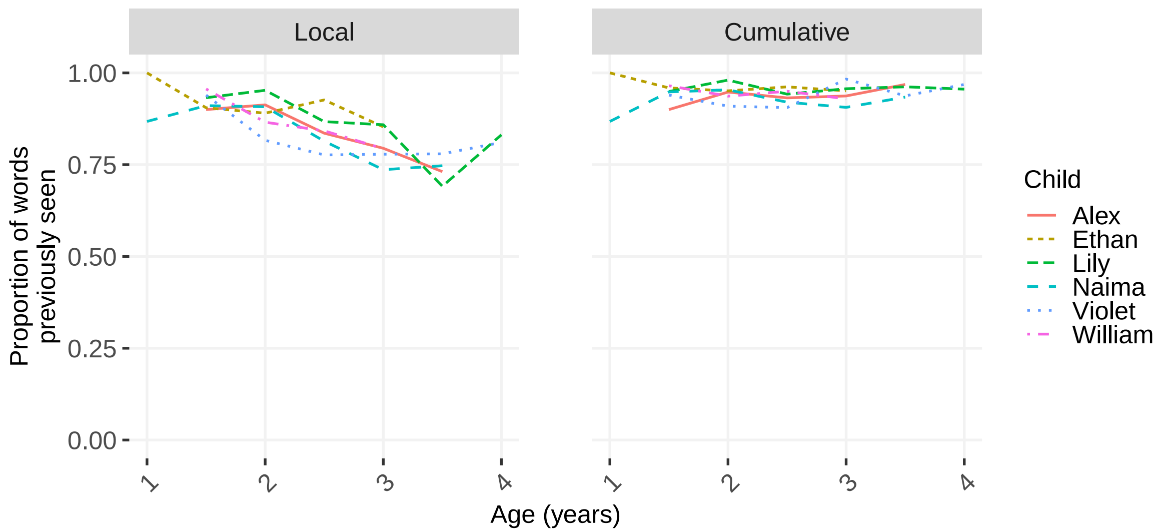

Figure 7. Proportion of words in the local child utterances seen in the training data across age using local (left) and cumulative (right) sampling.

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 reconstruct 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, with both shorter and longer memory of caregiver speech (i.e., local and cumulative input). Overall, and against our predictions, the current findings support the view that BTP is an age-invariant learning mechanism for 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 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 linguistic history into account. The change in word usage we find here could be partly due to a change in linguistic input not captured in the transcripts. The corpus we used is relatively dense: multi-hour at-home recordings made approximately every two weeks for 2–3 years. However, this corpus still only contained a small fraction of what each child heard during the represented periods of time (i.e., 2 hours of ~200 waking hours in a fortnight). Non-recorded caregiver speech may contribute an increasing amount of lexical diversity. Consider, for example, that input from peers containing different lexical items could have increased as children became old enough to independently socialize with other children or attend daycare or preschool (Hoff, 2010; Hoff-Ginsberg & Krueger, 1991; Mannle et al., 1992), which may help to explain the increased presence of words not found in the caregiver's speech. This problem is difficult to address directly since, even with cutting-edge tools and significant supporting resources, it is still nearly impossible to collect and transcribe a child's complete language environment (Casillas & Cristia, 2019; Roy et al., 2009). This effect could instead be simulated in future work by feeding speech from other children or adults into the model to mimic speech from peers and other caregivers. That said, our results show that the likelihood of previously unseen words actually decreased with age for the cumulative sample, suggesting that the "missing" words *are* present in caregiver speech, just not always in the recently recorded input.

An improvement in memory capacity with age could also provide a potential explanation for these findings on children's use of previously unheard words. Throughout childhood, including the first few years, SL-relevant cortical regions continue maturing (Casey et al., 2000; Diamond, 2002; Rodríguez-Fornells et al., 2009; Uylings, 2006) with concurrent increases in long-term memory (Bauer, 2005; Wojcik, 2013), working memory, and speed of processing (Gathercole et al., 2004; Kail, 1991). By ages three and four, the children in the current study may have been able to much more reliably draw upon information they were exposed to in the more distant past. If so, we would expect no significant increase in the use of previously unheard words as children get older with the cumulative sampling method—consistent with what we found here (Figure 7, right panel). This pattern of results indicates that children's developing memory could play an important role in the way they use environmental input statistics over age.

Abstraction and complex utterances

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

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

Going beyond the scope of this paper, a next step would be to explore how the CBL could be modified to augment its performance, particularly on more complex utterances. For example, the CBL model does not include the use of semantics when dividing the caregivers' speech into chunks or when reconstructing the child utterances. However, the meaning of what both caregiver 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 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, 2019; Roy et al., 2009).

Limitations and Future Work

Although the CBL was perfectly suited for this initial investigation (see Introduction), it is unclear how this model could be implemented at the neural level. In particular, the CBL does not specify how BTP (between chunks, and the running average) is stored in the brain, nor how the comparison mechanism that inserts chunk boundaries is implemented. The model's requirement for access to precise estimates of BTP between any two chunks may, with accumulated natural input, hugely increase its memory requirements. That said, these probabilities could potentially be approximated more efficiently in a neural net, which would also yield more graded, abstract chunks.

Perhaps more troubling is the BTP comparison mechanism, which presumably relies on functions of executive control, working memory, and/or long-term memory, and which is likely influenced by the child's speed of processing, all of which are known to change dramatically during the developmental period tested here (Bauer, 2005; Casey et al., 2000; Diamond, 2002; Gathercole et al., 2004; Kail, 1991; Rodríguez-Fornells et al., 2009; Uylings, 2006; Wojcik, 2013). Why, then, do we find no age effect here? We propose two possibilities that could be explored further: (a) while these memory, processing, and executive control functions *do* improve with age, they are already sufficient early on for the foundational computations of the model, and their increased functioning only comes into play once children begin to produce highly complex utterances; (b) caregiver linguistic input itself, perhaps via the child's signs of comprehension, closely parallels these maturational gains (e.g., via "fine-tuning"; Roy et al. (2009); Snow (2017)). Again, neural networks may be a natural option for exploring how changes in these maturational factors interact with changing input in the creation and storage of chunks. If further research did find that developmental change alters the CBL's ability to reproduce children's utterances, it would raise questions about the age-invariant influence of BTP over development. A similar approach could be taken to comparably investigate age-related change in the use of other mechanisms, including FTP.

In principle, these “next steps”—calling for the use of richer data—and potential neural-net implementations—to better simulate storage and processing limitations—could be explored using a number of different SL mechanisms for speech segmentation, comprehension, and production (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 fact, a number of existing models already take closer inspiration from neurocognitive maturational findings (e.g., Mareschal & French, 2017; Cleeremans & Elman, 1993; Perruchet & Vinter, 1998), and a side-by-side comparison of their longitudinal performance on natural language data with the CBL would be a worthwhile follow-up to the present research. Notably, while the CBL here performed above chance on average, there was still room to improve; another model may show even better performance, or the CBL might improve upon the addition of some of these maturational features.

Conclusion

In this study, we investigated whether the CBL model—a computational learner using one SL mechanism (BTP)—could reconstruct children’s spontaneous speech productions 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 reconstruct 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—BTP, but also other mechanisms—with other developing skills that also impact language learning.

Acknowledgements

We owe big thanks to Rebecca L. A. Frost and the MPI for Psycholinguistics' First Language Development group for insightful comments on earlier versions of this paper. This work was supported by an IMPRS fellowship awarded to IR and a Veni Innovational Research grant to MC (275-89-033).

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., 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.

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

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

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.

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.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English.

Journal of Memory and Language, *60*(1), 92–111.

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. (2019). A step-by-step guide to collecting and analyzing long-format speech environment (LFSE) recordings. *Collabra*, *5*(1), 24. https://doi.org/doi:10.1525/collabra.209

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 of sequence processing*. MIT press.

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356–371.

Conway, C. M., & Christiansen, M. H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 24–39.

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*(2), 137–173. https://doi.org/doi:10.1177/00238309060490020201

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. In D. Stuss & R. Knights (Eds.), *Principles of frontal lobe function* (pp. 466–503). New York: Oxford University Press.

Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, *11*(1/2), 131–152.

Emberson, L. L., Conway, C. M., & Christiansen, M. H. (2011). Timing is everything: Changes in presentation rate have opposite effects on auditory and visual implicit statistical learning. *The Quarterly Journal of Experimental Psychology*, *64*(5), 1021–1040.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71–83.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614.

Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70–74.

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

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

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

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

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

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

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

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

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.

Marr, D. (1982). *Vision.* San Francisco, CA: W. H. Freeman.

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. (2019a). Language learning as language use: A cross-linguistic model of child language development. *Psychological Review*, *126*, 1–51.

McCauley, S. M., & Christiansen, M. H. (2019b). 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.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*(04), 329–347.

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.

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/

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

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 Psychology*, *133*, 47–56.

Snow, C. E. (2017). Issues in the study of input: Finetuning, universality, individual and developmental differences, and necessary causes. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp. 179–193). John Wiley & Sons, Ltd.

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

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

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

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

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

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

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382.

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

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

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