Bursty, irregular speech input to preschoolers predicts vocabulary size

Research Highlights

1. We quantified the quantity and distribution ("burstiness") of speech directed to children aged 2-7 years using daylong, child-centered audio recorders.

2. We modeled how the quantity and burstiness of speech input related to children's age and concurrent receptive vocabulary size, assessed via a standardized test in the lab.

3. Speech input burstiness was a stronger correlate of age than speech input quantity (number of words), and burstiness was also a stronger predictor of vocabulary.

4. Results support Exposure-Consolidation models of cognitive development, and suggest that children's vocabulary learning benefits from opportunities to consolidate new linguistic information.

Abstract

Children learn language by listening to speech from caregivers around them. However, the type and quantity of speech input that children are exposed to changes throughout early childhood in ways that are poorly understood due to the small samples (few participants, limited hours of observation) typically available in developmental psychology. Here we used child-centered audio recorders to unobtrusively measure speech input in the home to 292 children, aged 2-7 years, over 555 distinct days (approximately 8,600 total hours of observation, or 29.62 hours/child). These large timescales allowed us to compare how different dimensions of child-directed speech input (quantity, burstiness) varied throughout early childhood. We then evaluated the relationship between each dimension of input and children's concurrent receptive vocabulary size. We found that the burstiness of speech input (spikes of words) was a stronger correlate with age than the quantity of speech input. Input burstiness was also a stronger predictor than input quantity for children's vocabulary size: children who heard spiky, more intense bouts of input had larger vocabularies. Overall, these results reaffirm the importance of speech input in the home for children's language development, and support exposure-consolidation models of early language development.

Keywords: language development, socialization, early childhood, home environment

1 Introduction

The rate of language development can vary widely, even among typically-developing children: precocious English-learning children may have vocabularies upwards of 200 words by 18 months (Robinson et al., 1990), while other children may not reach this milestone until age 24 months or later (Rescorla, 1989). One of the strongest predictors of these individual differences in children's language outcomes is the speech input that children hear in the home (Hoff, 2003; Huttenlocher et al., 2010; Rowe, 2012). This link between speech input and language development is apparent in children as young as 18 months: children who hear more speech from caregivers process words faster (Hurtado et al., 2008), grow larger vocabularies (Rowe, 2008), and produce more complex grammatical structures (Huttenlocher et al., 2010) well into early childhood. These effects are then cascading, with long-term academic implications; for example, preschoolers with larger vocabularies typically master critical precursors to literacy, such as phoneme and rhyme awareness, at an earlier age (Metsala, 1999).

Speech input to children is often referred to as CHILD-DIRECTED SPEECH (CDS), a speech register characterized by shorter and slower sentences, a simplified vocabulary, and exaggerated prosodic cues (Schwab & Lew-Williams, 2016; Soderstrom, 2007). Many characteristics of CDS have been proposed to convey benefits for children's language learning. For example, slower speech rates may delineate boundaries between words, which could help children parse individual lexical items from the fast-moving speech stream (Raneri et al., 2020), and disambiguate new versus already-learned word referents (Shi et al., 2023). Thus, not all speech heard in the home is equally beneficial for children's language learning—CDS is especially beneficial, more so than overheard speech, through early childhood (Foushee et al., 2016).

Quantitatively, CDS can be measured using a number of different units including the number of words, seconds or minutes of speech, and number of conversational back-and-forth turns between the child and the caregiver (Bergelson et al., 2019; Romeo et al., 2018; Schwab & Lew-Williams, 2016). Additional work has demonstrated how various qualitative dimensions of the input—diverse vocabulary, decontextualized language, extra-linguistic cues—also relate to vocabulary outcomes (assessed longitudinally 24-36 months: (Hirsh-Pasek et al., 2015), 18-54 months: (Rowe, 2012), and in 14-18 month-olds: (Cartmill et al., 2013)).

The connection between CDS and children's language is particularly well-documented for vocabulary: children who hear more speech input are more practiced at the skills required to process fast, variable speech in real time, which then aides in word learning throughout toddlerhood and the preschool years (Hurtado et al., 2008; Weisleder & Fernald, 2013). However, despite what has now been decades of work linking dimensions of CDS and children's vocabulary, many of those studies (including work that we have conducted) have been limited. Some findings were based on limited observational samples (e.g. 1 hour of observation in the home), or parent input to the child was elicited through semi-naturalistic, in-lab observation (Newman et al., 2016). These designs were made for reasons of feasibility: hand annotation of home-based audio/video recording can take up to 8x the recording's duration. Even when samples were taken over longer periods, often using child-centered audio recorders that children wear for 1 or more days, CDS estimates were averaged over daylong time scales (e.g. average number of turns/hour) (Romeo et al., 2018), or only smaller samples from the overall observation were analyzed (Cychosz et al., 2021).

Thus these previous designs have been limiting because communicative activity in children's lives is *temporally variable* (de Barbaro & Fausey, 2022), occurring in bursts throughout the day as children engage in different activities such as eating, dressing, and reading (Bang et al., 2022; Casillas et al., 2021; Mendoza & Fausey, 2021; Tamis-LeMonda et al., 2019). According to exposure-consolidation models of cognitive development (Bernier et al., 2013; Dionne et al., 2011; Henderson et al., 2013), these bursts, or in the case of word learning "lexical spikes," may facilitate children's language learning because children require time to consolidate linguistic information, such as new words, and hold it for later retrieval. We will refer to this as the Burst Hypothesis (for similar discussion regarding vocal turn taking, see the Interaction Burst Hypothesis (Elmlinger et al., 2023)).

Support for the Burst Hypothesis comes from randomized control sleep or "nap" designs, and other caregiver-centered predictors of language development. In a typical nap design for language research, children might be presented with a novel word learning task, assigned into a nap or wake group, and tested for retention after napping (or not). Nap designs have demonstrated that rest (i.e. an opportunity for consolidation without further input or learning) leads to stronger lexical memory consolidation effects in toddlers and preschoolers for nouns (16-month-olds: Horváth et al. (2015)) and verbs (2-year-olds: He et al. (2020); 3-year-olds: Sandoval et al. (2017)).

Additional support for the Burst Hypothesis comes from the joint attention literature. It is well-known that children who engage in more episodes of joint attention—caregiver-child coordinated attention on an object or event—learn words faster (Akhtar & Gernsbacher, 2007; Tomasello & Todd, 1983). But recent work suggests that the proportion of joint attention *bouts* (< 3 seconds) during caregiver-infant interaction at 9 months are uniquely predictive of vocabulary size 3 and 6 months later (Abney et al., 2017). Thus, the temporal dynamics of joint attention are predictive of 12- and 15-month-olds' vocabulary sizes *beyond mere quantity*.

The Burst Hypothesis combines ideas from sleep-based consolidation and joint attention bouts: it is not simply that *sleep* allows linguistic input to consolidate, facilitating learning and word-referent mapping, but that the *dose* and *distribution* of linguistic input to children matters as well. Children's vocabulary development benefits from increased quantities of CDS in the home. It is unknown if the distribution of input likewise matters. Evidence for the Burst Hypothesis would see a positive association between children who hera irregular, burstier speech input and vocabulary size, beyond effects of input quantity. These CDS bursts would provide the large quantities of linguistic input that previous work has demonstrated to be an important component of language development (Hoff, 2003; Rowe, 2008), but critically, the bursts would dose the input out in ways that better accommodate the time that children require to consolidate new linguistic information.

1.1 Lack of demographic diversity is particularly problematic for studies on children's language socialization

Another limitation of previous work is its highly homogeneous demographic samples, leading to research findings that may not be applicable to all cultural backgrounds, even within the United States. The lack of diverse samples is a concern in many areas of developmental science, but it has been particularly problematic for CDS studies. One of the seminal studies to document an effect of speech input for children's language examined how socioeconomic status (SES) impacted the number of cumulative words that children were exposed to by kindergarten (the "30-million Word Gap") (Hart & Risley, 1995). That work, and its subsequent findings, has been strongly critiqued for the poor representation of socioeconomic diversity (e.g. N=6 families in the low-SES group compared to N=23 in the mid-SES group) and the conflation of race/ethnicity and SES within groups. This overall poor representation of diverse cultural and socioeconomic backgrounds has been the source of much recent debate concerning the validity of the "Word Gap," with some arguing that the lack of diverse representation could invalidate early findings (Golinkoff et al., 2019; Ochs & Kremer-Sadl, 2020). To deny findings surrounding the Word Gap could be consequential for children's language development and school readiness. But it is likewise inappropriate to extrapolate models and measurements made predominantly over white, upper middle-class families to all other groups. To truly understand if the relationship between different measures of CDS and children's language is robust, our samples must be large, comprehensive, and diverse.

2 The Current Study

This study tests whether the *distribution* of speech input directed to children is a more relevant predictor of concurrent language development than the *quantity* of input (the Burst Hypothesis). We pose the following questions:

1. How does the overall dose (quantity) and distribution (burstiness) of CDS change from age 2 to 7 years?

In line with previous work, we anticipate that (1a) children will be exposed to less CDS as they grow older (Rowe, 2012), but that (1b) burstiness of CDS will be a stronger correlate of age than quantity.

2. What is the relationship between the overall dose (quantity) and distribution (burstiness) of CDS and concurrent vocabulary size?

2a. Bursty CDS will be positively associated with receptive vocabulary size, and will be a stronger predictor of vocabulary than input quantity alone.

Our approach allows us to robustly put the Burst Hypothesis to the test, and simultaneously address several of the methodological gaps that have made it difficult to accurately model the relationship between CDS and vocabulary development. First, we analyze a sample of daylong audio recordings from 292 unique children (approximately 8,600 hours of observation over 555 different days), which is magnitudes larger than many similar home-based studies of language development. Second, we take advantage of recent advances in natural language processing and machine learning techniques to feasibly parse the *entire duration* of each recording, allowing us to model the temporal fluctuations of CDS, in addition to the raw quantities/hour measure which is commonly reported. Finally, we intentionally curated a dataset to encompass more racial, ethnic, and socioeconomic diversity of families with preschoolers in the United States to make our results more robust and generalizable to families of all backgrounds. While we acknowledge that our sample is still skewed (towards higher SES), and lacking representation of some groups, we believe this dataset is an important first step to diversifying language socialization studies. We carefully acknowledge the limitations of our dataset in our interpretation of the data.

3 Methods

3.1 Participants

Participants were N=292 children (133 girls and 159 boys), exposed to American English (N=40 [13.70%] children were acquiring one additional language; see Supp. Materials I). All were typically-developing, per caregiver report, and passed a pure-tone hearing screening (0.5, 1, 2, and 4KHz) at study enrollment. Families were recruited from research databases and outreach to preschools in the Midwest and on the East coast. The data that we report on here come from larger longitudinal research programs where families were enrolled in multi-year studies on child language development. See Table 1 for demographic information. Importantly, for our goal of analyzing a diverse sample, 15.41% of mothers had a high school degree or less, and 29.79% had less than a college degree.

SES was instantiated as the highest level of maternal education achieved, in line with previous work on child language development (Hoff, 2003; Rowe, 2008). We binned education into five levels: 1) < high school, 2) high school diploma equivalent certificate or diploma, 3) technical-associate degree, some college (2+ years), or trade school, 4) college degree, and 5) graduate degree.

3.2 Data collection

Families completed daylong audio recordings at one or more timepoints. The child wore a small, lightweight Language ENvironment Analysis (LENA) recording device (2”x3”; 2 oz.) in a specialized vest for the entire day. Families received the materials either in the mail or upon a visit to the research lab along with written instructions. Recordings were completed on a typical, non-school day. Families were instructed to turn the device on in the morning when the child awoke and continue recording for the duration of the device battery (16 hrs.). During bathtime and other water activities, caregivers were told to place the recorder in a safe, dry place as close to the child as possible. The device continued recording while the child napped.

N=192 children contributed multiple home audio recordings, collected longitudinally at different timepoints. Of these, N=38 children (13.01% of sample) contributed 3-4 recordings and N=154 children (52.74%) completed two recordings. A repeated measures analysis of variance (ANOVA) demonstrated that there were no differences in the number of recordings (p>.05), or hours of observation (p>.05), by child age. Our statistical modeling takes into account repeated

Table 1

Demographic information.

	Mean (SD) Range
Gender (F, M)	133, 159
Age (mos) at enrollment	44.51 (15.63) 28-85
Maternal Ed.*	3.88 (1.14) 1-5
	number of households in each group
1: < high school	10
2: G.E.D. or high school diploma	35
3: associate degree, some college, or trade school	42
4: college degree	93
5: graduate degree	108
Ethnicity (N)†	
	number of participants in each group
Hispanic	32
Race (N)	
Asian	3
Black	51
Hawaiian/Pacific Islander	1
White	206
American Indian	1
Asian & white	2
Black & white	1
More than 1 race (unspecified)	7

*Maternal education was unreported for 4 children. Race/ethnicity information was unreported for 20 children. Maternal education was binned into discrete levels (1-5); see text for detail.

measures within families over time. See Table 2 for descriptive statistics of recordings and Figure 1 for a histogram of observation hours by age.

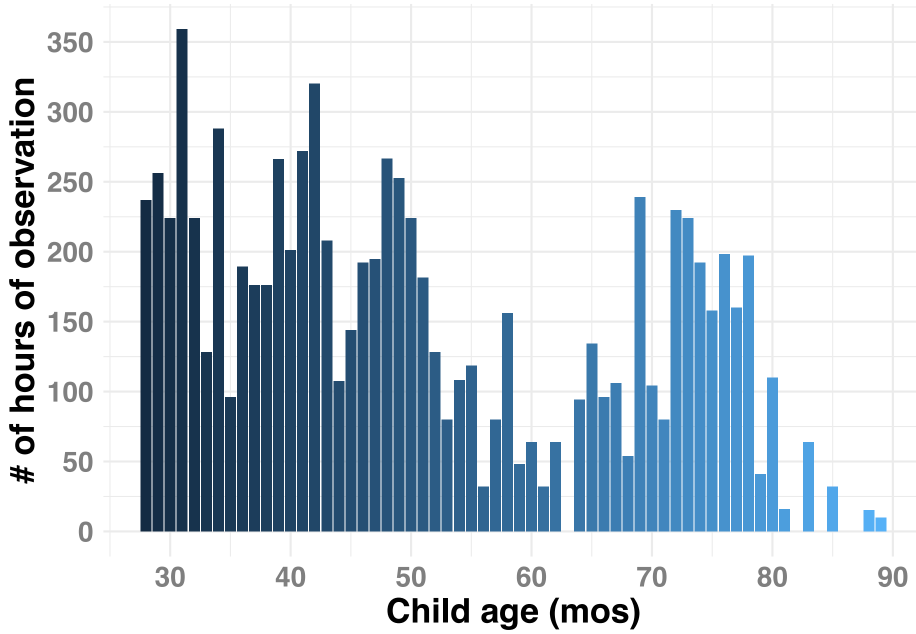

Figure 1. Total hours of observation via child-centered daylong audio recordings, by child age (in months).

N=285 children (97.6%) completed a vocabulary assessment, the Peabody Picture Vocabulary Test (PPVT-4) (Dunn & Dunn, 2007), concurrently with their home audio recording. For the children who completed multiple home audio recordings, we report on the relationship between the home environment and vocabulary outcome at the first timepoint. See Table 2 for descriptive statistics.

Deriving child-directed speech estimates. We first removed all recordings <5 hours (N=7). An additional recording was removed due to experimenter error during collection. N=555 audio recordings remained. Measures of children's speech input were semi-automatically derived from each recording using a combination of automatic speech recognition algorithms and pre-trained classifiers (Figure 2). All processing scripts are included in the project's Github repository (*anonymized for review:* `https://anonymous.4open.science/r/everyday_speech-E07F/`).

Table 2

Audio recording and vocabulary score statistics.

	Mean (SD)	**Range**
Recording Length (Hours)	15.58 (1.43)	5.19-16
# of CDS words/hr	1094 (494)	99.95-3564.84
Burst Amplitude (Words)	400 (133)	92.59-897.1
Burst Duration (30-second segments)	10 (5)	3.57-54
PPVT-4 (GSV) Score	128 (22.82)	53-182

CDS=child-directed speech. PPVT-4=Peabody Picture Vocabulary Test-4th edition. GSV=Growth Scale Value (a form of standardization that allows vocabulary scores to be compared between and within children (Dunn & Dunn, 2007)).

First each recording was divided into 5-minute epochs. LENA recordings have an associated metadata file ('interpreted time segment' file) which was mined to report the linguistic units within each epoch, such as adult word count (AWC), for use in the classifiers. We applied pre-trained classifiers to compute the probability that each epoch contained sleep and/or CDS. It was important to identify sleep epochs to remove children's daily naps, but also because some families continued recording after the child went to sleep at night. Complete technical details about the sleep and CDS classifiers' training data and verification with hand-coded data are available in Bang et al. (2023), but in brief: the LENA system contains a speaker diarization algorithm which assigns speaker tags and timestamps to audio clips (Xu et al., 2009). The probability of *sleep* (versus awake) was derived from a decision tree classifier on the basis of meaningful speech (primary split) and silence (secondary split) (achieved AUC (area under the curve)=.881). Epochs with >.9 probability of sleep versus awake were removed (mean computed probability of 'sleep' within sleep epochs=0.92 (SD=0.11); the probability of sleep followed a bimodal distribution such that an epoch either had a very high, or very low, probability of containing sleep). An average of 4.44 hours of sleep (SD=2.07) were removed/recording.

Next, we employed a classifier to estimate the probability that at least 50% of speech input

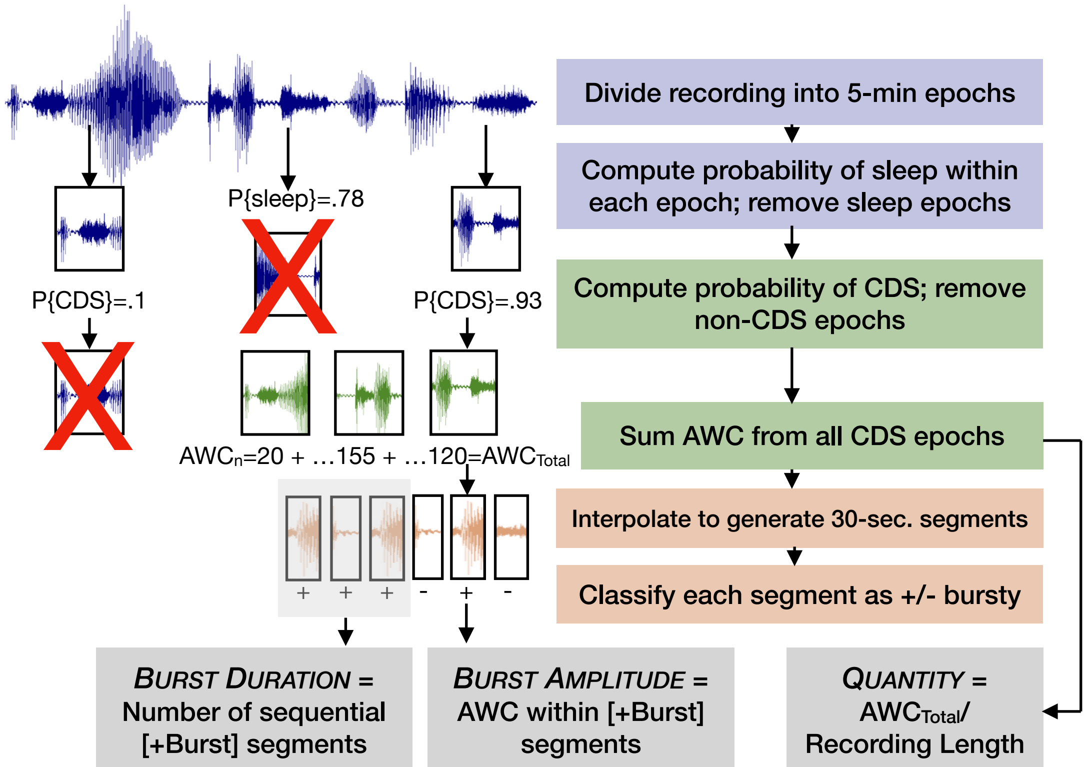

Figure 2. Flowchart illustrating audio processing steps and methods used to generate dose (quantity) and distribution (burst) metrics. CDS=child-directed speech. AWC=adult word count.

within each epoch was CDS (versus other-directed speech or 'ODS'). This probability estimate was derived from an eXtreme Gradient-Boosted tree based on measures of silence in the recording and features from the diarization algorithm (achieved AUC=0.72). Epochs with >.5 probability of CDS versus ODS were considered 'CDS' (following Bang et al. (2023)). The mean computed probability of 'CDS' within the CDS epochs was 0.67 (SD=.10)) and an average of 67.19% of non-sleep epochs within each recording were classified as 'CDS' (SD=10.41%). Using the LENA algorithm's AWC estimator, we then summed the total number of words within each epoch as our CDS word count. See Table 2 for summary statistics on the estimated number of CDS words/recording.

Our measure of input dose (quantity) is the *average number of CDS words/hour*: we divided the sum all of the words within the CDS epochs by the length of the recording.

Our measures of input distribution (burstiness) are *burst amplitude* and *burst duration*, computed as follows (see Figure 3):

1. To increase temporal resolution, we interpolated data points within CDS epochs by exploding each epoch by 10 to generate 30-second samples of CDS for the entire recording.

2. We computed the median number of CDS words from the entirety of each daylong recording. This was done dynamically so each recording had its own median CDS word count. Dynamic computation was chosen because bursts are relative, not absolute—what may be considered bursty for one child, age group, etc. may not be bursty for another.

3. We classified each segment as +/- "bursty," where a bursty segment has a CDS word count greater than 3 SDs from the recording's median. 3 SDs should ensure that only the most extreme segments are classified as bursty (under a normal distribution only approximately the top 1% of segments would be considered bursty).

4. For burst amplitude, we counted the number of words within each bursty segment and calculated the average amplitude over all bursts for the modeling.

5. For burst duration, we counted the number of sequential bursty segments and then computed the average number of sequential segments from bursts over the entire recording for the modeling.

4 Results

Data were analyzed in the RStudio computing environment (R version 4.2.1; RStudioTeam, 2024). All computing and statistical analyses are included in the project's GitHub repository (*anonymized for review:* `https://anonymous.4open.science/r/everyday_speech-E07F/`). Visualizations were made using `ggplot2` (Wickham, 2016) and modeling was conducted using

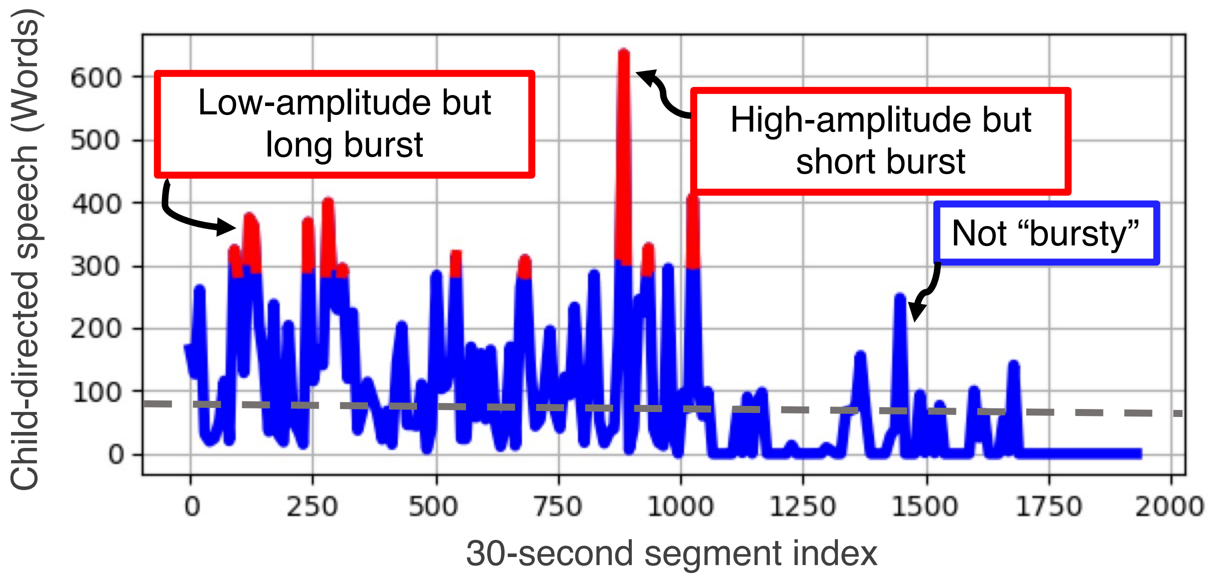

Figure 3. Illustration of measures of child-directed speech "burstiness" metrics from a single daylong audio recording. Red=30-second segments classified as +bursty spikes. Blue=segments classified as -bursty. Dashed line=median child-directed speech word count from the recording. Since burst amplitude and duration are evaluated separately, a child could have both high-amplitude and long-duration input bursts.

`lme4` and `lmerTest` packages (Bates et al., 2015; Kuznetsova et al., 2017); see project documentation for package versions.

To account for the repeated measures within children and over time, we fit linear mixed effects models where the baseline model always included a random intercept of Child. The baseline also always included fixed effects of Child Gender (contrast coded, 1=Female) and Maternal Education. Continuous predictors (Child Age, Maternal Education, Input Measures) were centered to facilitate model and effect size comparison. Data visualizations employ the non-transformed measures for clarity. Best model fit was determined from log-likelihood tests, AIC values, and model summaries. Input measures were log-normalized to ensure assumptions of normal distribution were met and because we assume that there may be more meaningful differences between, for example, burst amplitudes of 95-105 words than 885-895.

4.1 How does the dose (quantity) and distribution (burstiness) of CDS change from age 2 to 7 years?

To compare how quantity and burstiness of CDS vary by child age, we fit three baseline models, each predicting a different dimension of CDS (burst amplitude, burst duration, and input quantity; see Table 2 for descriptive statistics). Fixed effects included Child Gender and Maternal Education. To each model we added Child Age, which significantly improved a model predicting burst amplitude (χ^2=4.25, p=.04) and burst duration (χ^2=16.63, p<.001), but not input quantity (χ^2=0.35, p>.05; see Table 3 for model summaries). (AIC values likewise improved upon adding Child Age for burst amplitude and duration models, but not the input quantity model.) There was a negative effect of age for both the amplitude (β=-0.0022, p=0.04) and duration models (β=-0.0047, p<.001), suggesting that the amplitude and duration of CDS bursts decrease between 2 and 7 years; however, CDS quantity does not decrease with age in this sample.

4.2 How does the dose (quantity) and distribution (burstiness) of CDS correlate with concurrent vocabulary size?

Next, we compared how the quantity and burstiness of CDS predicted children's receptive vocabulary size. We first fit a baseline model with fixed effects of Child Age, Child Gender, Maternal Education, and Average Hourly CDS Words. Child Age was modeled to account for its strong relationship with word learning over this period. Hourly CDS Words was added to evaluate if the distribution of input parameters (Burst Amplitude and Burst Duration) predicted vocabulary scores above and beyond sheer input quantity. We fit two models to predict vocabulary scores using the two different distribution parameters; both improved their respective model fits (Burst Amplitude: χ^2=4.42, p=.04; Burst Duration: χ^2=4.12, p<.001). Specifically, both measures of input distribution were positive predictors of concurrent vocabulary scores (Burst Amplitude: β=4.54, p=0.04; Burst Duration: β=3.17, p=0.04; Figure 5), even after controlling for the quantity of input. Both measures also rendered the quantity of input

Table 3

The effect of child age on the quantity and burstiness of child-directed speech

	Burst Amplitude Model	Burst Duration Model
Intercept	β=-0.01	β=0.02
	CI=(-0.05, 0.04)	(-0.02, 0.07)
	t = -0.40	t = 0.90
	p = 0.69	p = 0.37
Age (mos)	$\beta=-0.002^{*}$	$\beta=-0.005^{***}$
	(-0.004, -0.0001)	(-0.01, -0.003)
	t = -2.05	t = -4.23
	p = 0.04	p < 0.001
Gender:Female	β=0.03	$\beta=-0.01$
	(-0.04, 0.10)	(-0.08, 0.06)
	t = 0.82	t = -0.29
	p = 0.42	p = 0.78
Mat. Ed.	$\beta=0.06^{***}$	β=0.02
	(0.04, 0.08)	(-0.01, 0.04)
	t = 5.50	t = 1.44
	p < .001	p = 0.16
Log Likelihood	-196.92	-198.48
AIC	405.84	408.96

All continuous variables are centered. *p<0.05; **p<0.01; ***p<0.001

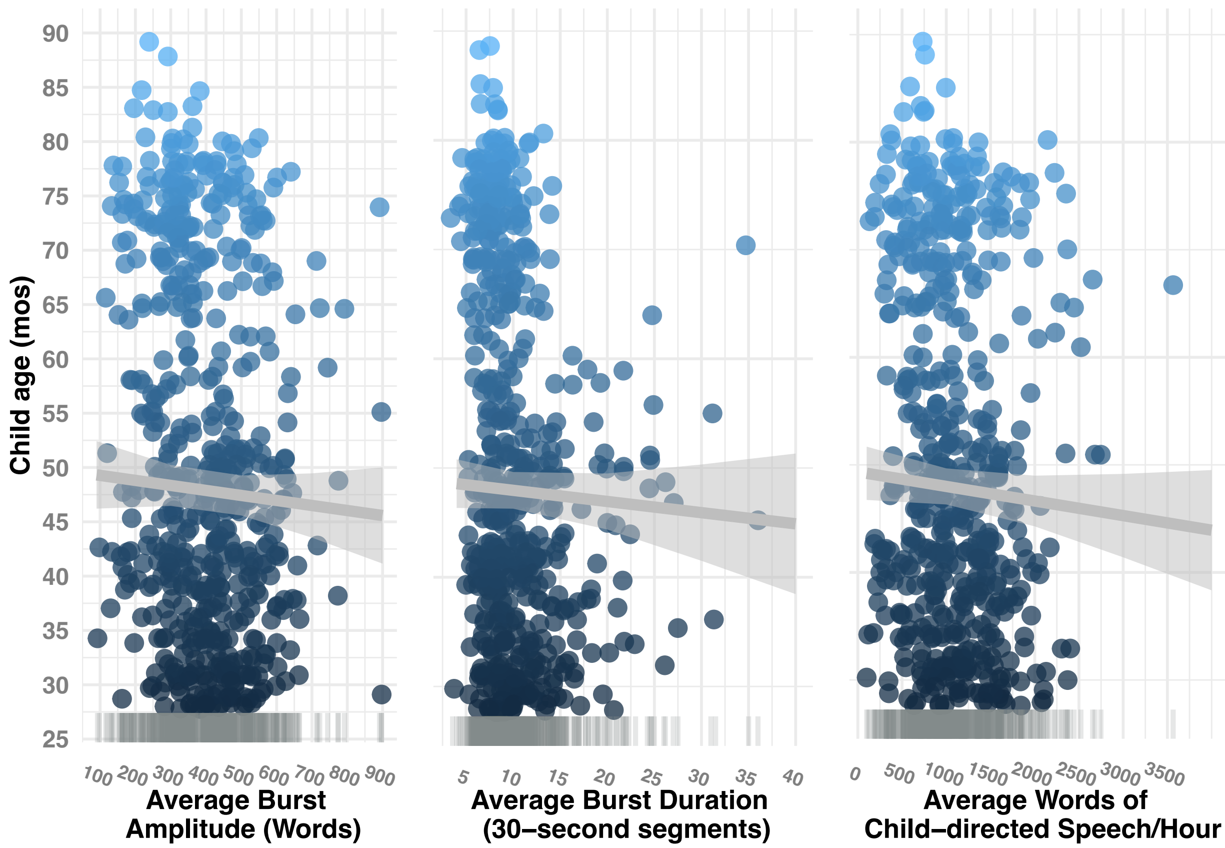

Figure 4. The relationship between input burstiness and child age. Each data point represents raw, non-transformed data from one daylong recording. The light gray line represents coefficient values from the associated model parameter in Table 3. Ribbons represent 95% confidence intervals around the model estimate. Each pile in the rug plots (gray, bottom) represents the value from one recording.

non-significant (See Table 4 for model summaries).

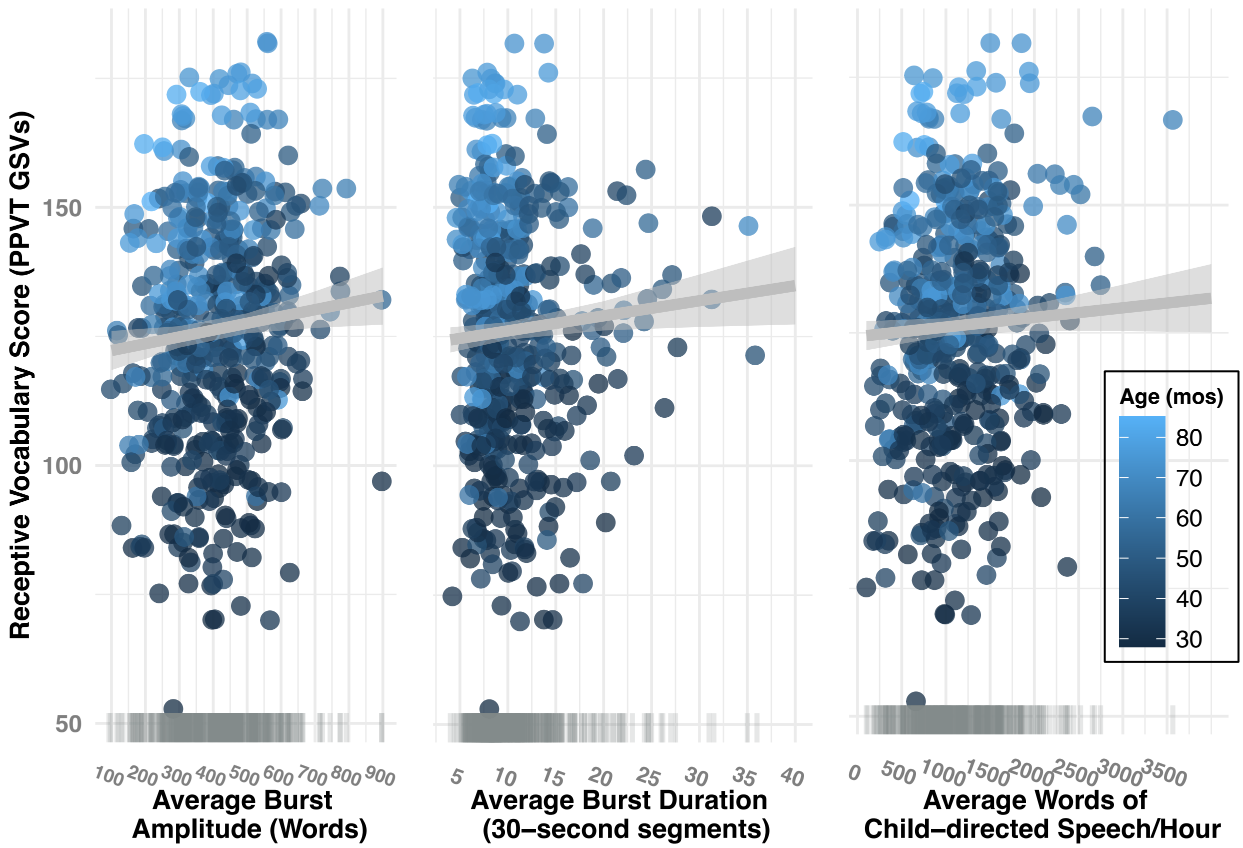

Figure 5. The relationship between input burstiness and receptive vocabulary size (PPVT-4 score). Each data point represents raw, non-transformed data from one daylong recording. The light gray line represents coefficient values from the associated model parameter in Table 4. Ribbons represent 95% confidence intervals around the model estimate. Each pile in the rug plots (gray, bottom) represents the value from one recording.

5 Discussion

This study examined whether the distribution of speech input to children in early childhood was a more relevant predictor of concurrent language development than quantity of input (the Burst Hypothesis). We used a novel pipeline of speech recognition and machine learning classification algorithms to analyze 555 daylong audio recordings of children's home language

Table 4

The effect of quantity and burstiness of child-directed speech on concurrent vocabulary scores

	Burst Amplitude Model	Burst Duration Model
Intercept	β=129.21***	β=129.07***
	CI=(126.94, 131.49)	(126.80, 131.33)
	t = 111.35	t = 111.83
	p < .001	p < .001
Bursty Measure	β=4.54*	β=3.17*
	(0.29, 8.80)	(0.10, 6.24)
	t = 2.10	t = 2.03
	p = 0.04	p = 0.05
Avg. Hrly Words	β=−0.22	β=1.58
	(−3.34, 2.90)	(−0.81, 3.96)
	t = −0.14	t = 1.30
	p = 0.89	p = 0.20
Age (mos)	β=1.49***	β=1.48***
	(1.40, 1.58)	(1.39, 1.57)
	t = 31.78	t = 31.57
	p < .001	p < .001
Gender:Female	β=1.96	β=1.95
	(−1.45, 5.37)	(−1.43, 5.34)
	t = 1.13	t = 1.13
	p = 0.27	p = 0.26
Mat. Ed.	β=5.37***	β=5.44***
	(4.31, 6.44)	(4.38, 6.49)
	t = 9.88	t = 10.10
	p < .001	p < .001
Log Likelihood	−1,853.41	−1,853.90
AIC	3,722.82	3,723.80

All continuous variables are centered. *p<0.05; **p<0.01; ***p<0.001

environments (>8,600 hours of observation). This allowed us to compare the distribution of speech input—measured as both the amplitude and duration of speech bursts—versus the quantity of input, to children. These input measures were then correlated with the children's concurrent receptive vocabulary scores. Our sample was both large (N=292 children) and socioeconomically representative, lending a robustness to our results that has been missing in much language socialization research. Results showed that both measures of input distribution (burst amplitude and duration) predicted (1) child age and (2) concurrent vocabulary above and beyond the quantity of input.

Based on effect sizes, burst duration was a stronger correlate of child age than burst amplitude. Burst duration and amplitude were both strong predictors of vocabulary size: children with longer bursts and spikier bursts in their input had larger vocabularies, even after controlling for overall input quantity. These results support the Burst Hypothesis, and more general exposure-consolidation theories of language and cognitive development (Bernier et al., 2013; Dionne et al., 2011; Henderson et al., 2013), specifically for word learning. The importance of bursts is not pre-determined—adult learners appear to benefit more from spaced, rather than massed, exposure ("spaced learning") (Smolen et al., 2016). The Burst Hypothesis functions under the observation that speech input to children ebbs and flows throughout day as children engage in various activities that have more (e.g. mealtime), less (e.g. independent play), or no (e.g. naps) speech input. Input bursts provide rich opportunities for children to learn, while ebbs give children the opportunity to consolidate the new referent information and entrench representations to facilitate later retrieval. For word learning, bursts may provide many repetitions of a word in quick succession (Schwab & Lew-Williams, 2016), including in similar sentential frames (Mintz, 2003) or via distinct referents in the environment (Perry et al., 2010)—all contexts that are known to bolster lexical acquisition via grammatical category development and/or word-referent mapping. Future work should explore how some of the potential components of lexical input make up individual speech bursts to better elucidate these details.

Contrary to our prediction, child age was not negatively correlated with input quantity. We believe that this null effect could be because while children are exposed to fewer CDS utterances as they grow older, their CDS consists of more diverse word types (Rowe, 2012), and greater morpheme to word ratios. So the effect of age was potentially masked by the differences in lexical content by age. Indeed, when we re-fit the same model predicting age from CDS quantity to only include children less than 40 months (N=165), we saw the predicted negative relationship between age and CDS quantity (β=-0.02, p=0.04).

Because these results are correlational, future work should continue to test the Burst Hypothesis, including extending to different areas of language development. Already Elmlinger et al. (2023) proposed that bursts of verbal interaction between caregivers and infants may encourage infants to reach milestones in early vocal development such as babbling. We extended our definition of bursty input to examine bursts in conversational turns between the child and caregiver(s) (for full analysis see Supp. Materials II), and adult speech that did not differentiate between CDS and ODS (Supp. Materials III). We replicated the finding that children with burstier conversational turns had larger vocabularies, but there was no relationship between burstiness in overall input (CDS+ODS) and vocabulary. We also conducted these supplementary analyses to be more confident in the CDS classifier because one limitation of this work is reliance on automatic speech recognition tools (e.g. speaker diarization algorithms). To train the classifiers we employed, Bang et al. (2023) constructed a thorough training dataset (through 5-fold cross-validation) and carefully compared the results of their classifiers with "gold standard" human annotation to derive sensitivity-specificity statistics. Nevertheless, the authors warn that "there is no guarantee of similar performance for families and settings dissimilar to the present dataset" (p. 227). We think of the present work as an important first step in taking advantage of unprecedentedly large datasets, while still understanding the limitations that semi-automatic pipelines bring to research design.

This analysis only compared the distribution of CDS throughout a child's day. However, sleep also allows preschoolers to consolidate new cognitive information (Giganti et al., 2014; Kurdziel et al., 2013), including linguistic (He et al., 2020; Horváth et al., 2015; Sandoval et al., 2017). We did not have data on the duration or frequency of the children's typical sleep and nap patterns. However, we did have estimates on the duration of sleep within each recording, which we used to examine if children with longer naps had larger vocabularies. This analysis was exploratory, not confirmatory. We limited the analysis to children's nap times, not overnight sleep (see Supp. Materials IV for detail). We analyzed children under 42 months, as we assumed that children above this age were unlikely to still be regularly napping. Estimated nap durations ranged from 0-55 minutes (M= 26.57, SD=17.52) for the N=124 remaining recordings (N=115 children). Nap duration was positively related to concurrent vocabulary (β=0.15, p=.03, controlling for child age, gender, and SES). This finding lends another piece of evidence in support of the idea that children benefit from increased opportunity to consolidate new linguistic information, in this case via sleep, something that has not previously been demonstrated using naturalistic observations of children's sleep behavior. It is important not to over-interpret this post-hoc analysis. We do not *know* that the children were resting during these periods—they could have been engaging in silent play in a quiet environment (e.g. without T.V.). But even if that were the case, it would still mean that the child was receiving minimal input and therefore had an opportunity to consolidate newly-acquired knowledge.

In conclusion, we find support for the relevance of the distribution of speech input, above quantity of input, for preschoolers' language development (the Burst Hypothesis). Our results should be extended to additional metrics of input and language development milestones, to better understand the dynamics of input and how they relate to children's overall development.

6 References

Abney, D. H., Smith, L. B., & Yu, C. (2017). It's Time: Quantifying the Relevant Timescales for Joint Attention. In *Proceedings of the 2017 Cognitive Science Society.*

Akhtar, N., & Gernsbacher, M. A. (2007). Joint Attention and Vocabulary Development: A Critical Look. *Language and Linguistics Compass*, *1*(3), 195–207. doi: 10.1111/j.1749-818X.2007.00014.x

Bang, J. Y., Kachergis, G., Weisleder, A., & Marchman, V. (2023). An automated classifier for periods of sleep and target-child-directed speech from LENA recordings. *Language Development Research*. doi: 10.34842/XMRQ-ER43

Bang, J. Y., Mora, A., Munévar, M., Fernald, A., & Marchman, V. A. (2022). *Time to talk: Multiple sources of variability in caregiver verbal engagement during everyday activities in English- and Spanish-speaking families in the U.S* (Preprint). PsyArXiv. doi: 10.31234/osf.io/6jzwg

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

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What Do North American Babies Hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724. doi: 10.1111/desc.12724

Bernier, A., Beauchamp, M. H., Bouvette-Turcot, A.-A., Carlson, S. M., & Carrier, J. (2013). Sleep and Cognition in Preschool Years: Specific Links to Executive Functioning. *Child Development*, *84*(5), 1542–1553. doi: 10.1111/cdev.12063

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283. doi: 10.1073/pnas.1309518110

Casillas, M., Brown, P., & Levinson, S. (2021). Early language experience in a Tseltal Mayan village. *Child Development*, *0*(0), 1–17.

Cychosz, M., Cristia, A., Bergelson, E., Casillas, M., Baudet, G., Warlaumont, A. S., . . . Seidl,

A. (2021). Vocal development in a large-scale crosslinguistic corpus. *Developmental Science*, *24*(5), e13090. doi: 10.1111/desc.13090

de Barbaro, K., & Fausey, C. M. (2022). Ten Lessons About Infants' Everyday Experiences. *Current Directions in Psychological Science*, *31*(1), 28–33. doi: 10.1177/09637214211059536

Dionne, G., Touchette, E., Forget-Dubois, N., Petit, D., Tremblay, R. E., Montplaisir, J. Y., & Boivin, M. (2011). Associations Between Sleep-Wake Consolidation and Language Development in Early Childhood: A Longitudinal Twin Study. *Sleep*, *34*(8), 987–995. doi: 10.5665/SLEEP.1148

Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody picture vocabulary test.* Pearson Assessments.

Elmlinger, S. L., Goldstein, M. H., & Casillas, M. (2023). Immature Vocalizations Simplify the Speech of Tseltal Mayan and U.S. Caregivers. *Topics in Cognitive Science*, *15*(2), 315–328. doi: 10.1111/tops.12632

Foushee, R., Griffiths, T. L., & Srinivasan, M. (2016). Lexical Complexity of Child-Directed and Overheard Speech: Implications for Learning. In *Proceedings of the 2016 Cognitive Science Society.*

Giganti, F., Arzilli, C., Conte, F., Toselli, M., Viggiano, M. P., & Ficca, G. (2014). The Effect of a Daytime Nap on Priming and Recognition Tasks in Preschool Children. *Sleep*, *37*(6), 1087–1093. doi: 10.5665/sleep.3766

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language Matters: Denying the Existence of the 30-Million-Word Gap Has Serious Consequences. *Child Development*, *90*(3), 985–992. doi: 10.1111/cdev.13128

Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young American children.* Baltimore, MD: Paul H. Brookes Publishing.

He, A. X., Huang, S., Waxman, S., & Arunachalam, S. (2020). Two-year-olds consolidate verb meanings during a nap. *Cognition*, *198*, 104205. doi: 10.1016/j.cognition.2020.104205

Henderson, L., Weighall, A., & Gaskell, G. (2013). Learning new vocabulary during childhood:

Effects of semantic training on lexical consolidation and integration. *Journal of Experimental Child Psychology*, *116*(3), 572–592. doi: 10.1016/j.jecp.2013.07.004

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., ... Suma, K. (2015). The Contribution of Early Communication Quality to Low-Income Children's Language Success. *Psychological Science*, *26*(7), 1071–1083. doi: 10.1177/0956797615581493

Hoff, E. (2003). The Specificity of Environmental Influence: Socioeconomic Status Affects Early Vocabulary Development Via Maternal Speech. *Child Development*, *74*(5), 1368–1378. doi: 10.1111/1467-8624.00612

Horváth, K., Myers, K., Foster, R., & Plunkett, K. (2015). Napping facilitates word learning in early lexical development. *Journal of Sleep Research*, *24*(5), 503–509. doi: 10.1111/jsr.12306

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children. *Developmental Science*, *11*(6), F31-F39. doi: 10.1111/j.1467-7687.2008.00768.x

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of Variability in Children's Language Growth. *Cognitive psychology*, *61*(4), 343–365. doi: 10.1016/j.cogpsych.2010.08.002

Kurdziel, L., Duclos, K., & Spencer, R. M. C. (2013). Sleep spindles in midday naps enhance learning in preschool children. *Proceedings of the National Academy of Sciences*, *110*(43), 17267–17272. doi: 10.1073/pnas.1306418110

Kuznetsova, A., Brockhoff, P., & Christensen, R. (2017). lmerTest Package: Tests in linear mixed-effects models. *Journal of Statistical Software*, *82*(13), 1–26.

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, *24*(6), e13122. doi: 10.1111/desc.13122

Metsala, J. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*(1), 3–19.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91–117. doi: 10.1016/S0010-0277(03)00140-9

Newman, R. S., Rowe, M. L., & Bernstein Ratner, N. (2016). Input and uptake at 7 months predicts toddler vocabulary: The role of child-directed speech and infant processing skills in language development. *Journal of Child Language*, *43*(5), 1158–1173. doi: 10.1017/S0305000915000446

Ochs, E., & Kremer-Sadl, T. (2020). Ethical Blind Spots in Ethnographic and Developmental Approaches to the Language Gap Debate:. *Langage et société*, *N° 170*(2), 39–67. doi: 10.3917/ls.170.0039

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn Locally, Think Globally: Exemplar Variability Supports Higher-Order Generalization and Word Learning. *Psychological Science*, *21*(12), 1894–1902. doi: 10.1177/0956797610389189

Raneri, D., Von Holzen, K., Newman, R. S., & Bernstein Ratner, N. (2020). Change in maternal speech rate to preverbal infants over the first two years of life. *Journal of Child Language*, *47*(6), 1263–1275. doi: 10.1017/S030500091900093X

Rescorla, L. (1989). The Language Development Survey: A Screening Tool for Delayed Language in Toddlers: Journal of Speech and Hearing Disorders: Vol 54, No 4. *Journal of Speech and Hearing Disorders*, *54*(4), 587–599.

Robinson, N. M., Dale, P. S., & Landesman, S. (1990). Validity of Stanford-Binet IV with linguistically precocious toddlers. *Intelligence*, *14*(2), 173–186. doi: 10.1016/0160-2896(90)90003-C

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-Million-Word Gap: Children's Conversational Exposure Is Associated With Language-Related Brain Function. *Psychological Science*, *29*(5), 700–710. doi: 10.1177/0956797617742725

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development, and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205.

Rowe, M. L. (2012). A Longitudinal Investigation of the Role of Quantity and Quality of Child-Directed Speech in Vocabulary Development: Child-Directed Speech and

Vocabulary. *Child Development*, *83*(5), 1762–1774. doi: 10.1111/j.1467-8624.2012.01805.x

RStudioTeam. (2020). *RStudio: Integrated Development for R.* Boston, MA: RStudio, Inc.

Sandoval, M., Leclerc, J. A., & Gómez, R. L. (2017). Words to Sleep On: Naps Facilitate Verb Generalization in Habitually and Nonhabitually Napping Preschoolers. *Child Development*, *88*(5), 1615–1628. doi: 10.1111/cdev.12723

Schwab, J. F., & Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *WIREs Cognitive Science*, *7*(4), 264–275. doi: 10.1002/wcs.1393

Shi, J., Gu, Y., & Vigliocco, G. (2023). Prosodic modulations in child-directed language and their impact on word learning. *Developmental Science*, *26*(4), e13357. doi: 10.1111/desc.13357

Smolen, P., Zhang, Y., & Byrne, J. H. (2016). The right time to learn: Mechanisms and optimization of spaced learning. *Nature Reviews Neuroscience*, *17*(2), 77–88. doi: 10.1038/nrn.2015.18

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532. doi: 10.1016/j.dr.2007.06.002

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine Language: Speech Directed to Infants During Home Activities. *Child Development*, *90*(6), 2135–2152. doi: 10.1111/cdev.13089

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First Language*, *4*(12), 197–211. doi: 10.1177/014272378300401202

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. doi: 10.1177/0956797613488145

Wickham, H. (2016). *Ggplot2: Elegant Graphics for Data Analysis.* New York: Springer-Verlag New York.

Xu, D., Yapanel, U., & Gray, S. (2009). *Reliability of the LENA Language Environment Analysis System in young children's natural home environment* (Technical Report lTR-05-2). Boulder, CO: LENA Research Foundation.