Cognition 230 (2023) 105275

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

How infants' utterances grow: A probabilistic account of early language development

Qihui Xu [a,*], Martin Chodorow [b], Virginia Valian [b]

[a] Department of Psychology, Graduate Center, City University of New York, New York City, NY, USA
[b] Department of Psychology, Hunter College, City University of New York, New York City, NY, USA

ARTICLE INFO

Keywords:
Language acquisition
Corpus linguistics
Computational models
Utterance length

ABSTRACT

Why are children's first utterances short and ungrammatical, with some obvious constructions missing? What determines the lengthening of children's early utterances over time? The literature is replete with references to a one-word, a two-word, and a later multiword stage in language development, but with little empirical evidence, and with little account for how and why utterances grow. To address these questions, we analyze speech samples from 25 children between the ages of 14 and 43 months; we construct distributions of their utterances of lengths one to five by age. Our novel findings are that multiword utterances of different lengths appear early in acquisition and increase together until they reach relatively stable proportions similar to those found in parents' input. To explain such patterns, we develop a probabilistic computational model, VIRTUAL, that posits an interaction between a) varying, increasing resources from various developmental domains and b) target utterance lengths mirroring the input. VIRTUAL successfully accounts for most of the empirical patterns, suggesting a probabilistic and dynamic process that is nonetheless compatible with apparent distinct milestones in development. We provide a new, systematic way of showing how developmental cascade theories could work in language development. Our findings and model also suggest insights into syntactic, semantic, and cognitive development.

1. Introduction

Children's first utterances are short and often have omission errors, such as "key open door" and "blue car broken down"; they gradually lengthen as children get older (Bloom, 1973; Bowerman, 1973; Braine & Bowerman, 1976; Brown, 1973; Goldin-Meadow & Singer, 2003). Utterance length has been considered a general indicator of language development and is frequently used to describe developmental changes in linguistic competence. Previous work suggests strong relations between utterance length and the development of syntax, semantics, and vocabulary (Blake, Quartaro, and Onorati, 1993; Brown, 1973; Devescovi et al., 2005; Rollins, Snow, and Willett, 1996; Scarborough, Rescorla, Tager-Flusberg, Fowler, and Sudhalter, 1991). For example, utterance length is correlated with a variety of syntactic milestones (e.g., Blake et al., 1993; Brown, 1973; Valian, 1991) and linguistic complexity (e.g., Ambridge, Kidd, Rowland, & Theakston, 2015; Le Normand, Moreno-Torres, Parisse, & Dellatolas, 2013). The determinants of utterance length also bear on the roles that various systems external to language, such as cognitive and motor systems, play in development (Berk & Lillo-Martin, 2012; Moore & Maassen, 2004). Yet the trajectory of children's development of utterance length as children move beyond single words is unknown. In the present work we a) characterize the developmental changes in utterance lengths starting very early in acquisition and b) model the underlying processes that could produce such changes.

The literature is replete with references to a one-word, a two-word, and a later multiword stage in language development, stages that are often considered as qualitatively different from each other (e.g., Bloom, 1973; Brown, 1973; Butcher and Goldin-Meadow, 2000; Herr-Israel and McCune, 2011; Leopold, 1949). But the data are inconsistent with respect to what very early combinations look like (are they limited to two words or a mixture of different lengths?), how early in development word combinations occur (two years old or earlier?), and how they change over time (in distinct stages or continuously? Braine & Bowerman, 1976; Herr-Israel & McCune, 2011; Leopold, 1949; Radford, 1990). If the development of utterance length over time is continuous, will it be compatible with apparent milestones as observed in the previous literature on the one-word and the two-word stages (Bloom, 1973;

* Corresponding author.
E-mail address: qxu@gradcenter.cuny.edu (Q. Xu).

https://doi.org/10.1016/j.cognition.2022.105275
Received 6 October 2021; Received in revised form 27 August 2022; Accepted 29 August 2022
0010-0277/Published by Elsevier B.V.

Brown, 1973; Butcher and Goldin-Meadow, 2000; Herr-Israel and McCune, 2011; Leopold, 1949)? Without knowledge of the trajectory, it will be difficult to understand the mechanisms governing the development of combinatorial speech.

In contrast to stage theories, a variety of theories view change as a developmental cascade shaped by various resources (Oakes and Rakison, 2019; see Iverson, 2021 for a review). Rather than abrupt qualitative changes, development is a gradual and cumulative consequence of the dynamic interactions between many domain-general and domain-specific developing resources or systems. Consistent with the idea of cascades, we propose a continuity and a nonlinearity in the development of utterances of different lengths. For the continuity, we expect that utterances of different lengths will co-occur during the earliest period of combinatorial speech, reflecting a fluctuating process in which less advanced and more advanced productions coexist and develop simultaneously. For the nonlinearity, we expect that the development of utterance length can be better described by a nonlinear trajectory, as opposed to a straight one. Observed milestones are compatible with underlying gradual changes in development. Note that while the term – nonlinearity - may be used differently in different fields, we use nonlinear to refer to patterns of change that are different from straight-line trajectories.

We highlight two possible sources of change. One is the growth of various resources over time including, among others, working memory (Newbury, Klee, Stokes, and Moran, 2016), lexical knowledge (Nóro and Mota, 2019), and articulatory capacity (Moore and Maassen, 2004). Children's utterances are thus initially short due to limited cognitive and/or linguistic resources, gradually becoming longer and more complete over time as those resources increase (Newberry et al., 2016; Moore and Maassen, 2004; Nóro and Mota, 2019).

A second source is adult language input, which provides target utterances of different lengths for children to aim for. Adult input affects many aspects of language development in the first year, including word meanings (Bergelson and Swingley, 2012), word segmentation (Christophe, Dupoux, Bertoncini, and Mehler, 1994), syntactic rules (Shi, Werker, and Morgan, 1999), and word order (Gervain, Nespor, Mazuka, Horie, and Mehler, 2008). The presence of different utterance lengths in the input potentially provides children with examples of lengths to model.

The two sources – increasing resources and an input distribution of model lengths – suggest a dynamic interaction between what children intend to say and what they can actually produce. Although children's early short and ungrammatical utterances, such as those missing obvious grammatical constructions, might be evidence of a lack of linguistic competence (Braine and Bowerman, 1976; Pine, Lieven, and Rowland, 1997), they might instead reflect limitations in the resources necessary for combinatorial speech (Valian, 1991). In tests of comprehension, children appear to have more advanced linguistic knowledge than their productions indicate (Goldin-Meadow, Seligman, and Gelman, 1976; Santelmann and Jusczyk, 1998; Shi et al., 1999; Shipley, Smith, and Gleitman, 1969; Valian, Hoeffner, and Aubry, 1996). Even within production, utterances conveying the same information can vary in length under different conditions; for example, Valian and Aubry (2005) showed that children produced more elements when given a second chance to repeat the same utterance, suggesting that they can produce longer utterances when their resources are augmented by prior lexical look-up and overall semantic analysis.

Given that both the increasing resources and an input distribution of model lengths have been shown to affect utterance development (e.g., Braine and Bowerman, 1976; Christophe et al., 1994; Newbury et al., 2016; Shi et al., 1999; Valian and Aubry, 2005), it is likely that utterance length development is a cumulative developmental cascade and reflects the interactive effects of those systems. We propose a model, which we call VIRTUAL (varying, increasing resources and target utterance (adult) lengths), of how the two sources interact to influence the development of utterance length. Without a resource limitation, children's utterance lengths would resemble the length distribution we see in adults; since two-year-olds are resource-limited in linguistic, cognitive, and/or biological domains, their utterances will often be shorter than they intend. In effect, the child reduces the intended utterance to the length that her resources can support. Because of the way the sources interact, the distribution of child utterance lengths is the product of probabilistic rather than an all-or-none processes. Similar to the cascade account for developmental change (Oakes and Rakison, 2019), VIRTUAL portrays underlying development as a gradual, continuous process that is nonetheless compatible with apparent distinct milestones.

In Study 1, we document the developmental trajectories of utterances of different lengths by examining early multiword utterances in spontaneous speech samples from 25 children (MacWhinney, 2000). If the development of utterance length is a gradual and cumulative consequence of many other developing resources and systems that is nonetheless compatible with apparent distinct milestones in observed behaviors, then, in Study 1, we should see the continuous and nonlinear development, rather than discrete stages, of utterances of various lengths even during the earliest period of combinatorial speech. In Study 2, we use VIRTUAL, a probabilistic model that embraces the underlying change with continuity and dynamic interactions, to simulate the behavioral patterns characterized in Study 1.

2. Study 1

Study 1 documents the distribution of early utterance lengths and examines their trajectories of change over time. We establish when combinations begin and how many words children combine to determine whether the development of different utterance lengths is continuous and simultaneous or in separate stages. We then track with regression analysis how that number changes across development. To determine whether the development of different utterance lengths is linear or nonlinear, we contrast a linear and a nonlinear regression model.

2.1. Method

2.1.1. Data

We screened all the longitudinal English corpora in CHILDES (MacWhinney, 2000), selecting 24 children whose data fit the following criteria: a) their recordings started no later than 23 months, b) their corpus contained at least one hour-long recording per month, and c) the interval between consecutive sessions was $\leq$2 months (Bloom, 1970; Braunwald, 1971; Brown, 1973; Demuth and McCullough, 2009; Higginson, 1985; Jones and Rowland, 2017; MacWhinney, 1991; Parsons, 2006; Post, 1992; Theakston, Lieven, Pine, and Rowland, 2001). Due to sparse samples at some age points, we focused on the range between 14 and 43 months; each age point consisted of data from at least four children. We similarly tracked the adult productions for each child.

Among the 24 children, recordings of eight children – the *earlier* group – started no later than 16 months. The remaining 16 children – the *later* group – started no later than 23 months (Fig. 1). Fig. 2 shows the number of word tokens produced by each child in the *earlier* group from 14 to 43 months of age (for children in the *later* group, see Appendix A).

2.1.2. Procedure

Utterances that are counting or routines were excluded. Imitations were excluded if they immediately repeated a parent's whole utterance or a subset of it. Utterances with repetitive words at adjacent positions, unintelligible markers, or filler words without specific meanings were shortened. See Table 1 for statistics and examples. To identify filler words without specific meanings, we first used the part-of-speech tags provided by CHILDES (MacWhinney, 2000) to retrieve words tagged as ‘co’ (i.e., communicators). Next, we asked seven annotators to judge whether the words have meaning. A word was considered a filler word without specific meaning if at least one annotator judged it to be.

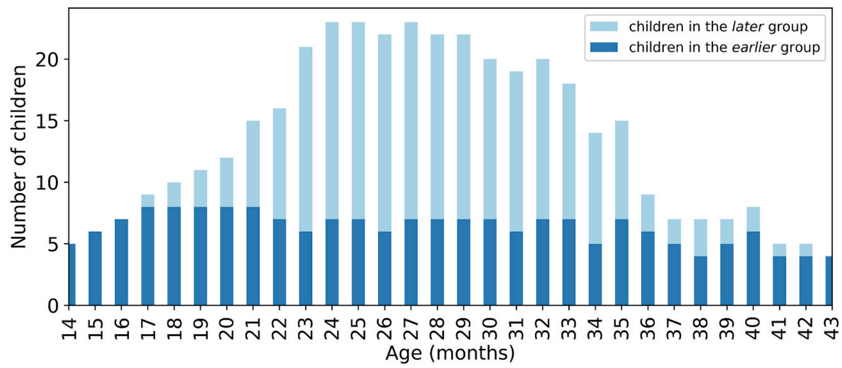

Fig. 1. Number of children contributing data at each age point. The darker blue bars on the bottom represent the children in the *earlier* group and the lighter ones on the top represent the children in the *later* group. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

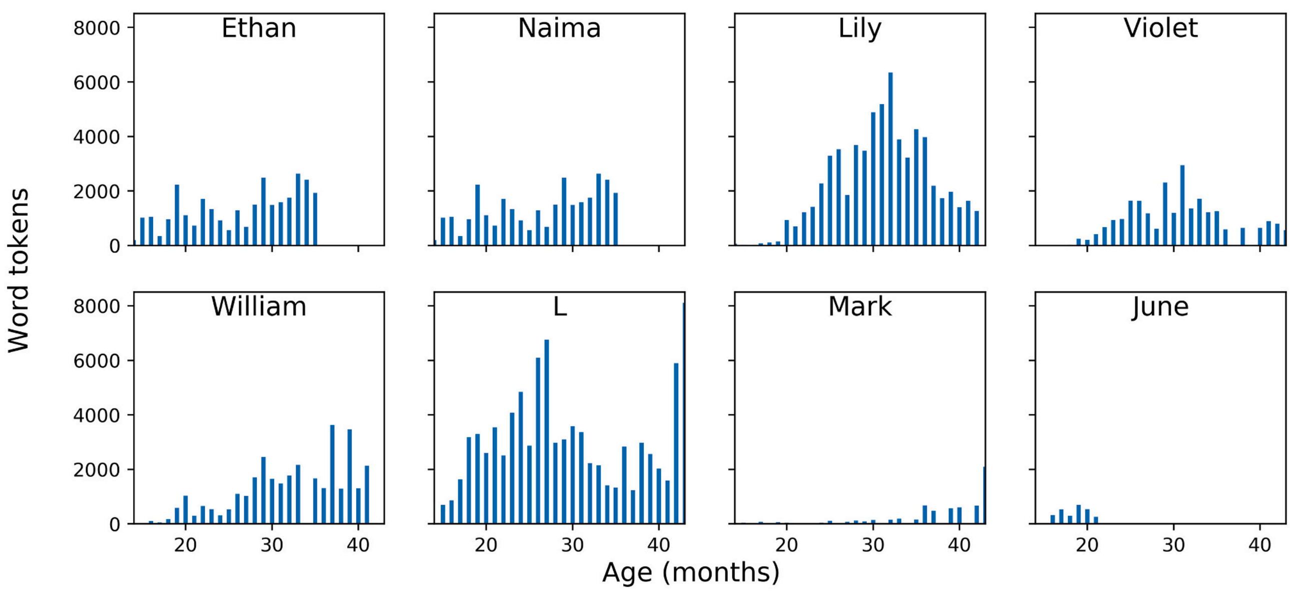

Fig. 2. Number of word tokens by age for children in the *earlier* group. See Appendix A for children in the *later* group.

Table 1
Count of utterances excluded or used and examples for each filtering criterion.

		N (%)	Criterion	Criterion-wise *N* (%)	Example
Utterances excluded	Utterances only containing illegal elements	93,011 (19%)	–	–	"xxx."; "Ah."
	Utterances of lengths 1–5	26,835 (5%)	Immediate imitations	21,133 (4%)	MOT: "I see the red car." CHI: "Red car."
			Counting	3565 (1%)	"One two three."
			Routine	2078 (0.4%)	"a c a d."
	Utterances of lengths 6 or more	31,466 (6%)	–	–	"Sunny to go fire it's sunny."
Utterances used	Utterances not shortened	293,407 (59%)	–	–	"Key open door."
	Utterances shortened	56,384 (11%)	Repetitions	20,209 (4%)	"Mommy mommy eat." - > "Mommy eat."
			Unintelligible markers	34,957 (7%)	"xxx yyy bee." - > "bee."
			Filler words	7582 (2%)	"ah open wide" - > "open wide"

Note. N (%): count of the utterances excluded or used and the percentage of those utterances out of all utterances before cleaning; Criterion-wise *N* (%): count of the utterances of each criterion and the percentage of those utterances out of all utterances before cleaning. Percentages under Criterion-wise *N* (%) do not add up to the percentage under *N* (%), because a given utterance may meet more than one criterion for exclusion or shortening. The utterances only containing illegal elements are the ones with only unintelligible markers or filler words, as shown in the example. The example for immediate imitations includes the child's (CHI) utterance and its immediately preceding adult's (MOT) utterance. The examples for shortening include both the original and the reduced utterances.

We focused on utterances of one, two, three, four, and five words, which account respectively for 35%, 20%, 17%, 12% and 7% of children's utterances (a total of 91%) and a total of 70% of adults' utterances. We extracted utterances of those lengths from the data of each child and adult, and constructed proportional distributions of length as a function of age.

2.1.2.1. Capturing early word combinations. We analyzed and compared the distributions for children in the *earlier* and the *later* groups, examining data aggregated across all children and data from each individual child. To ensure that each point for each group still consisted of data from at least four children, the age range for the *later* group was narrowed down to 20–35 months but only in this analysis (i.e., the analysis based on the two groups combined still focused on the age range between 14 and 43 months).

2.1.2.2. Tracking developmental changes. This analysis used combined data of both the *earlier* and the *later* groups for a larger sample size. We used linear segmented regression to quantify nonlinear change. The linear segmented regression model consists of two linear segments separated by one inflection point. Note that the two linear segments and an inflection point are a simplified description of a nonlinear shape rather than an indication of abrupt switch in behavior. We use the term "inflection point", also referred to as "breakpoint" or "changepoint" in other literature (Muggeo, 2016; Muggeo, Atkins, Gallop, and Dimidjian, 2014), to avoid a possible misinterpretation that the change at this point indicates any abrupt switch. Compared with other models that propose a nonlinear change, such as polynomial or exponential regression, segmented linear regression retains the advantages of linear regression such as simplicity and interpretability. It is also more in line with our theoretical interest because it can find the inflection point where the most apparent change occurs, allowing it to model not only developmental changes but also the steady, adult-like state that the child ultimately attains (e.g., Bloom, 1973; Brown, 1973; Butcher and Goldin-Meadow, 2000; Herr-Israel and McCune, 2011; Leopold, 1949; Oakes and Rakison, 2019). Alternative models, such as polynomial regression, will continue to change as a function of age and never converge on a steady state.

For each utterance length, we fitted a mixed-effects segmented regression (Muggeo, 2016; Muggeo et al., 2014) to the proportion of utterances by age to determine whether a single linear regression line or two linear regression lines with an inflection point between them fitted the data significantly better. The fixed effects are intercept and age for a single linear regression, and intercept, inflection point, age before the inflection point, and age after the inflection point for a segmented regression. The random effect is all the individual children from both the *earlier* and the *later* group with random intercepts and slopes. If the proportion of utterances of a certain length increases rapidly until an age (the inflection point) at which the slope changes significantly (e.g., stops increasing and then remains constant or flattens), a segmented regression will identify the inflection point and the slopes before and after the inflection point. If there is an inflection point, a model with two linear regression lines will fit the data significantly better than a model with a single linear regression line. The slopes of the two regression lines represent the developmental rates by age before and after the inflection point.

We use the inflection points estimated on each individual child's data from the mixed-effects regression model, and then compare the inflection points of each utterance length to determine where the most apparent developmental transitions occur.

2.2. Results

2.2.1. Capturing early word combinations

At the age of 14 months, 19% of utterances produced by children in the *earlier* group ranged from two to four words (an average of 5 multiword utterances per child with five children ($N = 5$) contributing to the data; examples are "back on" and "truckie went by"), and one-word utterances account for 81% (78 one-word utterances per child; $N = 5$). At 15 months, two-, three-, four, and five-word utterances accounted for, respectively, 15%, 6%, 2%, and 1% of utterances (69 multiword utterances per child; $N = 6$; examples are "key open door" and "a microphone mommy microphone"), increasing at 20 months to 22%, 12%, 7%, and 2% (311 multiword utterances per child; $N = 8$; examples are "keep that open for Lucy" and "see baby in there") (Fig. 3A). Most of the individual children followed the overall pattern (Fig. 4). For example, for six of the eight children with data at 15 months of age, two used utterances of lengths two-to-five, two used lengths two-to-four, and one used length three; one child used only single word utterances. When children produce shorter utterances, they produce longer utterances as well. Appendix B presents examples of multiword utterances produced by the *earlier* children before 20 months.

At 20 months, the children in the *later* group produced utterances with lengths two, three, four, and five words, respectively accounting for 34%, 14%, 5%, and 1% of the total (217 multiword utterances per child; $N = 4$). Their percentages were higher than those for the *earlier* children because even their early recordings occur at a later age (Fig. 3C). At 23 months, where more children's data were available, the comparable percentages were 25, 10, 4, and 1 (384 multiword utterances per child; $N = 15$). The proportion of longer utterances increased after that point. For the 15 individual children's data at 23 months, ten used two- to five-word utterances, four used two- to four-word utterances, and the remaining one used two- and three-word utterances. Adults from the *earlier* and the *later* groups were similar to each other, with relatively constant proportions of utterances of different lengths (Fig. 3B and D). At an older age, children's utterance length distributions became similar to those of adults'.

2.2.2. Tracking developmental changes

Significant inflection points occurred for each utterance length, as shown by segmented regressions (Table 2 and Fig. 5). For example, before 25.4 months of age, one-word utterances as a proportion of one- to five-word utterances decreased by 4.5% for each additional month ($b = -0.045$, 95% CI [-0.049, -0.040]). At 25.4 months (95% CI [24.0, 26.9]), the inflection point occurred: one-word utterances now decrease only 0.9% per month ($b = -0.009$, 95% CI [-0.021, 0.003]). The segmented regression fitted the data significantly better than did a linear regression (likelihood ratio test, $\chi^2(2) = 165.77$, $p < .001$). Each multiword utterance length first increased and then either decreased or increased at a much slower rate than before (Fig. 5). Fig. 6 shows developmental changes of one-word utterances for individual children. In sum, the proportions of early utterance lengths by age display significant nonlinear transitions.

Inflection point timing varied as a function of utterance length, as shown by repeated measures ANOVA using the inflection points estimated for each individual child's data ($F(2, 46) = 246.09$, $p < .001$, $\eta_p^2 = 0.92$); with Greenhouse-Geisser correction). Pairwise post-hoc comparisons with Bonferroni corrections found significant inflection point differences between utterances of two and three words ($t(23) = -6.50$, $p < .001$, $d = -1.33$, *95%* CI [-4.66, -1.79]), between three and four words ($t(23) = -10.70$, $p < .001$, $d = -2.19$, *95%* CI [-6.95, -4.00]), and between four and five words ($t(23) = -17.81$, $p < .001$, $d = -3.64$, *95%* CI [-4.20, -3.03]). The longer the multi-word utterance, the later the inflection point.

2.3. Discussion

We have two notable and novel findings. First, combinatorial speech appears very early – around 14–15 months. Previous research concentrates on older children, starting at around 19 months or later. Although the data are sparse at the earliest ages, at no period did children produce

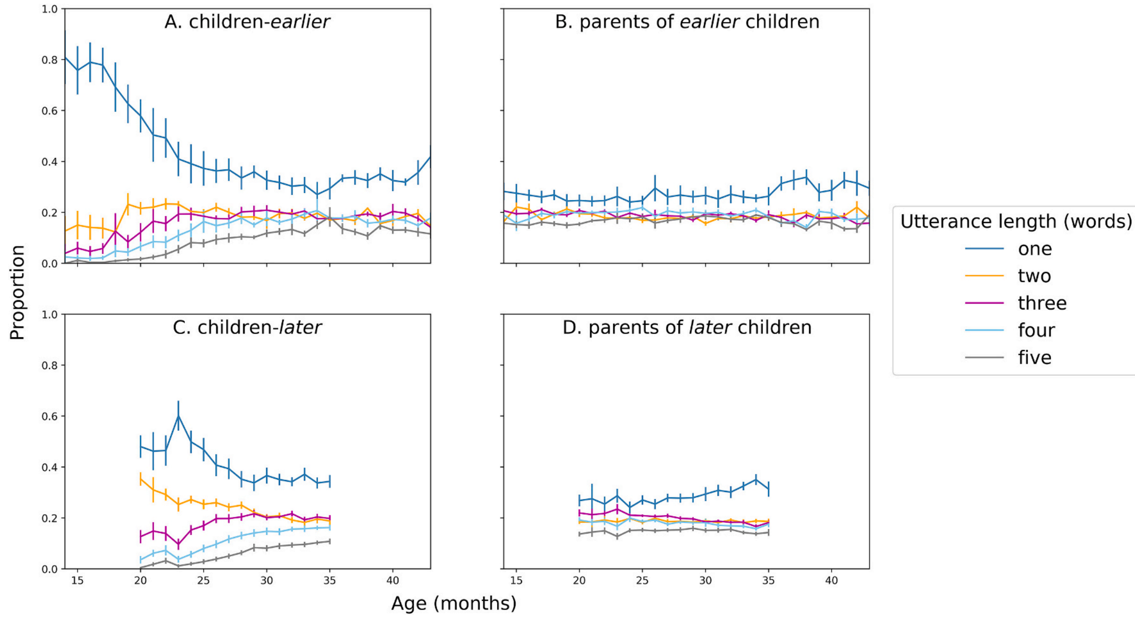

Fig. 3. Proportional distributions of utterances by length and age aggregated across children (A) and parents (B) in the *earlier* group, and children (C) and parents (D) in the *later* group. X-axis indicates children's age in months when being recorded. Y-axis indicates proportions of utterances of different lengths (i.e., one to five words) out of all utterances of one to five words. Error bars are standard errors of the means. Parents in each group are the parents of the children in that group. To ensure that each point for each group still consisted of data from at least four children, the age range for the *later* group was narrowed down to 20–35 months.

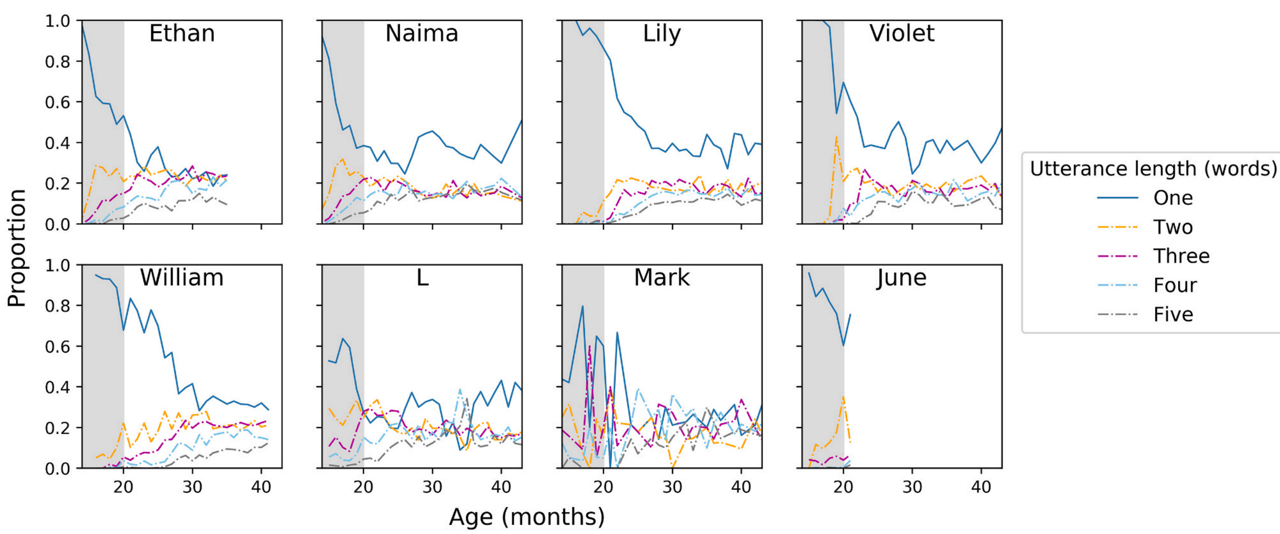

Fig. 4. Proportions of utterances by length and age of individual children in the *earlier* group. Y-axis indicates proportions of utterances of different lengths (i.e., one to five words) out of all utterances of one to five words. The grey area indicates the ages between 14 and 20 months.

only one- and two-word utterances. Instead, they produce two-, three-, four-, and even some five-word utterances. Children consistently used shorter utterances but tried longer ones, until the production of the next longer utterance became more stable. That pattern has not previously been described (Bloom, 1973; Butcher & Goldin-Meadow, 2000; Herr-Israel & McCune, 2011; Leopold, 1949; Scollon, 1976).

One might wonder if the eight children in the *earlier* group were precocious, rather than simply having been recorded earlier. What suggests that the children in the *earlier* group were of a piece with the children in the *later* group is that all of the children in the later group had numerous multiword utterances. That suggests that earlier recordings of the older children would have revealed the same early onset of multi-word productions that we see in the *earlier* children.

The second notable finding is that the development of utterances of different lengths is nonlinear. The development of one-word utterances is characterized by a rapid decrease and finally a plateau. Longer utterances increase (more gradually) and then plateau. The longer the utterance, the more time it takes to plateau. By hypothesis, that is because longer utterances require more resources than shorter utterances, and those resources both fluctuate during the short term and gradually increase over the long term. When children's utterance proportions plateau, we also observed a similarity of the utterance length distributions between children and parents. By hypothesis, that is showing that child utterance length development is influenced by parents' input. Our findings are consistent with the two resources we proposed - increasing resources and an input distribution of model lengths.

We use segmented regression as a simple tool for testing the nonlinear change and the average rates of change before and after the point of maximum change. Other characterizations are possible, for example, one that would use local minima or maxima for locating

Table 2
Mixed-effects segmented regressions for one- to five-word utterances.

Utterance length (words)	IP	CI_{IP}	*b* Before *b* After	CI_{b_Before} CI_{b_After}	χ^2
One	25.42	[23.96, 26.88]	−0.045– 0.009	[−0.049, −0.040] [−0.021, 0.003]	165.77***
Two	21.38	[19.86, 22.90]	0.012– 0.004	[0.009, 0.016] [−0.012, 0.003]	71.00***
Three	24.95	[22.81, 27.09]	0.0140 .003	[0.012, 0.016] [−0.002, 0.009]	90.26***
Four	30.91	[29.65, 32.17]	0.0120 .000	[0.011, 0.013] [−0.005, 0.005]	117.69***
Five	34.83	[33.40, 36.26]	0.008– 0.004	[0.008, 0.009] [−0.010,0.002]	111.37***

Notes. IP represents the inflection point. CI_{IP} represents the lower and upper limits of the 95% confidence interval of IP. *b* is the unstandardized regression coefficient, i.e., the change in proportion per month (slope) of utterances of the given length. For each utterance length, the *b* Before is estimated when the predictor is the ages before the inflection point, whereas the *b* After is estimated when the predictor is the ages after the inflection point. CI_{b_Before} and CI_{b_After} represent the 95% confidence intervals of *b* Before and *b* After, respectively. χ^2 indicates the chi-square of the difference between the segmented regression model and the linear model.

*** $p < .001$.

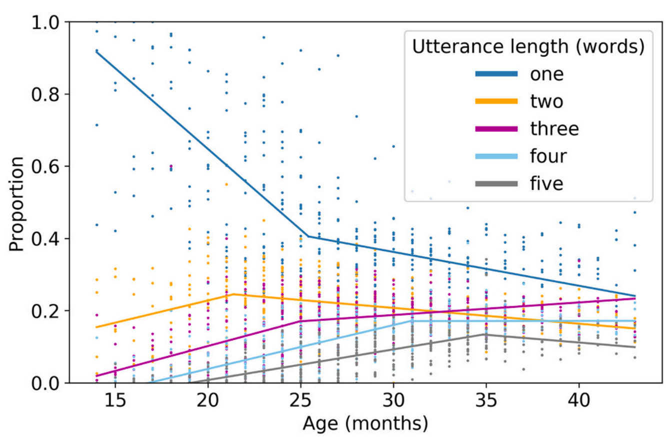

Fig. 5. Mixed effects segmented regressions of the proportions of one to five-word utterances by age. The dots represent individual children's data at each age point. Note that the two linear segments and an inflection point are a simplified description of a nonlinear shape rather than an indication of abrupt switch in behavior.

nonlinear shifts, but segmented regression is the simplest characterization. The current study analyzed utterances of each length separately while ignoring their interdependence – they all represent proportions of the same pool of utterances. However, we do not think that this will affect our findings that multiword utterances appear early and develop continuously and nonlinearly. The fact that they plateau at different time points rules out the possibility that the change in utterances of one length was solely an artifact caused by the change in utterances of another length; utterances of different lengths are interdependent in their proportions, but their development is at least partially independent.

The nonlinear changes in behavior where the utterance proportions plateau would seem to provide a challenge to a continuous underlying process. We also see age differences for the nonlinear changes of different utterance lengths. In Study 2 we ask whether an underlying continuous process is compatible with the nonlinearity of development. We also investigate formally what roles varying resources (e.g., cognitive, linguistic, or motor resources) and adult input play in shaping such a developmental trajectory of utterance length that we observe in Study 1. Our aim in Study 2 is to model the data of Study 1 to understand the underlying mechanisms.

3. Study 2

In Study 2 we model the underlying changes that may give rise to a) the early appearance of utterances longer than two words, b) the gradual increase of utterances by length, and c) the nonlinear transitions to more stable proportions of utterance length. We take a first step to provide a formal model, VIRTUAL, of the hypothesis that children's increases in utterance lengths are a function of a distribution of target lengths given by the input and continuously varying and developing resources which may involve multiple domains such as linguistic, cognitive, and biological constraints. We also test the theoretical and empirical adequacy of those two underlying assumptions.

3.1. Method

3.1.1. Data

We used the data from Study 1, including both the children's and their parents' utterances.

3.1.2. Model

We propose VIRTUAL for modeling the data. VIRTUAL is a probabilistic computational model that posits only two sources: continuously varying and developing cognitive or linguistic resources and a probability distribution of target utterance lengths obtained from the input. The idea underlying the first source is that the probability of success in producing a given length depends on the level of available resources (R). R varies from moment to moment and is normally distributed around a mean that increases as the child's age increases, as illustrated in Fig. 7. For an utterance to be produced, R must be sufficient to produce at least one word; larger values of R make longer utterances possible.

The idea underlying the second source is that each time a child generates an utterance, the child "intends" to produce an utterance of some target length (T). The distribution of target lengths is determined by the distribution of lengths in the adult input, shown in Fig. 8A. For consistency with Study 1, we set the range of lengths from 1 to 5. Note that the parents' length distribution is roughly constant across a wide range of ages, as shown in Fig. 8B. Since resources increase with age, longer utterance lengths will also increase until the child's distribution matches the parent's.

We propose that the probability of the child producing an utterance of length N at each timestep, $P(N)$, is as shown in Eq. 1. There are two ways in which this can happen: (a) The target utterance length is equal to N, $P(T = N)$, and the resource level is adequate to support producing N words, $P(R \geq N)$; (b) The target length is longer than N, $P(T > N)$, but the maximum length that R can support is N, $P(R_{MAX} = N)$, in which case the child will reduce the utterance length from T to N. $P(R_{MAX} = N)$ equates to $P(N \leq R < N + 1)$ in the probability calculation, representing resources that can maximally support length N and are not adequate for length $N + 1$. The denominator in the equation, Z, is a normalizing constant to guarantee that the $P(N)$ values are probabilities. Z is equal to the sum of the numerators in $P(N)$, for $N = 1$ to 5.

$$P(N) = \left(\left[P(T = N)^{*}P(R \geq N)\right] + \left[P(T > N)^{*}P\left(R_{MAX} = N\right)\right]\right)^{*}\left(\frac{1}{Z}\right) \tag{1}$$

3.1.3. Procedure

Using Eq. 1, we calculate the probabilities of producing utterances of lengths one to five words at each age point (month). Similar to Study 1, we construct proportional distributions of length as a function of age. Simple segmented regressions (Muggeo, 2003) are fitted to the distributions by age to locate inflection points and to measure the slopes of

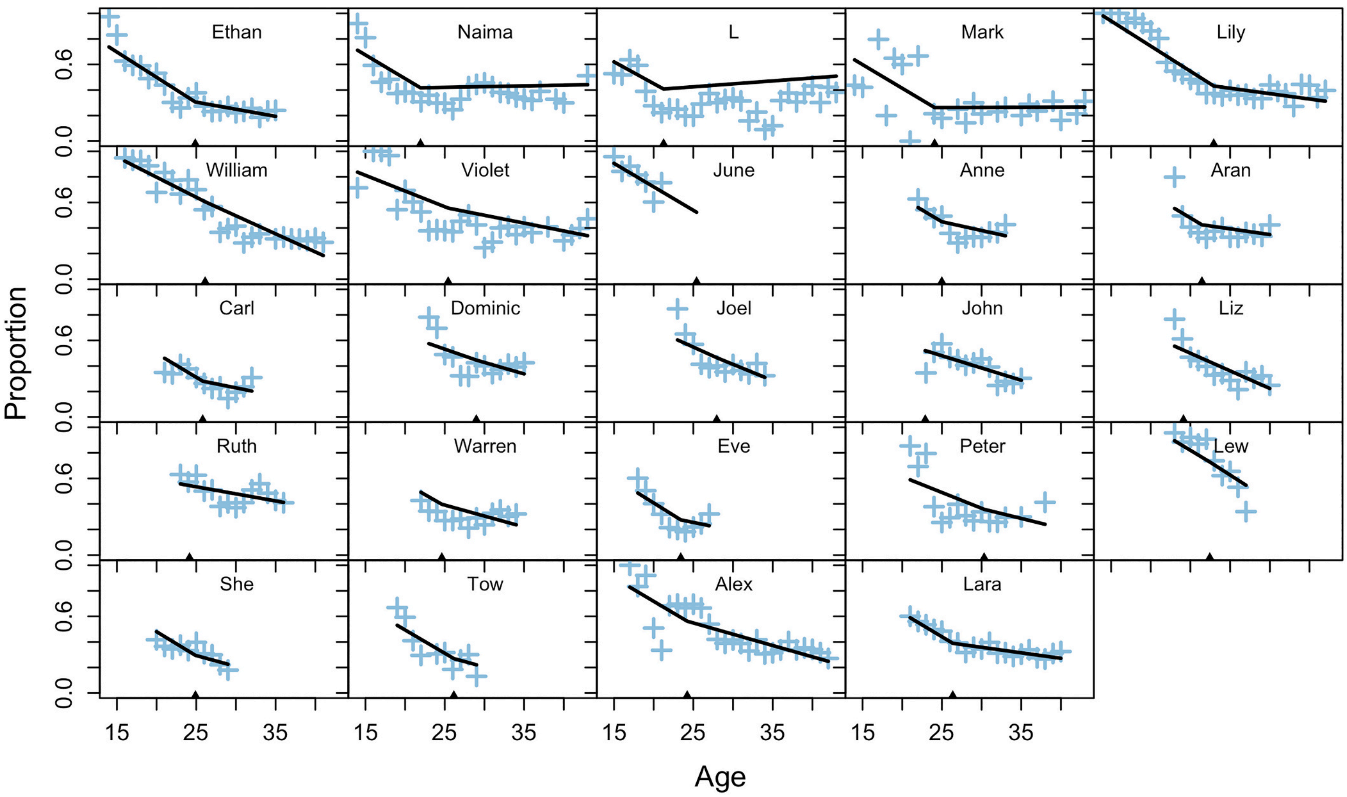

Fig. 6. Individual children's development of one-word utterances, predicted by the mixed-effects segmented regression model. The triangle on the x axis denotes the segmented regression inflection point for each child.

the regression lines before and after the inflection points. Finally, we compare the model predictions to the empirical results from Study 1. Model predictions of proportions <0.001 (one per thousand utterances) are excluded from the analysis due to the unreliability of empirical evidence for such rare events. As a result, agepoints of five-word utterances in the model output start from 17 months.

The standard deviation of the R distribution is fixed at 1.5. The initial mean value of R at 14 months (the earliest time point in Study 1) is set to 0 and is increased by 0.35 for each successive month. Those parameters are tuned empirically to provide the best overall fit of the data (i.e., the number of predicted coefficients that fall within the 95% confidence interval of the empirical results). The parameter values do not affect the fundamental framework of the model, and they are open to fine tuning (see Discussion for more details). We show in Appendix C that the change in parameters will only produce quantitative rather than qualitative differences in VIRTUALs predictions; the greater the initial mean, the lower the proportion of one-word utterances at 14 months (Fig.C.1). The smaller the SD, the sharper the change in utterances of different lengths (Fig.C.2). The greater the incremental value, the earlier the development reaches a stable, adult-like state (Fig.C.3).

We test the model against the children's empirical distribution. We also separately test the adequacy of the two sources, target lengths and increasing resources, of VIRTUAL and the interaction between them.

3.2. Results

Fig. 9 compares the empirical pattern from Study 1 (A and C) to the probabilistic model (B and D). The model qualitatively captured the four dominant features of the empirical pattern. The model simulated 1) the early dominance of utterances of length 1, and 2) the early appearance of utterances of lengths 2–5. 3) Each length has an inflection point, and 4) longer utterances have later inflection points.

Table 3 quantitatively compares the segmented regressions of the model and the empirical data. Most of the inflection points and slopes predicted by VIRTUAL fell within the 95% confidence interval of the empirical results. For example, one-word utterances as a proportion of one-to five-word utterances first decreased significantly by age, $b = -0.036$, reached an inflection point at 26.6 months, and remained almost constant after the inflection point, $b = -0.003$. The predicted inflection point and the slope before the inflection point fell within the 95% confidence intervals of those from the empirical data for the inflection point (25.4, 95% CI [24.0, 26.9]) and the slope after the inflection point ($b = -0.009$, 95% CI [-0.021, 0.003]).

3.3. Testing VIRTUAL's assumptions

We test the adequacy of the two sources – target lengths and increasing resources – of VIRTUAL and the interaction between them.

3.3.1. Target lengths

In VIRTUAL, the target distribution is determined by the distribution of lengths in the input. To test the necessity to posit a target distribution, we removed the target length terms from VIRTUAL so that the probability of producing an utterance of length N is based solely on the probability that R_{MAX} is equal to N, as in Fig. 10A.

3.3.1.1. Methods. We applied the same proportional distribution analysis that we used for VIRTUAL, with the same parameters of R for the initial mean, standard deviation, and increment size. Similar to VIRTUAL, parameter values for the no-target-distribution model were tuned empirically.

3.3.1.2. Results. The no-target-distribution model (Fig. 10B) shows that as age increases, each length becomes dominant for a period of time and then declines in prominence only to be replaced by the next (higher) length. This is strikingly different from the empirical pattern of

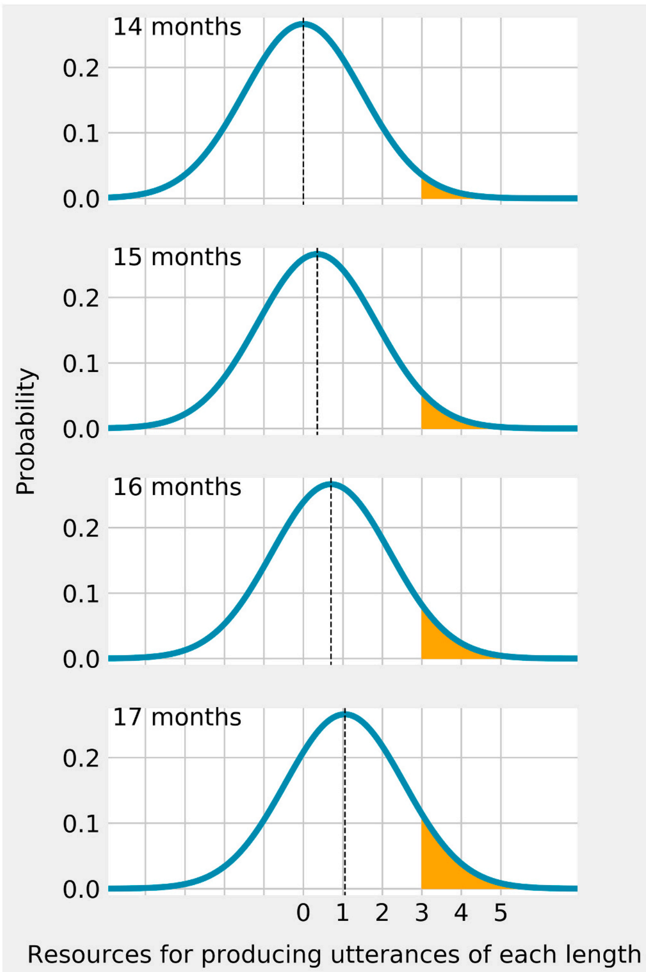

Fig. 7. The R distribution mean increases as age increases, shown here in timesteps of one month. At each timestep, the R distribution provides the probability of having resources for supporting the production of the utterances of each length. The shaded area illustrates the probability of having resources for supporting the production of three-word or longer utterances, which increases by age.

children's length distributions (see Fig. 10C). Parameter tuning does not qualitatively change its predictions. This test thus confirms the necessity for VIRTUAL to include a target utterance length distribution in explaining children's utterance length development.

3.3.2. *Resource distribution*

To test the source of increasing developmental resources, we could not compare the model without the resource distribution to the children's data since the development of speech production is known to be at least partially driven by the growth of resources such as working memory, lexical knowledge, or articulatory capacity (Moore and Maassen, 2004; Newbury et al., 2016; Nóro and Mota, 2019). Instead, we test the plausibility of the idea that resources are continuously varying, normally distributed, and increasing. Since R must be sufficient to produce at least one word, $P(R \geq 1)$, for any utterances to be produced, the change in R predicts the rate of increase in the total number of utterances children will produce. Therefore, we counted the total number of utterances that children produce and compared it to the model's prediction that the probability of an utterance of any length, $P(R \geq 1)$, first rapidly increases and then levels off. We expect to see a correspondence between the change in $P(R \geq 1)$, an increase and then a leveling off, and the development of the total number of utterances that children produce.

3.3.2.1. Methods. We used the empirical data to analyze the mean frequency of all the utterances by age. Since utterance frequency is an unstandardized measurement and is susceptible to noise, we focus on the Providence data: that database has sessions of fixed duration and contains many early recordings before 20 months. To ensure that each age point consists of data from at least four children in the corpus, we focus on the range between 14 and 41 months.

Data were standardized to make the model prediction and the empirical data comparable. For the model prediction, Z scores were calculated at each timestep based on the mean and standard deviation across the whole age range. For the empirical data, Z scores were separately calculated for each child based on the child's mean and standard deviation across the age range. To quantify the change, we again applied segmented regressions for the empirical data (Muggeo, 2016; Muggeo et al., 2014).

3.3.2.2. Results. The change in the frequency of utterances predicted by VIRTUAL is similar to the development of the children's utterance production (Fig. 11). Segmented regressions show the inflection point of VIRTUAL's predictions (22.6) to be within the 95% confidence interval ([16.1, 27.4]) of the empirical results. The results are consistent with the

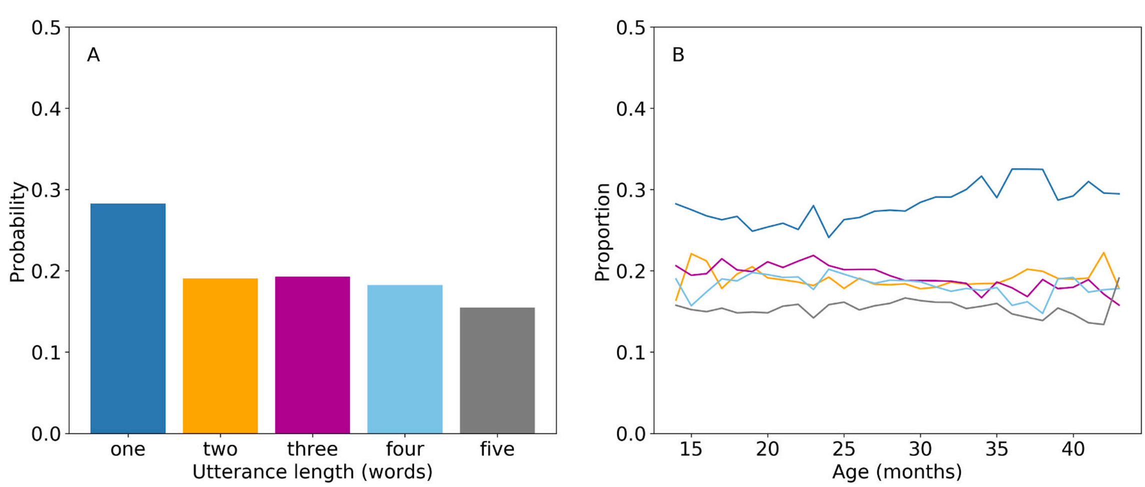

Fig. 8. (A) distribution of "target" utterance lengths, and (B) plot of parents' utterance lengths by child age. The distribution of "target" utterance lengths (A) consists of the mean proportions of the utterance lengths in the parents' distribution (B), which remains relatively stable over time.

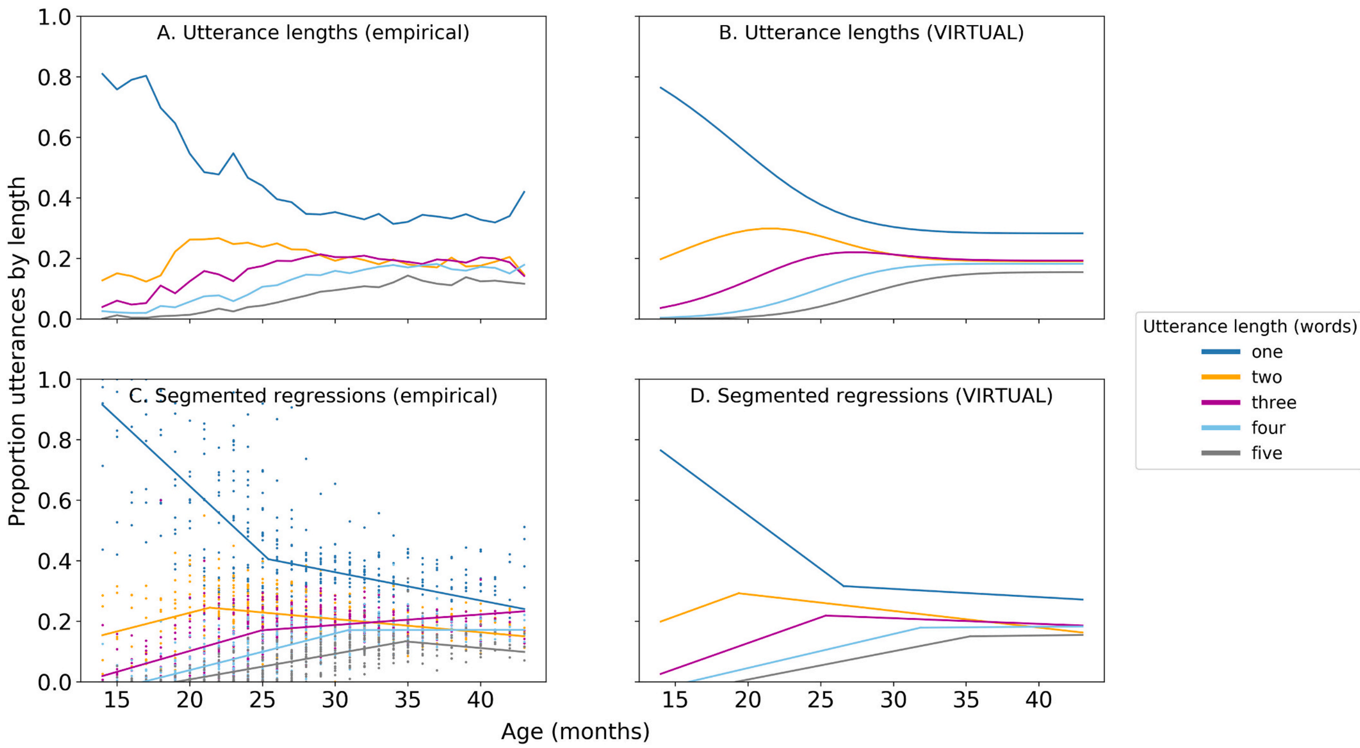

Fig. 9. A) Empirical and B) VIRTUAL distributions of utterance lengths, and segmented regressions fitted to the C) empirical and D) VIRTUAL distributions.

Table 3
Segmented regression results for the model predictions and for the empirical data.

Utterance length (words)	Model predictions		Empirical results	
	IP (months)	*b* Before IP *b* After IP	IP (months)	*b* Before IP *b* After IP
One	26.58[a]	−0.036 −0.003[a]	25.42	−0.045 −0.009
Two	19.36	0.017 −0.005[a]	21.38	0.012 −0.004
Three	25.35[a]	0.017 −0.002[a]	24.95	0.0140 .003
Four	31.86[a]	0.011[a] 0.000[a]	30.91	0.0120 .000
Five	35.28[a]	0.009[a] 0.001[a]	34.83	0.008 −0.004

Notes. IP represents the inflection point. *b* is the unstandardized regression weight, i.e., the change in proportion (slope) of utterances of the given length per month. For each utterance length, the *b* Before is estimated when the predictor is the ages before the inflection point, whereas the *b* After is estimated when the predictor is the ages after the inflection point.

[a] Parameter falls within the 95% confidence interval of the corresponding empirical results.

idea that resources are continuously varying, normally distributed, and increasing.

3.3.3. *Interaction between the two sources*

VIRTUAL posits an interaction between the two sources: a target utterance is shortened in production when the available resources that the child has are not adequate to support its length. To examine this parameter, we created a no-reduction model. That is, the target utterances are either successfully produced when the resource level is adequate, or no utterance is produced when resources are not adequate. This no-reduction model omits the length-reduction term (the second term) in the numerator in Eq. 1 (See Fig. 12).

The equations of the two models differ in their numerators, which are analogous to the count of one-to-five-word utterances in the empirical data. In the empirical data, the proportion of N-word utterances is the frequency of N-word utterances divided by the total frequency of all the one- to five-word utterances. Likewise, the numerator in the formula for *P(N)* in the two models is divided by the probability that an utterance of any length between one and five will be produced (i.e., the denominator).

With the length reduction term in its equation, VIRTUAL predicts an increase in the frequency of short utterances followed by a slight decrease in frequency before a leveling off, whereas the no-reduction model predicts monotonically increasing frequencies (Fig. 12B). This is because, in VIRTUAL, the reduction mechanism assigns extra utterances to the short lengths at early ages, but as children increase their resources for producing longer utterances, the number of short utterances declines since the children do not need to perform as many reductions as before. In contrast, the numerator of the no-reduction model relies only on the available developmental resources, which can only increase and do not decrease.

3.3.3.1. Methods. To test which of the two models fits the data better, we constructed utterance frequency distributions of the empirical data and of the numerators of the two models. Similar to the test of the resource distribution, we focused on children's data from the Providence corpus with the same age range, and the data were standardized to make the model predictions and the empirical data comparable. Since the reduction process primarily affects the short utterances, we focused on one-and-two-word utterances. Similar to VIRTUAL, parameter values for the no-reduction models were tuned empirically.

3.3.3.2. Results. As shown in Fig. 12C, all children's data, except for William's, clearly show an increase, a slight decrease, and finally a leveling off. Such a pattern in the empirical data is consistent with VIRTUAL's prediction as opposed to that of the no-reduction model. Parameter tuning does not qualitatively change the predictions of the no-reduction model. This test thus confirms the necessity for VIRTUAL to assume a reduction of the target utterance in production when the

A. The VIRTUAL model

$$P(N) = ([P(T = N) * P\ (R \geq N)]\ * [P(T > N)\ *\ P(R_{MAX} = N)])\ *\ \left(\frac{1}{Z}\right)$$

The no-target-distribution model

$$P(N) = ([P(R_{MAX} = N)])\ *\ \left(\frac{1}{Z}\right)$$

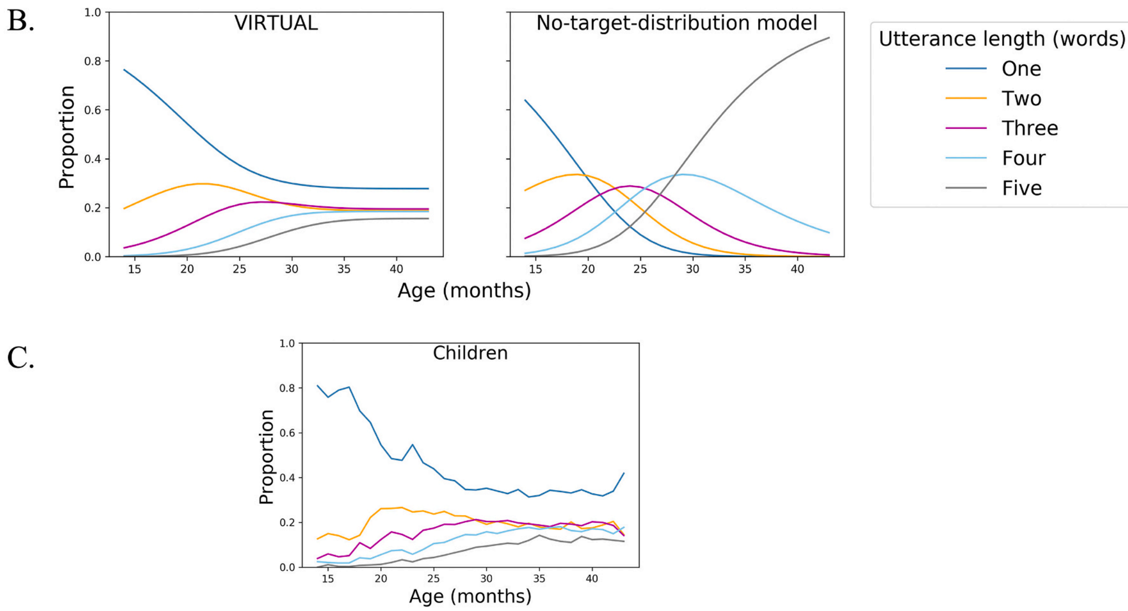

Fig. 10. Testing the distribution of target utterance lengths. A) The formulae of the VIRTUAL model and the no-target-distribution model, where the terms in blue represent the parts relevant to the target-distribution assumption; B) The utterance length distributions predicted by VIRTUAL and the no-target-distribution model. C) The children's empirical utterance length distribution by age. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

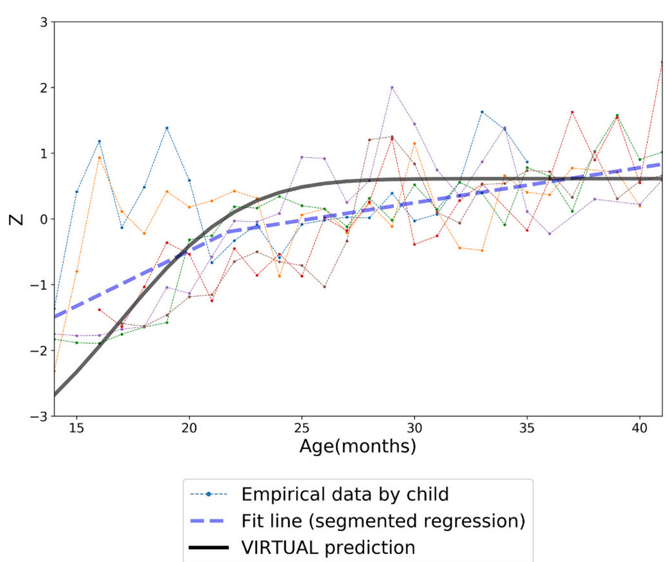

Fig. 11. Testing the developmental resources. Empirical data by child: standardized average frequency of utterances as a function of age, based on the empirical data. Fit line: standardized utterance frequency predicted by mixed-effects segmented regression. VIRTUAL prediction: standardized probability of having resources for producing utterances of any length.

available resources that the child has are not adequate to support its length.

3.4. Discussion

VIRTUAL, our probabilistic model of the developmental increase in the length of children's utterances, simulates most of the qualitative characteristics of the empirical data. It correctly shows the early dominance of one-word utterances and the simultaneous presence of multi-word utterances. It shows nonlinear changes in the development of utterance length, with longer utterances taking more time to plateau. Most of the quantitative comparisons fall within the confidence intervals of the empirical data. We show that the assumptions in the model – a) continuously varying and growing resources, b) a target distribution, and c) the interaction between the two sources – are necessary to account for the data. The alternative models without the target distribution of utterance lengths or the length-reduction process fail to predict the empirical patterns. The varying and increasing resources of VIRTUAL align well with a central feature of the empirical data it is analogous to, the change of utterance frequency with age.

VIRTUAL is not perfect. Some of the predicted inflection points and slopes in the segmented regression analyses do not fit the confidence intervals of the empirical data. We might improve the fit by changing or adding model parameters, such as providing a different shape of the resource distribution or dynamically varying developmental rates, but that improvement would complicate the model. Since VIRTUAL allows a fine-tuning of the parameter values, future studies could tune the

A. The VIRTUAL model

$$P(N) = ([P(T = N) * P\,(R \geq N)]\ * \boldsymbol{[P(T > N)\ *\ P(R_{MAX} = N)]})\ * \left(\frac{1}{z}\right)$$

The no-reduction model

$$P(N) = ([P(T = N) * P\,(R \geq N)]) * \left(\frac{1}{Z}\right)$$

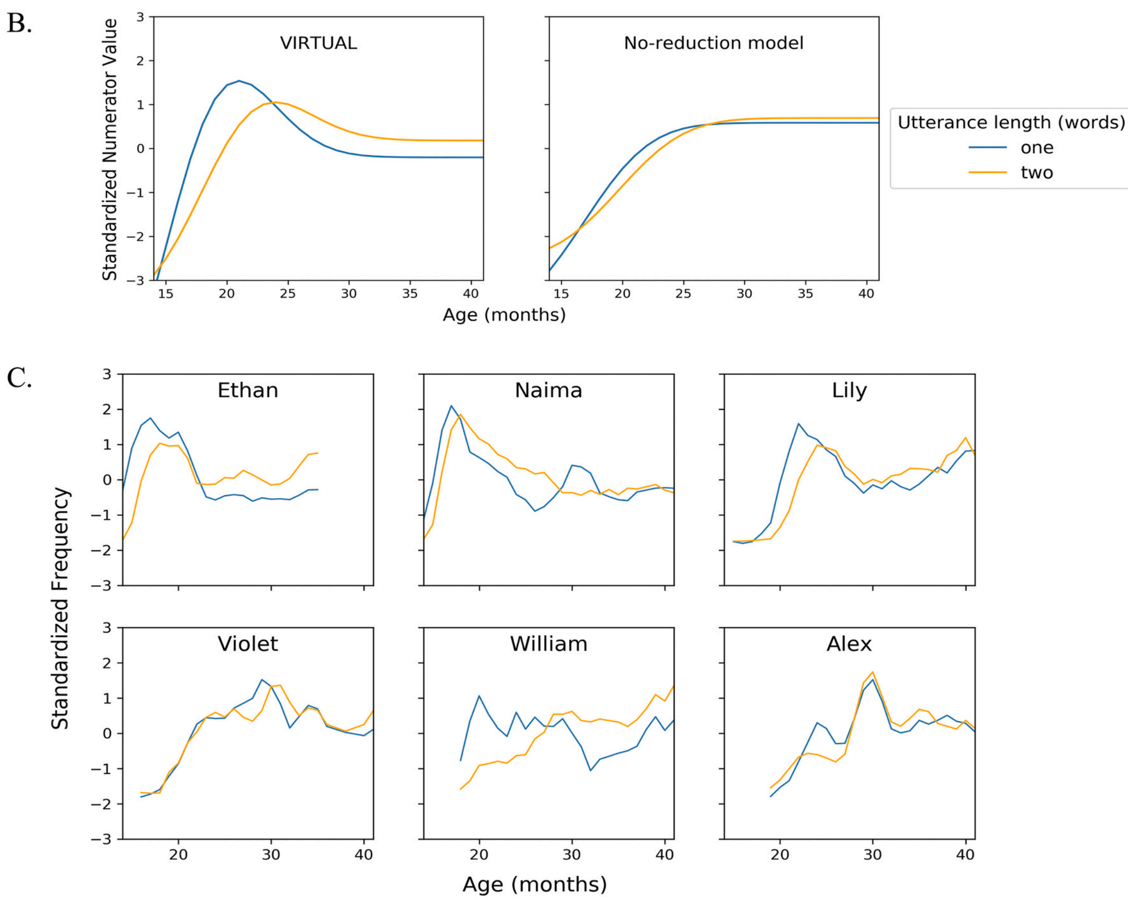

Fig. 12. Testing the reduction process. A) The formulae of VIRTUAL and the no-reduction model, where the term in blue represents the length reduction process; B) The standardized numerator values by length and age, predicted by VIRTUAL and the no-reduction model. The standardized numerator value is analogous to the standardized utterance frequency in the empirical data; C) Standardized average frequency of utterances by length and age, based on the empirical data of the individual children from the Providence corpus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parameters to better understand language development of different populations. For example, different children may have different starting points, developmental rates, and variance of the resource distribution. We show that VIRTUAL's success owes to its assumptions rather than to artifacts of tailoring parameters; changing parameter values results in quantitative rather than qualitative differences, and tailoring an alternative model that lacks one of VIRTUAL's components does not lead to success, no matter how the parameter values are tweaked. That gets to the second point, which is that VIRTUAL is falsifiable despite its flexibility in parameter tuning. Where the data's qualitative patterns are different from VIRTUAL's predictions, VIRTUAL will fail. Indeed, there are individual children (e.g., "Lew" as shown in Fig. 6) whose data and overall pattern look very different from other children's. Such a disparity may be because of sampling bias or language acquisition factors that VIRTUAL does not account for. Future studies should validate VIRTUAL's predictive power and gain more insights on the commonality and variability among individuals in language development.

The model is silent on the nature of the resources that are involved in early utterance production. We view utterance length as a general index of language development, and a cumulative consequence of a dynamic interaction between many different resources. In early long utterances, in particular, there may be many processes occurring simultaneously. (See, for example, in Appendix B, "lock door key lock door" at 18 months). An increase in length can simultaneously reflect syntactic development (e.g., ability to produce subject-verb-object sequences, such as in "key lock door"), pragmatics (e.g., topicalization of a verb phrase, perhaps for emphasis), and biological capacity (e.g., breath and motor control that allows production of more words in a single utterance). Only experimental work can detail the nature of the cognitive and linguistic resources that are likely to constrain production during language acquisition (Berk & Lillo-Martin, 2012; Moore & Maassen, 2004; Valian, Hoeffner, & Aubry, 1996). Further, we cannot rule out the possibility that resources are not normally distributed: there are unlimited possible alternative distributions of the developmental resources. We show, however, that the assumption of a normal distribution of resources yields a correspondence between the model and the empirical data. Such a benchmark could benefit future studies on the developmental of cognitive and linguistic resources for speech production.

Despite its limitations, VIRTUAL is a psychologically plausible model linking theories of developmental cascades with behavioral change. First, it instantiates the correlation seen in experimental work between a growth in working memory and vocabulary size and the development of speech production (e.g., Newbury et al., 2016; Nóro and Mota, 2019). Second, it represents one aspect of that sensitivity in the component specifying the perception of spoken utterance lengths in adult language, which extends the previous findings that infants show sensitivity to their native language at prosodic, lexical, and syntactic levels (Bergelson and Swingley, 2012; Christophe et al., 1994; Gervain et al., 2008; Shi et al., 1999). Future studies could investigate why the target distribution of utterance lengths that mirror parents' language contributes to child utterance length development and where it comes from. Learning or imitating adult utterances could lead to adult-like utterance length distributions. On the other hand, the target utterance length distributions may reflect some intrinsic linguistic properties such as communication efficiency, when a short utterance suffices for a simple event, and intelligibility, when more words are needed for a more complex event (see Gibson et al., 2019 for a review of communicative efficiency).

Third, VIRTUAL attributes the disparity between what children might intend to say and what they do say to restricted developmental resources, which corresponds to claims that children's comprehension is broader and deeper than their production, and that their productions are limited by resources (Goldin-Meadow et al., 1976; Santelmann and Jusczyk, 1998; Shi et al., 1999; Shipley et al., 1969; Valian et al., 1996; Valian and Aubry, 2005).

4. General discussion

The present research provides the first systematic documentation of children's very early combinatorial speech, and offers a simple, but novel, probabilistic account of the developmental process underlying the behavioral changes in utterance length. Our empirical data demonstrate previously unnoticed developmental patterns. Our computational model captures the dominant features of the empirical data and highlights possible mechanisms that drive utterance length development. The findings support a cascade account of development: the change in utterances of various lengths resembles a continuous and simultaneous process more than a sequence of discrete stages (Iverson, 2021; Oakes and Rakison, 2019).

Taken together, the findings from the empirical data and the computational modeling support three main points. First, as soon as children produce two-word utterances, they also produce longer utterances. Although the development of utterances of different lengths is not linear, it also cannot be neatly divided into a one-word, a two-word, and a multi-word stage as many previous studies have done (e.g., Brown, 1973; Herr-Israel and McCune, 2011). All lengths are present simultaneously.

VIRTUAL, which incorporates a continuously varying and increasing resource distribution, successfully models the coexistence of shorter and longer utterances. The shape of the resource distribution makes the probability of producing longer utterances always present to some degree, and the increment of the mean increases the probabilities as development proceeds. That explains why there is no period where children exclusively produce short utterances: resource availability during language production is probabilistic rather than all-or-none.

Second, the development of utterance length is nonlinear. For lengths from two to five words, the proportions in the child's output increase and then plateau, with longer utterances showing later inflection points. VIRTUAL models the nonlinear change without relying on underlying discrete qualitative changes. Rather, changes arise from an interaction between continuously varying, developmentally increasing resources and a probability distribution of target utterance lengths in discrete units (words).

The growth of multiword utterances reflects the effect of increasing resources on a resource-limited process: with increasing resources, the proportion of multiword utterances rises. A shorter utterance length will have its inflection point earlier in development because its resource requirement is satisfied sooner than that of a longer utterance. The plateaus, on the other hand, reflect a data-limited process; increasing the resource does not affect the proportion because it is limited by the target distribution.

Third, VIRTUAL represents an interaction of two sources, increasing developmental resources and adult input, over time. With adequate resources, the production of an utterance of a given length is primarily determined by the distribution of utterance lengths in the parent's input. When resources are low, as they are early in development, children have to exclude some of the elements from the utterance that they would otherwise have produced, yielding an ungrammatical and shorter utterance. This account is consistent with the evidence cited above (e.g., Shi et al., 1999; Valian and Aubry, 2005) that children know more than they say.

Our findings and model can be used in future research to probe more aspects of language development. For example, the finding that multiword utterances of different lengths appear early and increase together may provide insights on child syntactic development. Unlike suggestions that syntactic acquisition begins around 24 months, plus or minus two (e.g., Radford, 1990), our data show that one hallmark of syntax – word combinations (e.g., "key open door" and "blue car broken down") – occurs considerably earlier and more dynamically. Moreover, VIRTUAL suggests that children know more than they say, which implies that syntactic skills develop faster and are more advanced than previously thought. We propose more attention to the syntactic competence behind the very first word combinations. On the other hand, because utterance length is an indirect measure of syntactic development, similar MLUs can reflect different degrees and types of development (Leonard and Finneran, 2003; Rollins et al., 1996). Future study should chart the appearance of various kinds of constituents and syntactic structures as length increases.

According to VIRTUAL, longer utterances are reduced to shorter ones due to limited developmental resources. Future studies should examine which elements are most likely to be omitted in this process, and why. In addition, the parameters in VIRTUAL might be tuned to reflect differences between typical and atypical development, and to model individual differences in language acquisition.

CRediT authorship contribution statement

Qihui Xu: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Martin Chodorow:** Conceptualization, Methodology, Investigation, Supervision, Writing - review & editing. **Virginia Valian:** Conceptualization, Investigation, Supervision, Project administration, Writing - review & editing.

Data availability

Data and codes are available on https://osf.io/69yc4/

Acknowledgments

We thank Kyle Gorman, Sandeep Prasada, Charles Yang, and the members of the Language Acquisition Research Center at Hunter College, City University of New York for valuable comments and feedback. We are very grateful to our two reviewers.

Appendix A

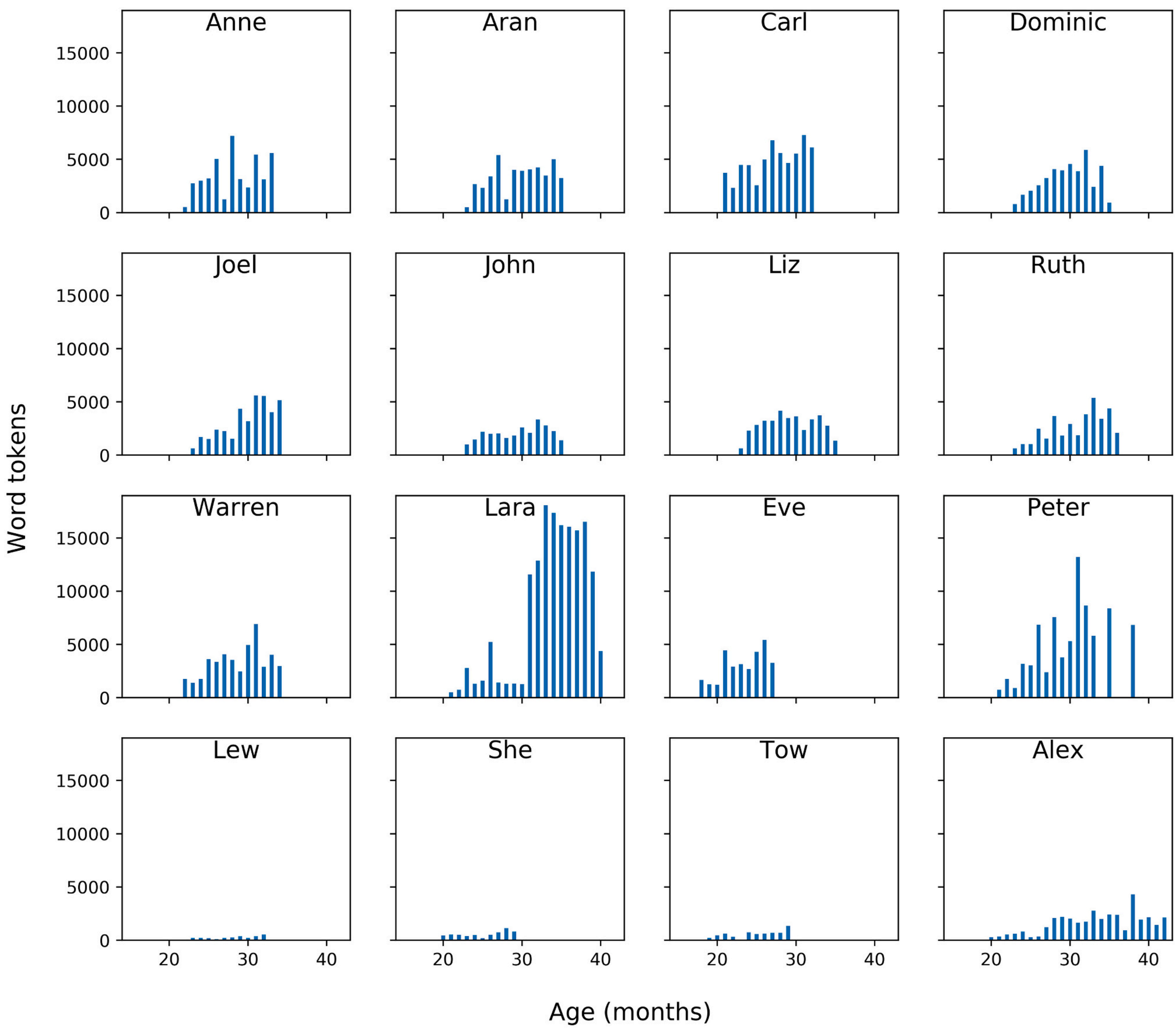

Fig. A.1. Number of word tokens by age for children in the *later* group. Note that the range of the y-axis is greater than in the graphs of the *earlier* group due to the particularly large amount of data for some of these children (e.g., Lara and Peter).

Appendix B

Table B.1
Examples of multiword utterances produced by the earlier children before 20 months. Sampling was balanced by age such that each age contains 10 utterances.

Age	Child	Utterances
14	Ethan	"Hi tape."
		"Back on."
		"A byebye."
	Naima	"Truckie went by."
		"Baby clothes."
		"Water mommy."
	Violet	"Dada nana."
	Mark	"Nice kitty."
		"Hi Tee."
		"That my bottle."
15	Ethan	"Open shut."
		"People here."
		"Gate down."
	Naima	"Water dirty water."
		"Meow mommy."
		"Catch mommy."
	L	"Deedee Joann."
		"Hi baby."
	Mark	"Say puppy."
	June	"It's a party."
16	Ethan	"Eat cereal."
		"Big bird."
		"That hole."
	Naima	"Doggy yogurt."
		"Truckie noisy truckie."
		"Mommy yogurt."
	William	"A worm."
	L	"Out go."
	June	"It's a ball."
		"It's a duck."
17	Naima	"Somebody hiding somebody."
		"Book yankee doodle book."
	Lily	"Bumble bee."
		"Pooh Bear."
	William	"Go quack."
		"This book."
	L	"I want."
		"Ride it round here."
	Mark	"I wanna get down."
	June	"It's a shoe."
18	Ethan	"Lock door key lock door."
		"What is that car motorcycle."
	Naima	"Sweeping them."
	Lily	"Poohbie Pooh."
	Violet	"Daddy's guitar."
	William	"Alright is Sarah what."
	L	"Mama bit."
		"Den then den then swimming."
	Mark	"Me a uppie."
	June	"And then it's time for."
19	Ethan	"Not for me."
	Naima	"Mom blowing on it."
		"Mommy having a sit."
	Lily	"I've got your tummy tum."
	Violet	"Piggy and me."
	William	"Have a day."
	L	"Bike Sue."
		"Yeah dump jump Baura Laura."
	Mark	"Pee_pee say pee_pee."
	June	"There's Potatohead's nose."

Appendix C

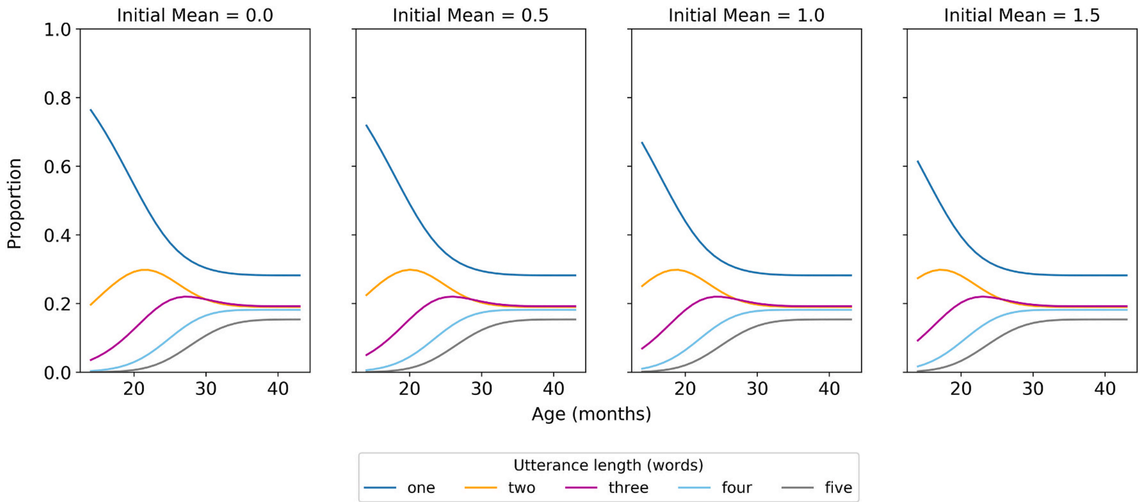

Fig. C.1. Varying Mean, SD = 1.5, Incremental value = 0.35. The greater the initial mean, the lower the proportion of one-word utterances at 14 months.

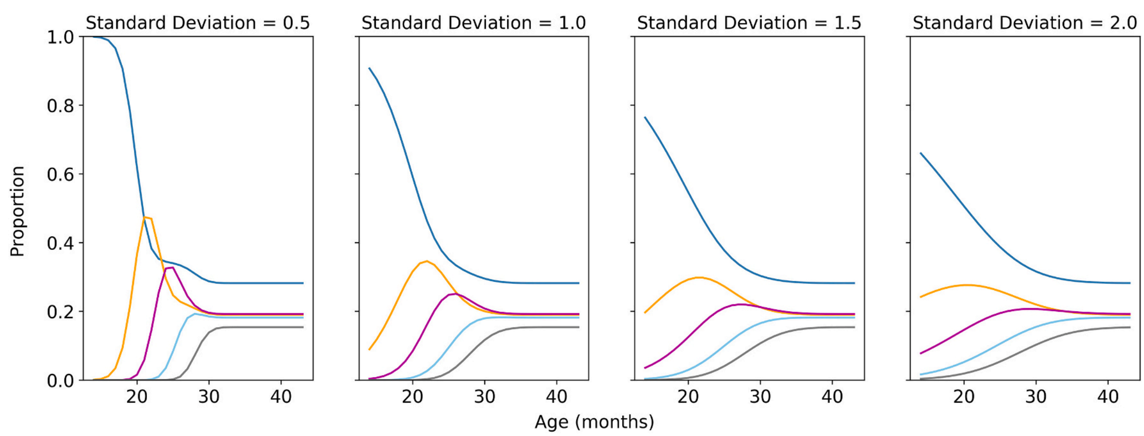

Fig. C.2. Mean = 0, Varying SD, Incremental value = 0.35. The smaller the SD, the sharper the change of utterances of different lengths.

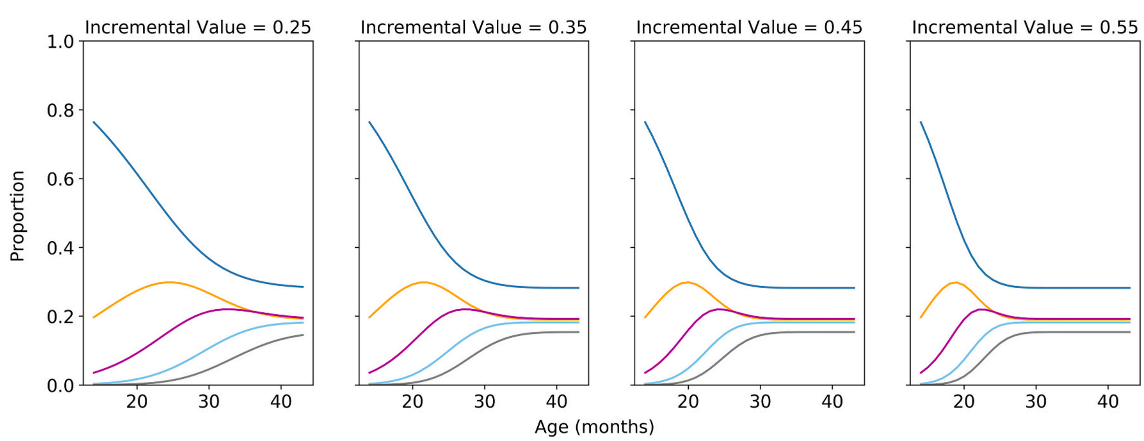

Fig. C.3. Mean = 0, SD = 1.5, Varying Incremental Value. The greater the incremental value, the earlier the development reaches a stable, adult-like state.

References

Ambridge, B., Kidd, E., Rowland, C., & Theakston, A. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language, 42*, 239–273.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258.

Berk, S., & Lillo-Martin, D. (2012). The two-word stage: Motivated by linguistic or cognitive constraints? *Cognitive Psychology, 65*(1), 118–140.

Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language, 20*(1), 139–152.

Bloom, L. (1970). *Language Development: Form and Function in Emerging Grammars*. Cambridge, MA: MIT Press.

Bloom, L. (1973). *One Word at a Time: The Use of Single Word Utterances Before Syntax*. The Netherlands: Mouton: The Hague.
Bowerman, M. (1973). Structural relationships in children's utterances: Syntactic or semantic? *Cognitive Development and Acquisition of Language*, 197–213.
Braine, M. D., & Bowerman, M. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, 1–104.
Braunwald, S. R. (1971). Mother-child communication: The function of maternal-language input. *Word, 27*(1–3), 28–50.
Brown, R. (1973/2013). *A First Language: The Early Stages*. Harvard University Press.
Butcher, C., & Goldin-Meadow, S. (2000). Gesture and the transition from one- to two-word speech: When hand and mouth come together. In D. McNeill (Ed.), *Language and Gesture (Language Culture and Cognition)* (pp. 235–258). Cambridge: Cambridge University Press.
Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *The Journal of the Acoustical Society of America, 95*(3), 1570–1580.
Demuth, K., & McCullough, E. (2009). The prosodic (re)organization of children's early English articles. *Journal of Child Language, 36*, 173–200.
Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J., & Bates, E. (2005). A crosslinguistic study of the relationship between grammar and lexical development. *Journal of Child Language, 32*(4), 759–786.
Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive Psychology, 57*(1), 56–74.
Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23*(5), 389–407.
Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. *Cognition, 4*(2), 189–202.
Goldin-Meadow, S., & Singer, M. A. (2003). From children's hands to adults' ears: gesture's role in the learning process. *Developmental Psychology, 39*(3), 509.
Herr-Israel, E., & McCune, L. (2011). Successive single-word utterances and use of conversational input: A pre-syntactic route to multiword utterances. *Journal of Child Language, 38*(1), 166–180.
Higginson, R. P. (1985). *Fixing-assimilation in language acquisition*. Unpublished doctoral dissertation. Washington State University.
Iverson, J. M. (2021). Developmental variability and developmental cascades: Lessons from motor and language development in infancy. In *Current Directions in Psychological Science*, 0963721421993822.
Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology, 98*, 1–21.
Le Normand, M. T., Moreno-Torres, I., Parisse, C., & Dellatolas, G. (2013). How do children acquire early grammar and build multiword utterances? A corpus study of French children aged 2 to 4. *Child Development, 84*(2), 647–661.
Leonard, L. B., & Finneran, D. (2003). Grammatical morpheme effects on MLU. *Journal of Speech, Language, and Hearing Research, 46*(4), 878–888.
Leopold, W. F. (1949, May). Original invention in infant language. In *, Vol. 3, No. 1. Symposium: A Quarterly Journal in Modern Literatures* (pp. 66–75). Taylor & Francis Group.
MacWhinney, B. (1991). *The CHILDES Project: Tools for Analyzing Talk*. Hillsdale, NJ: Erlbaum.
MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
Moore, C. A., & Maassen, B. (2004). Physiologic development of speech production. In B. Maassen, R. Kent, H. Peters, P. van Lieshout, & W. Hulstijn (Eds.), *Speech Motor Control in Normal and Disordered Speech* (pp. 191–209). Oxford University Press.
Muggeo, V. M. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine, 22*(19), 3055–3071.
Muggeo, V. M., Atkins, D. C., Gallop, R. J., & Dimidjian, S. (2014). Segmented mixed models with random changepoints: A maximum likelihood approach with application to treatment for depression study. *Statistical Modelling, 14*(4), 293–313.
Muggeo, V. M. R. (2016). Segmented mixed models with random changepoints in R. Working Paper. Available at https://www.researchgate.net/publication/292629179.
Newbury, J., Klee, T., Stokes, S. F., & Moran, C. (2016). Interrelationships between working memory, processing speed, and language development in the age range 2–4 years. *Journal of Speech, Language, and Hearing Research, 59*(5), 1146–1158.
Nóro, L. A., & Mota, H. B. (2019). Relationship between mean length of utterance and vocabulary in children with typical language development. *Revista CEFAC, 21*.
Oakes, L. M., & Rakison, D. H. (2019). *Developmental Cascades: Building the Infant Mind*. Oxford University Press.
Parsons. (2006). *Positional Effects in Phonological Development: A Case Study*. Memorial University of Newfoundland. M.A. Thesis.
Pine, J. M., Lieven, E. V., & Rowland, C. F. (1997). Stylistic variation at the "single-word" stage: Relations between maternal speech characteristics and children's vocabulary composition and usage. *Child Development, 68*(5), 807–819.
Post, K. (1992). *The language learning environment of laterborns in a rural Florida community*. Unpublished doctoral dissertation. Harvard University.
Radford, A. (1990). *Syntactic theory and the acquisition of English syntax: The nature of early child grammars of English*. Oxford: Willey-Blackwell.
Rollins, P. R., Snow, C. E., & Willett, J. B. (1996). Predictors of MLU: Semantic and morphological developments. *First Language, 16*(47), 243–259.
Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*(2), 105–134.
Scarborough, H. S., Rescorla, L., Tager-Flusberg, H., Fowler, A. E., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language-disordered groups. *Applied PsychoLinguistics, 12*(1), 23–46.
Scollon, R. (1976). *Conversations with a one-year-old: A case study of the developmental foundation of syntax*. Honolulu: University Press of Hawaii.
Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*(2), B11–B21.
Shipley, E. F., Smith, C. S., & Gleitman, L. R. (1969). A study in the acquisition of language: Free responses to commands. *Language*, 322–342.
Theakston, A. L., Lieven, E. V. M., Pine, J. M., & Rowland, C. F. (2001). The role of performance limitations in the acquisition of verb-argument structure: An alternative account. *Journal of Child Language, 28*, 127–152.
Valian, V. (1991). Syntactic subjects in the early speech of American and Italian children. *Cognition, 40*(1–2), 21–81.
Valian, V., & Aubry, S. (2005). When opportunity knocks twice: Two-year-olds' repetition of sentence subjects. *Journal of Child Language, 32*(3), 617–641.
Valian, V., Hoeffner, J., & Aubry, S. (1996). Young children's imitation of sentence subjects: Evidence of processing limitations. *Developmental Psychology, 32*(1), 153–164.