Bursty, irregular speech input to children predicts vocabulary size

Margaret Cychosz*[1], Rachel R. Romeo[2,3], Jan R. Edwards[2], and Rochelle S. Newman[2]

[1]Department of Linguistics, University of California, Los Angeles

[2]Department of Hearing and Speech Sciences, University of Maryland, College Park

[3]Department of Human Development and Quantitative Methodology, University of Maryland-College Park

Author Note

*Author to whom correspondence should be addressed: Margaret Cychosz, 2101H Campbell Hall, University of California, Los Angeles, 90095. Email: mcychosz4@ucla.edu.

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), acquiring English in the United States, 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

Research Highlights

1. We computed 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.

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; Merz et al., 2020; Romeo et al., 2018; Rowe, 2012). This link between speech input and language development is apparent throughout childhood, and is observed in children as young as 18 months: children who hear more speech from caregivers process words faster (Hurtado et al., 2008), produce more complex grammatical structures (Huttenlocher et al., 2010), and have stronger pre-literacy skills and activation in critical language-related regions of the brain well into middle childhood (Merz et al., 2020; Sheridan et al., 2012). Those who hear more de-contextualized language or rare words from caregivers also grow larger vocabularies (Beals, 1997; Rowe, 2012). 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, 2016a; Soderstrom, 2007). Many characteristics of CDS have been proposed to convey benefits for children's language learning. For example, in toddlerhood, 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). Later, in early to middle childhood, CDS from central caregivers could function as a narration tool, and a method to introduce children to infrequent lexical items, as their vocabularies continue to grow (Beals, 1997; Rowe, 2012). Nevertheless, the majority of this work has been conducted over limited samples (generally middle- to upper-class North American or European), meaning that our conception of speech input to children has focused on how input occurs in these populations and, to the limited extent that more diverse samples are examined, how their speech input to children differs from the western, European model. When alternative definitions of input are employed in other cultural contexts, quantitative and qualitative aspects of speech input change substantially (Scaff et al., 2024) and constructs such as overheard speech are more important for developmental outcomes (Foushee & Srinivasan, 2024). So while work on CDS has demonstrated that not all speech heard in the home is equally beneficial for children's language learning—CDS is especially beneficial, more so than overheard speech—an important caveat remains that these findings were drawn from a limited range of North American/European cultures who are not broadly representative of child language socialization.

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; Merz et al., 2020; Romeo et al., 2018; Schwab & Lew-Williams, 2016a). Additional work has demonstrated how various qualitative dimensions of the input—diverse vocabulary, decontextualized language, extra-linguistic cues—also relate to vocabulary (assessed longitudinally 24-36 months: Hirsh-Pasek et al. [2015], 14-18 month-olds: Cartmill et al. [2013], 18-54 months: Rowe [2012], and in 60-month-olds: Beals [1997]).

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 (Hurtado et al., 2008; Weisleder & Fernald, 2013). They are also presented with more word learning opportunities via the use de-contextualized language or more infrequent words in the input directed to them (Beals, 1997; Rowe, 2012). These aspects of speech input explain how CDS aides in word learning throughout early and middle childhood.

Previous research designs in this area have tended to take short samples, or collapsed over entire days (see Section 1.1 below). As such, our understanding of the dynamics of language input has been limited 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, independent play, and reading (see Tamis-LeMonda et al. [2019], Mendoza and Fausey [2021], and Bang et al. [2022] for examination of time-of-day effects in infancy, and Casillas et al. [2021], Goldenberg et al. [2022], and Holme et al. [2022] for preschool to early middle childhood). 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 in infancy, see the Interaction Burst Hypothesis (Elmlinger et al., 2023) and Suarez-Rivera et al. (2023) who report on bursts of infant-caregiver physical proximity using video recordings).

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 early childhood 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]). Support for the hypothesis can also be seen when evaluating children up to age 5 years who are transitioning out of the nap stage: those children who still regularly napped showed stronger lexical consolidation following a nap (no effect of napping was seen for children who had already transitioned out of napping) (Esterline & Gómez, 2021).

Additional support for the Burst Hypothesis comes from the joint attention literature. It has long been known that children who engage in more episodes of joint attention—caregiver-child coordinated attention on an object or event—develop larger vocabularies (evidence from infants: Tomasello and Todd [1983] and preschoolers: Eriksson [2019]; see Akhtar and Gernsbacher [2007] for overview). But there has likewise been a tradition in this line of work to examine the *distribution* of joint attention episodes, for example differentiating between the quantity versus duration of episodes between a caregiver and child (Rudd, Cain, & Saxon, 2008) or the occurrence of episodes across different contexts (Adamson, Bakeman, & Deckner, 2004). Most recently, Abney et al. (2017) assessed joint attention bouts between caregivers and their 9-month-olds during 5-minute interactions. The authors carefully distinguished between the *quantity* (bouts/minute) versus *distribution* (duration of bouts) within each interaction. Shorter joint attention bouts throughout the interaction, especially those <2 seconds, were predictive of vocabulary size 3 and 6 months later (Abney et al., 2017). Caregiver-child dyads with shorter average bout durations also had a larger quantity of bouts overall. Thus, the temporal dynamics of joint attention are predictive of children's language outcomes, notably vocabulary size, *beyond mere quantity.* This line of work within the joint attention literature has demonstrated that quantity and distribution of caregiver-child interaction reflect different underlying constructs during language acquisition, and suggests that these components could be teased apart in studies on language input as well to explain variability in vocabulary outcomes.

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 hear 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 from infancy to middle childhood (Hoff, 2003; Merz et al., 2020; Rowe, 2012), but critically, the bursts would dose the input out in ways that better accommodate the time that children require to consolidate new linguistic information. As such, this approach assumes that speech events within a child's day take a heavy-tailed or Pareto distribution that reflects bursts of intense activity followed by longer bursts of inactivity. Indeed, such a distribution of discrete event activity has been found for innumerable natural, social, and human behavioral phenomena including electronic activity (internet traffic, correspondence, print job submission; Barabási [2005] Harder and Paczuski [2006]) and natural disasters (Bak, Christensen, Danon, & Scanlon, 2002). In the current study, we draw from this work and instantiate burstiness as both the height of each burst (how many words per unit of time) and the duration of each burst (how many sequential bursts in a row). We elaborate further upon these definitions of burstiness, and their justification, in Section ."

1.1 Limitations of previous work on children's language socialization

Despite what has now been decades of work linking dimensions of CDS and children's language outcomes such as vocabulary size, 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) (Sheridan et al., 2012), 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 or words/hour) (Merz et al., 2020; Romeo et al., 2018), or only smaller samples from the overall observation were analyzed (Bergelson et al., 2019). These methods and the lack of appropriate speech recognition tools have made it difficult to assess the ebbs and flows of speech input to children throughout entire days at a large scale.

Another limitation of previous work on language socialization is its highly homogeneous demographic samples, leading to research findings that may not be applicable to different 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; Sperry, Sperry, & Miller, 2019). 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?

1a. 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 CDS quantity alone.

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

2a. We anticipate that burstiness of CDS will be positively associated with receptive vocabulary size, and will be a stronger correlate of vocabulary than CDS quantity alone.

Our approach allows us to robustly put the Burst Hypothesis to the test, and simultaneously begin addressing some 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). This dataset vastly exceeds the size of most home-based language development studies. 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 in the United States to make our results more robust and representative of families from different backgrounds. It is important to acknowledge, however, that our sample is still quite skewed (towards higher SES), and lacking representation of some racial and ethnic groups. For example, the percentage of Hispanic (10.96%) and Asian-identifying children (1.71%) within our sample is well below the national averages (19.1 and 6.3%, respectively) (Bureau, 2020). Thus, this dataset is a first step to diversifying language socialization studies. We carefully acknowledge the limitations of our dataset in our interpretation of the data.

We chose the relatively wide age range of 2 to 7 years (with an emphasis on 2-6 years; only 4 recordings came from 7-year-olds) because it is of substantial theoretical importance to understand how not just the quantity, but the *distribution* and *characteristics* of CDS change as children age. Much of the field's current interest in understanding the roles and limitations of CDS for language development revolves around how and when children learn from CDS versus overheard speech. To that end, there is an interesting line of more traditional laboratory work in older children examining novel word learning outcomes in overheard versus directed speech learning contexts (e.g. Foushee et al. [2021]). And much of the cross-cultural work that has focused on expanding definitions and conceptions of speech input to children has additionally examined children encompassing or surpassing this age range (e.g. 0;6-4;10 months: Scaff et al. [2024]; birth-11;0: Cristia et al. [2019]). This age range likewise transcends the transition to formal school (for some children; many children within this range in the United States would have been attending preschool for years), when the impact of the home language environment could be expected to differ. We carefully consider these possibilities in our interpretation of the data and evaluate interactions of our main effect with age in our statistical modeling.

3 Methods

3.1 Participants

Participants were N=292 children (133 girls and 159 boys), exposed to American English (N=84 [28.77%] children were acquiring one additional language and/or variety of American English such as African-American English; 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. N=48 children were attending elementary school at the time of study enrollment. 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.

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. 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. However, 36.99% of mothers had a graduate degree, so despite our attempts at diversification, our sample is still skewed in this way. See Supp. Materials I for tables de-aggregating SES, race, and ethnicity.

Since one of our hypotheses focuses on age effects, we wanted to ensure that our data were robust to potential inadvertent confounds of age with SES. We fit a linear model to evaluate a potential relationship between maternal education attainment and child age and found no

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.

significant effect (β=-0.06, t=-2.67, p=.08). To further ensure the robustness of our results, we additionally included maternal education as a covariate in all of our models before evaluating the main effect of interest; see Results.

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. Our statistical modeling takes into account repeated measures within families over time. See Table 2 for descriptive statistics of recordings and Figure 1 for a histogram of observation hours by age.

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.

3.3 Deriving child-directed speech estimates

We first removed all recordings <5 hours (N=7) since previous validation work has demonstrated that the LENA system performs markedly worse on recordings this length and

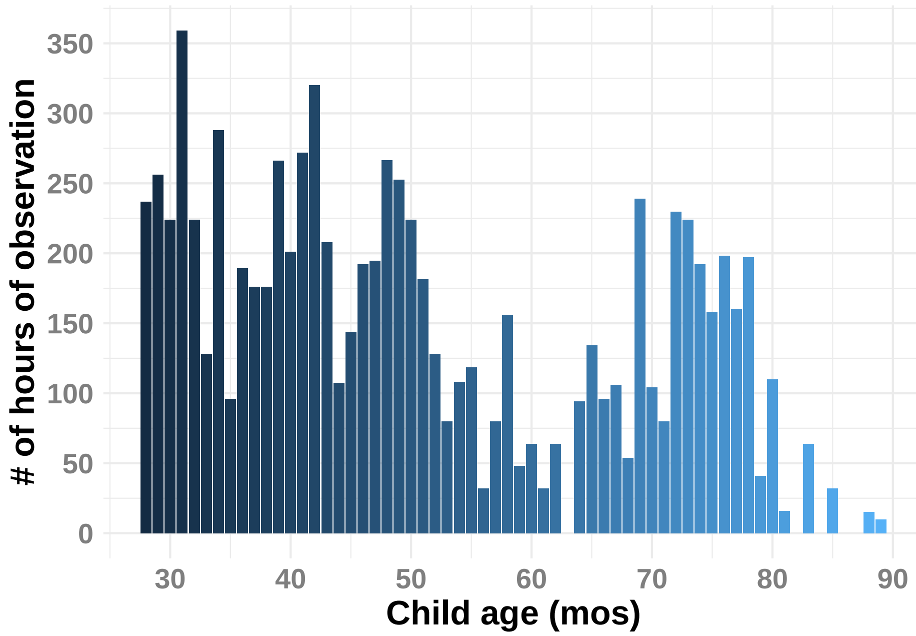

Figure 1. Total hours of observation via child-centered daylong audio recordings, by child age (in months).

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 bursty 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)). 30-second bursty segments were generated by exploding each CDS epoch from the recording by 10. See Section 3.3 for additional detail. *Although the recording duration spans a large range, only N=11 recordings (1.98% of the sample) were <10 hours.

shorter (Xu et al., 2009). Two of these removed recordings came from the same child. However, that child completed 4 recordings total, so the child still had 2 recordings remaining in the dataset. 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/`).

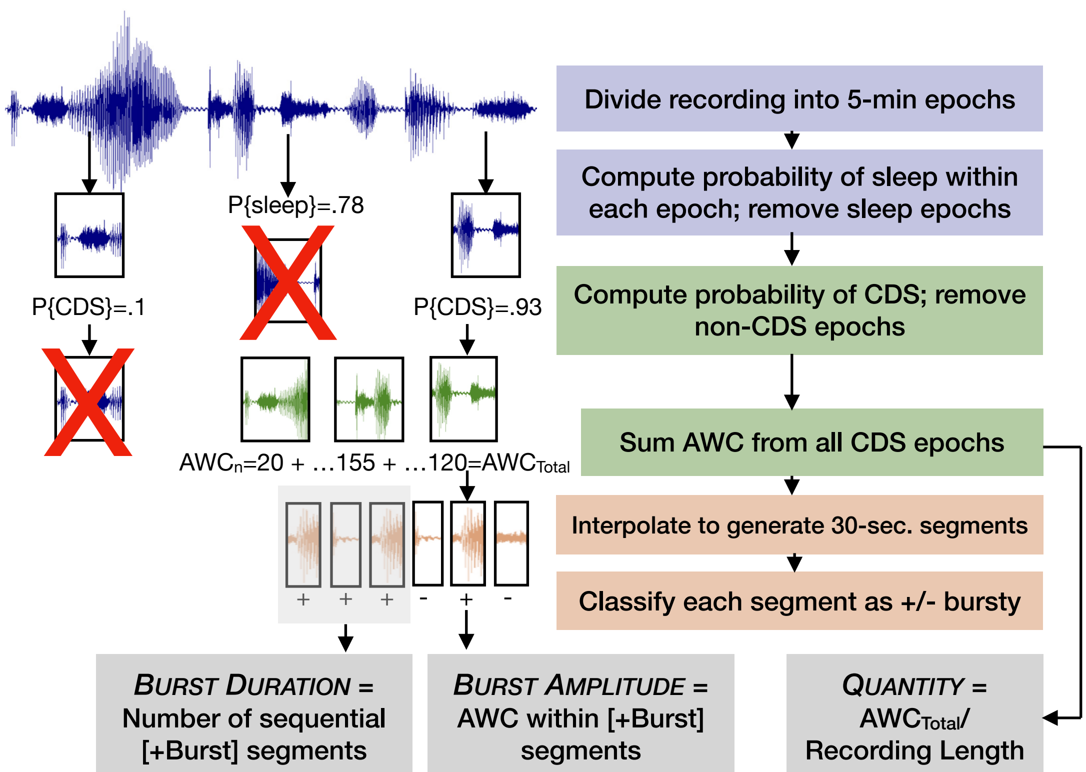

Figure 2. Flowchart illustrating audio processing steps and methods used to generate dose (quantity) and distribution (burst) metrics. CDS=child-directed speech. AWC=adult word count.

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 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). (Note that manual coding criterion for Bang et al. was .7 (i.e. 70% of a 5-minute segment had to contain CDS) while the criterion for the classifier was .5 (i.e. the classifier distinguished between 5-minute segments with >50% CDS from those with at least 50% ODS. This discrepancy in manual versus automated classification criteria could deflate the reported sensitivity and specificity metrics. In other words, the classifier might actually perform better than initially reported.)

Because the original Bang et al. (2023) classifier was trained over data from infants aged 17- to 28-months, we manually validated the classifier's performance over a subset of recordings from children in the current study (N=8 children aged, 3-, 4-, 5-, and 6-years-old, for a total of 32 recordings). We closely followed the methodology of Bang et al. to make sure that our results were comparable with that work. First, we divided each of the recordings into 5-minute segments. Next, we identified the portions of "extended" sleep in the recording using vocal activity (AWC counts) as a clue: 2 consecutive 5-minute segments with low AWC counts (AWC=0) suggested the presence of sleep which was then listened to and coded manually. As in Bang et al., coders listened to the portions prior to and following the low AWC count sections to verify when the sleep started and stopped. Then, again following Bang et al., we sorted the AWC counts in descending order and coders manually annotated each 5-minute segment as containing at least 70% target child-directed speech (tCDS) (or at least inclusive of the target child) versus other-directed speech (other children, excluding the target child, and adults). Stated in another way: a segment had to contain at least 70% tCDS to be classified as such. We continued this process until 12 5-minute segments of tCDS had been identified in each recording. This is the only way in which our methods differed from Bang et al.: they manually coded 10-minute segments as tCDS/ODS and then divided those segments into 5-minute segments for the validation; we coded 5-minute segments from the beginning. For sleep, our manual annotation showed extremely high sensitivity (0.9671) and specificity (0.9591) (overall accuracy=0.9635). For tCDS, we found comparably high sensitivity (72.75) and specificity metrics (77.78) for tCDS identification (overall accuracy= 74.22) compared to Bang et al. (Note that metrics for sleep were derived over just 16 children, 4/age group.)

Having validated the classifier, we then proceeded with computing CDS computation. 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*. Burst amplitude is a construct that reflects the number of words uttered within lexical spikes, which we refer to as "bursty segments," and thus captures how spiky the child's input is. (As such, burst amplitude is completely distinct from the physical concept of sound pressure level or sound intensity that is commonly measured in studies on audio processing, music, and speech.) However, burst amplitude does not capture how temporally *sustained* bursts in a child's input might be. Therefore, we additionally employ burst duration which is a construct that reflects the number of sequential bursty segments. In the section below, we provide extensive definitions and explanations of these measures.

To compute burst amplitude and burst duration, we employed a variety of amplitude threshold algorithms, a genre of algorithms that detect events in a signal exceeding pre-determined thresholds. Such approaches are commonly employed within neuroscience to, for example, detect synaptic events via bursts of oscillations of interest (Duguid et al., 2012; Hwang & Copenhagen, 1999). The measures were computed as follows (see also 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 mean number of CDS words from the entirety of each daylong recording. This was done dynamically so each recording had its own mean 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 mean. 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 *total* words within each bursty segment (not just those words beyond the bursty limit) and calculated the average amplitude over all bursts for the modeling. Each child thus had one overall burst amplitude measure.

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. Each child thus had one overall burst duration measure.

An alternative measure of burstiness, employed in disciplines such as network science, instead computes burstiness as an inter-event time distribution often referred to as a 'burstiness parameter' (Goh & Barabási, 2008; Kim & Jo, 2016). Such an approach to burstiness quantification applied to CDS could instead have quantified the time lag between each word or utterance spoken by an adult to the child, with shorter time lags between individual words or utterances amounting to spikier bursts. In theory, computation employing an amplitude threshold algorithm versus the inter-event time distribution technique would not impact the measure of burstiness. However, the data that we employ here is not sufficiently fine-grained to provide timestamps of each individual word or utterance spoken by an adult to the target child—30-second segments are instead employed—making it difficult, in practice, to carry out the inter-event time distribution approach at this stage.

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 `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

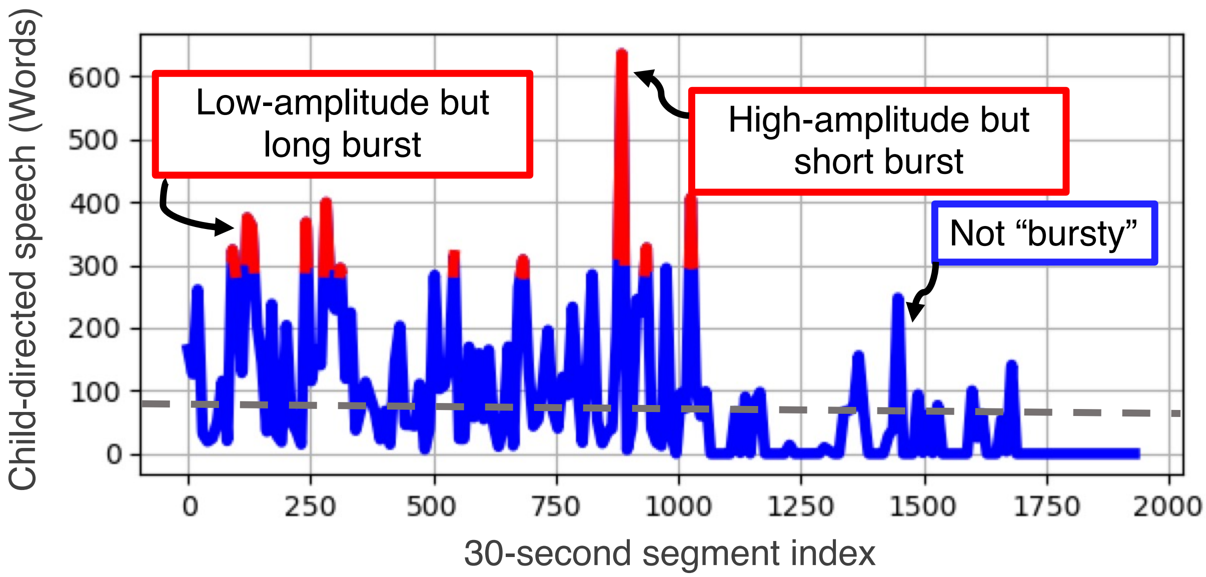

Figure 3. Illustration of measures of child-directed speech "burstiness" metrics from a single daylong audio recording. Red=30-second segments classified as +bursty spikes. Blue=segments classified as -bursty. Dashed line=mean child-directed speech word count from the recording. Burst amplitude refers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

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. For the former, each additional word token has the potential to alter the meaning, comprehensibility, or salience of a caregiver's utterance; for the latter, the addition or subtraction of a few tokens would not have as pronounced an impact on the overall understanding or significance.

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 (Figure 4).

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 refers to the number of words from adults in "bursty" segments and burst duration to the number of sequential "bursty" segments.

	Burst Amplitude Model	Burst Duration Model	Input Quantity Model
Intercept	β=-0.01	β=0.02	β=-0.02
	CI=(−0.05, 0.04)	(−0.02, 0.07)	(−0.09, 0.05)
	t = −0.40	t = 0.90	t = −0.62
	p = 0.69	p = 0.37	p = 0.54
Age (mos)	β=−0.002	β=−0.005	β=−0.001
	(−0.004, −0.0001)	(−0.01, −0.003)	(−0.004, 0.002)
	t = −2.05	t = −4.23	t = −0.59
	p = 0.04*	p < 0.001***	p = 0.56
Gender:Female	β=0.03	β=−0.01	β=0.04
	(−0.04, 0.10)	(−0.08, 0.06)	(−0.06, 0.15)
	t = 0.82	t = −0.29	t = 0.84
	p = 0.42	p = 0.78	p = 0.40
Mat. Ed.	β=0.06	β=0.02	β=0.10
	(0.04, 0.08)	(−0.01, 0.04)	(0.07, 0.13)
	t = 5.50	t = 1.44	t = 6.36
	p < .001***	p = 0.16	p < .001***
Log Likelihood	−196.92	−198.48	−378.27
AIC	405.84	408.96	768.53
Adjusted R-squared for fixed effects	0.077	0.077	0.073

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

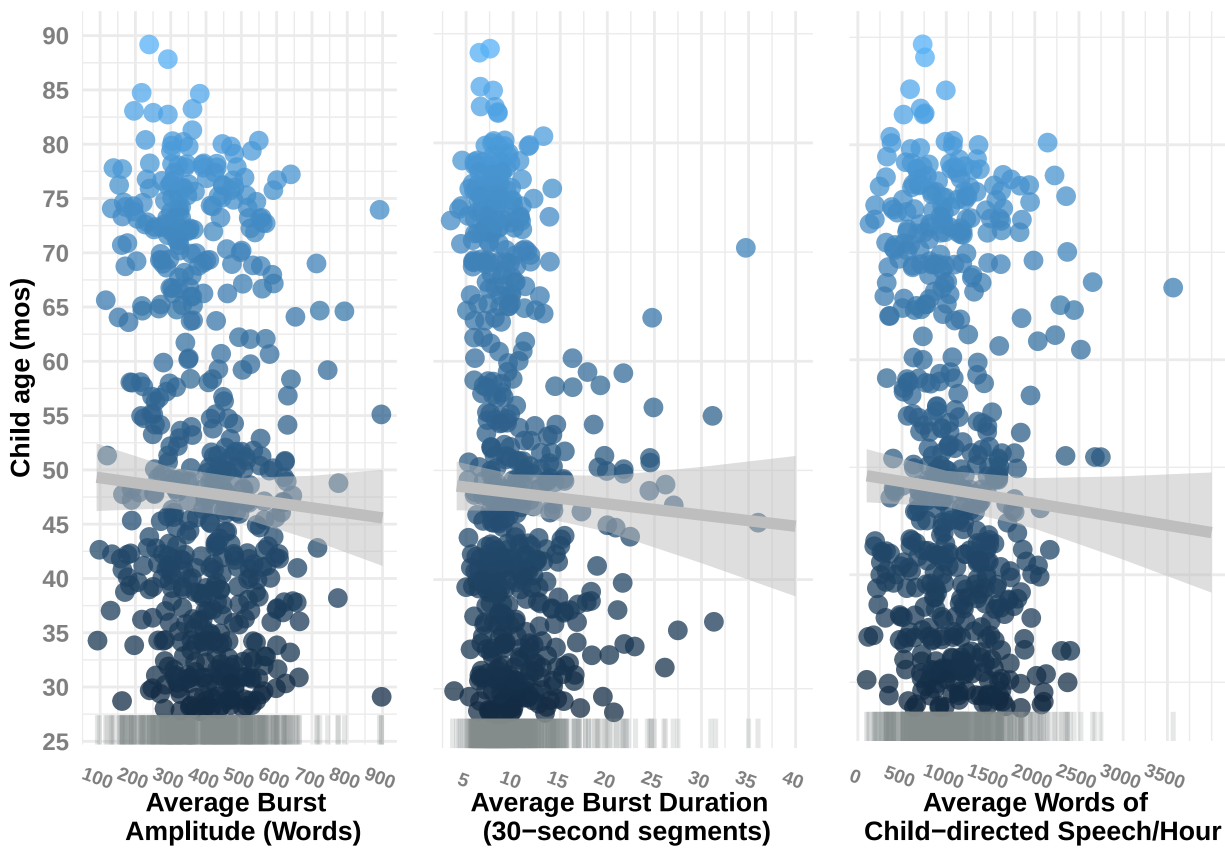

Figure 4. The relationship between input burstiness and child age. Each data point represents raw, non-transformed data from one daylong recording. Burst amplitude refers to the number of words from adults in "bursty" segments and burst duration to the number of sequential "bursty" segments. The light 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). Interactions of the main effects with Child Age did not significantly improve the model fits, suggesting that the effects of input distribution were consistent over this age range.

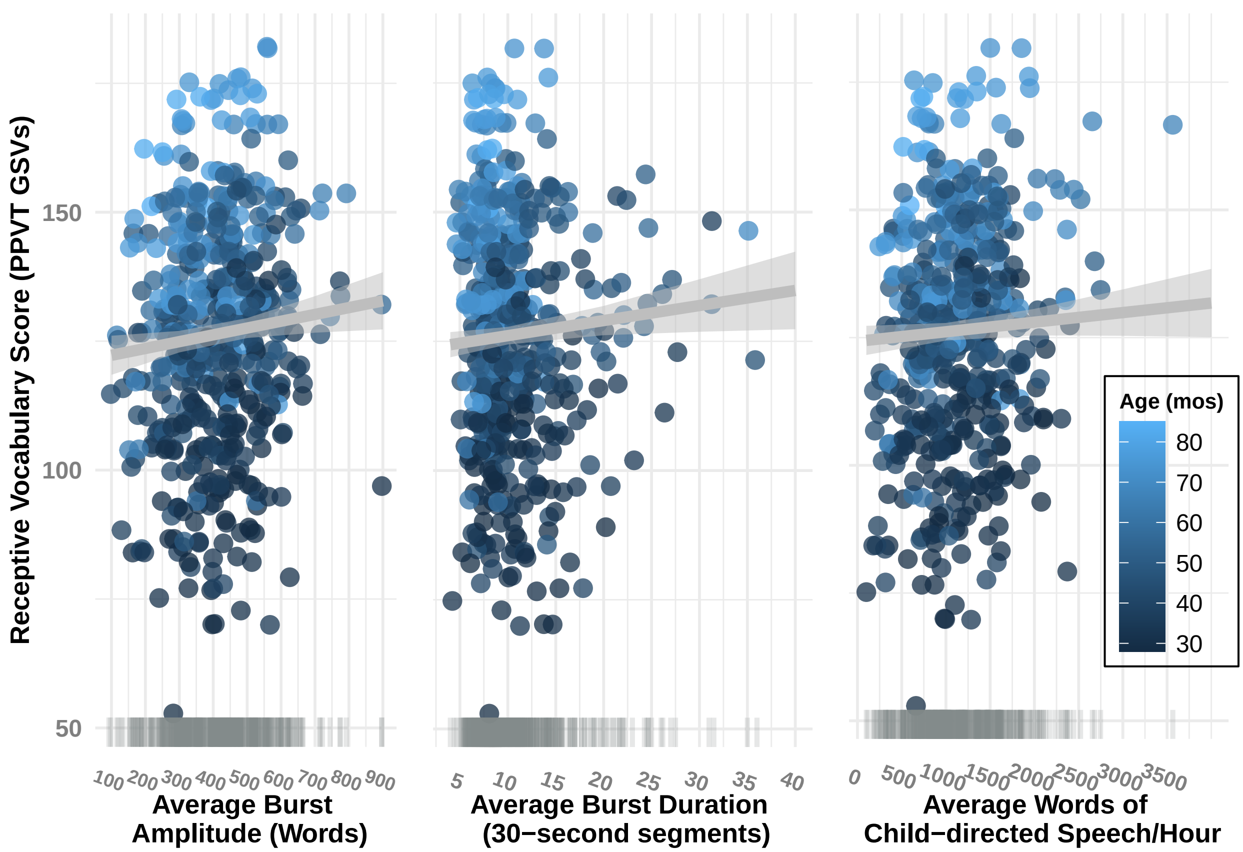

Figure 5. The relationship between input burstiness and receptive vocabulary size (PPVT-4 score). Burst amplitude refers to the number of words from adults in "bursty" segments and burst duration to the number of sequential "bursty" segments. Each data point represents raw, non-transformed data from 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.

Table 4

The effect of quantity and burstiness of child-directed speech on concurrent vocabulary scores. Burst amplitude refers to the number of words from adults in "bursty" segments and burst duration to the number of sequential "bursty" segments.

	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
Adjusted R-squared for fixed effects	0.585	0.590

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

4.3 Relationships with nap duration

In addition to the distribution of CDS throughout a given day, we also hypothesized a potential relationship between sleep and vocabulary size since sleep allows preschoolers to consolidate new cognitive and linguistic information (Giganti et al., 2014; He et al., 2020; Horváth et al., 2015; Kurdziel et al., 2013; Sandoval et al., 2017). This analysis was exploratory, not confirmatory. We limited the analysis to children's nap times, not overnight sleep (see Supp. Materials II for detail). We analyzed children under 42 months, as children above this age are unlikely to still be regularly napping. Estimated nap durations ranged from 5-55 minutes (M= 28.65, SD=16.47) for the N=115 remaining recordings (N=108 children). Nap duration was positively related to concurrent vocabulary (β=0.15, p=.03, controlling for child age, gender, and SES).

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 TV). 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.

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 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 diverse in some ways (though certain demographics such as maternal education level attainment are still over-represented). Thus, the sample takes an important first step towards a robustness in research results that has been missing in much language socialization research. Results showed that (1) child age correlated more strongly with input distribution (burst amplitude and duration) than input quantity and (2) input distribution predicted concurrent vocabulary above and beyond the quantity of input. Thus, these findings challenge the traditional emphasis on the sheer quantity of language input in early childhood as the primary driver of language development. Instead, they highlight the importance of the distribution of input, suggesting that it's not just about how much language a child hears, but also the temporal distribution of how that language is presented to them.

Based on effect sizes, burst duration and amplitude were equally strong correlates with child age. The measures were also both equally 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 the 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, 2016b) as opposed to hearing the same word repeated over a longer period (Yurovsky et al., 2013). In terms of mechanisms, it is possible that such lexical bursts could also reference distinct exemplars in children's environments (Perry et al., 2010), or occur in similar sentential frames (Mintz, 2003)—both 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).

It is important to acknowledge how these results could be expected to generalize to other cultural and language learning settings. The children examined here have a generally “WEIRD” upbringing (Western Educated Industrial Rich and Democratic) which, in the context of language socialization and development typically means nuclear families consisting of 1-2 central caregivers, with 1-2 additional siblings (or less), living in geographically isolated housing units, who are exposed to speech in primarily indoor environments. In comparison, children in some other socio-cultural settings may be exposed to a greater number of distinct speakers in their everyday lives if childcare is distributed between multiple community members. Or, for children from larger families, a larger proportion of language input could originate from older siblings than adult caregivers (Loukatou et al., 2022) or could occur in overlap instead of speech directed in isolation to an individual child (Scaff et al., 2024). How might the Burst Hypothesis play out in some non-WEIRD settings? (Non-WEIRD settings, it should be mentioned, are far from uniform and differ from one another just as much as WEIRD vs. non-WEIRD settings.)

Take the example of a child exposed to similar *amounts* of overall speech as a child in the current U.S. sample, but with speech that is (1) less frequently directed right to the child and (2) more evenly distributed over the course of the day because there are more unique speakers. The distribution of speech input would have smaller peaks and valleys, and thus might be considered less “bursty.” However, adapting the Burst Hypothesis to this setting may additionally require a re-formulation of our conception of language input (Scaff et al., 2024). Perhaps the overall speech exposure is less bursty, but speech directed to the target child and/or their sibling(s) still occurs in bursts. Alternatively, perhaps in this setting input from individual adult caregivers occurs in bursts, but input from siblings does not and this dimension could have important explanatory power for development in this cultural context.

The bottleneck to examining the Burst Hypothesis, and other input-related hypotheses for language development, in additional samples is no longer a lack of diverse, naturalistic child language corpora, but the time and linguistic expertise required to, for example, code for different speech registers (Bergelson et al., 2019; Cychosz et al., 2021). However, by broadening our definition of "learnable" linguistic input beyond CDS, which recent work has started to do (Scaff et al., 2024), researchers in this area could actually start overcoming this issue. For example, reliable vocal activity and speaker (gender) diarization algorithms now exist for child-centered audio data, which is sufficient to compare different distributions of speech input over a child's day. In short: we may be looking for (certain types of Westernized) CDS—a significant time and energy investment—when measuring distribution of the input over all speech could be sufficient to examine constructs such as the Burst Hypothesis in other, more diverse samples.

The combination of multiple speech recognition systems employed here allowed for more precise and detailed analysis of language input than previously possible, and opens up new avenues for investigating the mechanisms underlying language learning. In particular, 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. It remains somewhat unclear how language input shapes early phonological development (Cristia, 2020) (with some early evidence of SES effects upon perceptual outcomes within the first year of life [Singh et al. 2023]). The effect of additional dimensions of language input for perceptual outcomes, including the distribution of input examined here, could be an important next step.

We were also interested in extending the definition of bursty input to include bursts in conversational turns between the child and caregiver(s) (for full analysis see Supp. Materials III), and adult speech that did not differentiate between CDS and ODS (Supp. Materials IV). In our analyses, we replicated the "Bursty" effect: children with burstier conversational turns had larger vocabularies. However, there was no relationship between burstiness in overall input (CDS+ODS) and vocabulary. This finding reinforces the need to extend our understanding of input distribution to more diverse samples, where overheard speech may prove more important for language learning. As mentioned above, in such samples, bursty input in *any* register—child-directed or overheard—may be related to vocabulary outcomes while in this sample only bursty CDS was related.

We conducted these supplementary analyses in part 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 conclude that there was a positive association between nap duration and vocabulary size (Section 4.3). 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.

Finally, this work helps elucidate some of the complex relationships between SES and language development by highlighting the role of input distribution as a potential mediator of the relationship. By focusing on input distribution, in addition to quantity, this work highlights a potential avenue through which socioeconomic disparities in language development may manifest (Romeo et al., 2018; Rowe, 2008). It suggests that the structure of language input, instantiated as burst amplitude and duration here, may play a mediating role for the effects of SES on language outcomes. Understanding this relationship is essential for addressing disparities in language skills often observed between children from different socioeconomic backgrounds upon school entry. If input distribution, as suggested by the Burst Hypothesis, is a factor for vocabulary development in samples such as the one examined here, interventions aimed at improving language outcomes among children from socioeconomically-disadvantaged backgrounds may benefit from focusing on the distribution of language input rather than just the quantity. This could involve providing caregivers with guidance on how to engage in more interactive and responsive language interactions with their children, promoting more varied language experiences that mirror the patterns found to be beneficial in the study. However, much more work needs to be conducted on the temporal distribution of speech input before any such policy recommendations or changes could be made.

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.

Author Contributions

Conception: MC & RN. Data collection: JE & RR. Corpus organization: MC. Data analysis: MC. Manuscript writing: MC. Manuscript editing: all. Funding: all.

Ethics approval statement

All study procedures were approved by the relevant institutional review boards. Caregivers of all participants gave their informed consent before study enrollment.

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.*

Adamson, L. B., Bakeman, R., & Deckner, D. F. (2004). The Development of Symbol-Infused Joint Engagement. *Child Development*, *75*(4), 1171–1187. doi: 10.1111/j.1467-8624.2004.00732.x

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

Bak, P., Christensen, K., Danon, L., & Scanlon, T. (2002). Unified Scaling Law for Earthquakes. *Physical Review Letters*, *88*(17), 178501. doi: 10.1103/PhysRevLett.88.178501

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

Barabási, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, *435*(7039), 207–211. doi: 10.1038/nature03459

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

Beals, D. E. (1997). Sources of support for learning words in conversation: Evidence from mealtimes. *Journal of Child Language*, *24*(3), 673–694. doi: 10.1017/S0305000997003267

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

Bureau, U. C. (2020). *QuickFacts: United States. United States Census Bureau.* (Tech. Rep.).

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. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, 1–23. doi: 10.1017/S0305000920000549

Cristia, A. (2020). Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition. *Developmental Review*, *57*, 100914.

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-Directed Speech Is Infrequent in a Forager-Farmer Population: A Time Allocation Study. *Child Development*, *90*(3), 759–773. doi: 10.1111/cdev.12974

Cychosz, M., Villanueva, A., & Weisleder, A. (2021). Efficient estimation of children's language exposure in two bilingual communities. *Journal of Speech, Language, and Hearing Research*, *64*(10), 3843–3866. doi: 10.1044/2021_JSLHR-20-00755

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

Duguid, I., Branco, T., London, M., Chadderton, P., & Häusser, M. (2012). Tonic Inhibition Enhances Fidelity of Sensory Information Transmission in the Cerebellar Cortex. *The Journal of Neuroscience*, *32*(32), 11132–11143. doi: 10.1523/JNEUROSCI.0460-12.2012

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

Eriksson, F. (2019). *Joint attention and language acquisition* (Unpublished master's thesis). Stockholm University, Stockholm, Sweden.

Esterline, K., & Gómez, R. L. (2021). The Role of Sleep in Retention of New Words in Habitually and Non-Habitually Napping Children. *Brain Sciences*, *11*(10), 1320. doi: 10.3390/brainsci11101320

Foushee, R., & Srinivasan, M. (2024). Infants who are rarely spoken to nevertheless understand many words. *Proceedings of the National Academy of Sciences*, *121*(23), e2311425121. doi: 10.1073/pnas.2311425121

Foushee, R., Srinivasan, M., & Xu, F. (2021). Self-directed learning by preschoolers in a naturalistic overhearing context. *Cognition*, *206*, 104415. doi: 10.1016/j.cognition.2020.104415

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

Goh, K.-I., & Barabási, A.-L. (2008). Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, *81*(4), 48002. doi: 10.1209/0295-5075/81/48002

Goldenberg, E. R., Repetti, R. L., & Sandhofer, C. M. (2022). Contextual variation in language input to children: A naturalistic approach. *Developmental Psychology*, *58*(6), 1051–1065. doi: 10.1037/dev0001345

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

Harder, U., & Paczuski, M. (2006). Correlated dynamics in human printing behavior. *Physica A: Statistical Mechanics and its Applications*, *361*(1), 329–336. doi:

10.1016/j.physa.2005.06.079

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

Holme, C., Harding, S., Roulstone, S., Lucas, P. J., & Wren, Y. (2022). Mapping the literature on parent-child language across activity contexts: A scoping review. *International Journal of Early Years Education*, *30*(1), 6–24. doi: 10.1080/09669760.2021.2002135

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

Hwang, T. N., & Copenhagen, D. R. (1999). Automatic detection, characterization, and discrimination of kinetically distinct spontaneous synaptic events. *Journal of Neuroscience Methods*, *92*(1-2), 65–73. doi: 10.1016/S0165-0270(99)00095-3

Kim, E.-K., & Jo, H.-H. (2016). Measuring burstiness for finite event sequences. *Physical Review E*, *94*(3), 032311. doi: 10.1103/PhysRevE.94.032311

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.

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2022). Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language*, *49*(6), 1173–1192. doi: 10.1017/S0305000921000623

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

Merz, E. C., Maskus, E. A., Melvin, S. A., He, X., & Noble, K. G. (2020). Socioeconomic Disparities in Language Input Are Associated With Children's Language-Related Brain Structure and Reading Skills. *Child Development*, *91*(3), 846–860. doi: 10.1111/cdev.13239

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. 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.

Rudd, L. C., Cain, D. W., & Saxon, T. F. (2008). Does improving joint attention in low-quality child-care enhance language development? *Early Child Development and Care*, *178*(3), 315–338. doi: 10.1080/03004430701536582

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

Scaff, C., Casillas, M., Stieglitz, J., & Cristia, A. (2024). Characterization of children's verbal input in a forager-farmer population using long-form audio recordings and diverse input definitions. *Infancy*, *29*(2), 196–215. doi: 10.1111/infa.12568

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

Schwab, J. F., & Lew-Williams, C. (2016b). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*(6), 879–886. doi: 10.1037/dev0000125

Sheridan, M. A., Sarsour, K., Jutte, D., D'Esposito, M., & Boyce, W. T. (2012). The Impact of Social Disparity on Prefrontal Function in Childhood. *PLoS ONE*, *7*(4), e35744. doi: 10.1371/journal.pone.0035744

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

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy*, *28*(4), 708–737. doi: 10.1111/infa.12545

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

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019). Language Does Matter: But There is More to Language Than Vocabulary and Directed Speech. *Child Development*, *90*(3), 993–997. doi: 10.1111/cdev.13125

Suarez-Rivera, C., Pinheiro-Mehta, N., & Tamis-LeMonda, C. S. (2023). Within arms reach: Physical proximity shapes mother-infant language exchanges in real-time. *Developmental*

Cognitive Neuroscience, *64*, 101298. doi: 10.1016/j.dcn.2023.101298

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.

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive Processes in Cross-Situational Word Learning. *Cognitive Science*, *37*(5), 891–921. doi: 10.1111/cogs.12035