Comparing word diversity versus amount of speech in parents' responses to infants' prelinguistic vocalizations

Steven L. Elmlinger
Psychology Department
Cornell University
Ithaca, New York - USA
sle64@cornell.edu

Deokgun Park
Computer Science & Engineering
University of Texas
Arlington, Texas – USA
deokgun.park@uta.edu

Jennifer A. Schwade
Psychology Department
Cornell University
Ithaca, New York - USA
jas335@cornell.edu

Michael H. Goldstein
Psychology Department
Cornell University
Ithaca, New York – USA
mhg26@cornell.edu

***Abstract*—Our prior research posits that the prelinguistic vocalizations of infants may elicit caregiver speech which is simplified in its linguistic structure. Caregivers' speech clearly contributes to infants' development; infants' communicative and cognitive development are predicted by their ambient language environment. There are at least two sources of variation in infants' language environment: the number and the diversity of words infants hear. We compare the change in total number of words (tokens) to the diversity of words against one another. Distributions of words of differing sizes are difficult to compare to one another because the size of the distribution largely determines the word diversity of the distribution. A novel approach to minimizing the challenges of comparing distributions of words is applied to data which were previously reported. We also conducted a new simulation study to estimate the probability that these results are expected by chance. We found that the linguistic structure of caregivers' responses to infants' prelinguistic vocalizations has fewer word types as compared to infant-directed but non-contingent speech. Our new method shows that contingent word distributions remain simplified as the number of total words sampled increases. By vocalizing, infants elicit caregiver speech which is simpler in structure and may be easier to learn.**

Keywords—Parent-infant interaction, prelinguistic vocal production, conversational turn-taking, speech environment, simulation

I. Introduction

Caregivers' behavior, which is organized around the nascent vocalizations of their offspring, is crucial for communicative development. Vocal learning in songbirds [1], marmosets [2], and humans [3] is facilitated by social feedback from adults that is contingent on the immature vocalizations of offspring. An initial step in human infants' gradually developing vocal communication is the formation of expectations that their immature vocalizing reliably elicits social input [4,5]. No two caregivers talk to their children in exactly the same way and differences in infants' social environment influence the nature of their communicative development. Variability in the linguistic structure of parents' speech to prelinguistic infants predicts vocabulary growth [6]. Early communicative and language development is guided by the form and timing of caregivers' responsiveness [7] and infant-directed speech [8]. However, the role infants play in eliciting these behaviors from their parents is only just beginning to be investigated [9]. In particular, little attention has been paid to the linguistic patterning of caregivers' speech in response to infants' vocalizations.

New studies have found that infant vocalizations facilitate the production of more simplified talk from adult caregivers [10]. This response moves the complexity of caregivers' speech into a range that may facilitate infant learning. The lexical and syntactic structure of caregiver speech is simplified in response to infants' vocalizations. Caregivers uttered fewer unique word types, fewer words per utterance and higher proportions of utterances which contained only a sinlge word when talk was contingent on vocalizations. At present, it is unclear if this effect of simplification is stable when larger corpora (i.e., larger samples of talk) are analyzed. Ongoing efforts employ sophisticated techniques to distinguish genuine word diversity effects from effects due to the sample size of words among different corpora (i.e., samples of talk) [11]. The goals of the current work are to better understand how the simplification of talk (specifically lexical diversity) may scale with the size of the sample of talk, and to more precisely quantify the simplification of caregivers' contingent speech.

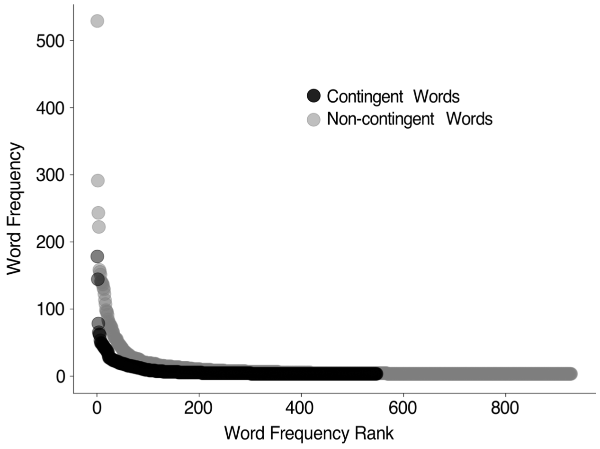

Figure 1. A word-rank by word-frequency graph of contingent and non-contingent words.

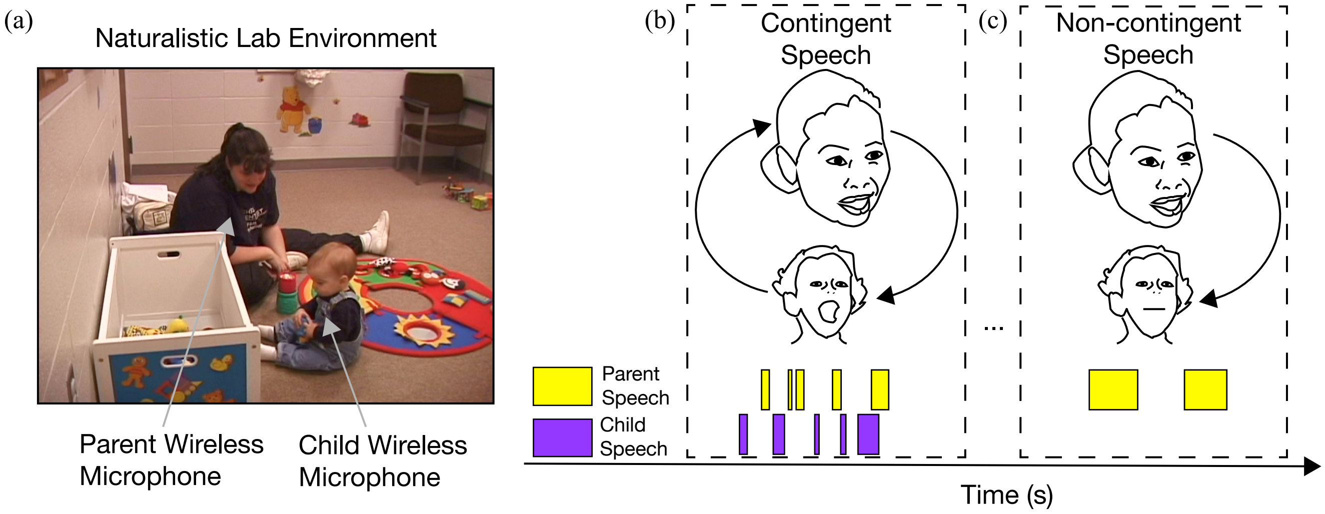

Figure 2. (a) Caregivers and infants were recorded wirelessly with microphones while they played with toys in the lab. Graphical representation of turn-taking data collated into categories of contingent caregiver speech (b) and non-contingent speech (c). Speech within turn-taking was categorized as contingent if it occurred within 2 seconds of infants' non-cry vocalization.

New challenges associated with interpretation and analysis of increasingly large-scale data collection efforts are being uncovered [12]. These challenges are being met with new techniques to address them. A central problem in understanding word diversity at scale is the fact that word frequency rank distributions are non-Gaussian (i.e., do not form a normal distribution). For example in Elmlinger et al., 2019, word ranks show that few words are very frequent and many words are infrequent (Figure 1, also see Piantadosi, 2014 for a detailed review of near-Zipfian distributions in natural language [13]). Heavy-tailed distributions such as these are characteristic of the objects encountered throughout infants' early visual experience, the labels of those objects infants hear, and the object-referent mappings children begin to understand first [14]. These distributions are not amenable to the summary statistics traditionally used in psychology because data-clustering around the center does not occur [15].

Derivations of caregiver type-token ratios (TTR), a ratio comparing the count of unique words (types) to the count of words (tokens), have recently been shown to be predictive of children's language outcomes [6, 16]. There are at least two issues surrounding TTR measures which remain understudied. First, because TTR nonlinearly scales with sample size, comparisons of TTR measures which were calculated from two different subsets (corpora) of words with unequivalent sample sizes are difficult to interpret [17]. Second, TTR measures are highly sensitive to the sample time period. Counting up the number of unique words from one corpus, for example, does little to illuminate the number of word types in another corpus which differs in total number of words. Initial attempts at addressing these issues are starting to yield new techniques [11]. Sampling techniques can give researchers insight into the sample size at which diversity of talk may be expected to generalize. Here we contribute an additional sampling technique alongside the central question posed in the present work: how does the word count relate to the number of unique words that infants' vocalizations elicit from their caregivers?

Recordings of naturalistic parent-infant interactions at large scales afford new opportunities for connecting the real-time structure of interaction to phenomena such as language learning which emerges over longer developmental time scales. The everyday learning context of infants' linguistic environment operates at multiple timescales. Learning happens in the moment when caregivers organize responses contingently around infants' babbling [18]. Learning also occurs incrementally over longer timescales through sequences of vocal turn-taking bouts [19]. The extent to which shorter and longer timescales can be analytically compared to one another has received little attention. The present study connects the multiple time scales through the analysis of caregiver speech during unstructured play sessions and simulates the change in word diversity as a function of talk which was either coordinated around infants' vocalizations or not.

II. Methods

A. Participants

In this study, thirty caregiver-infant pairs participated. The mean infant age was 9 months 20 days with a range of 9 months 12 days to 10 months 4 days. We recruited these subjects from birth announcements in advertisements and announcements in local newspapers. As a gift for their participation, families received a t-shirt or a bib. The participants reported in the present research were also reported on in previous studies [3,10, 20].

B. Apparatus

All recording sessions took place in a naturalistic environment which consisted of a twelve foot by eighteen foot playroom comprised of a toy box, toys and animal posters. This environment afforded infants free range to play and explore around the room as they wished. In the room were three digital cameras which were remote-controlled by experimenters capturing the video recordings. Infants wore overalls which concealed a wireless microphone (Telex

FLM-22; Telex Communications, Inc., Burnsville, MN) along with a transmitter (Telex USR-100). Before each session, wireless lapel microphones (Telex FLM-22) were affixed to caregivers' shirts. Caregiver microphones were connected to transmitters which were hidden in a pouch around their waist (Telex USR-100) (Figure 2a). Distinct audio channels were utilized in the recording of infants' vocalization and caregiver speech, respectively.

C. Procedure

Each participant engaged in 30-minute play sessions in the lab. During these sessions, parents were asked to play like they would at home, which resulted in unstructured free-play.

D. Speech Transcription

The speech that parents produced was completely transcribed (see [10] for reliability measures). If parents' utterances were separated by silence for longer than two seconds and/or if the pitch contours exhibited were terminal, they were segmented into separate utterances [21]. Following prior corpus transcription conventions, inflections were disregarded (*dog*, *dogs*, and *doggy* = *dog*) [22]. If parents' utterances occurred before two seconds after the offset of infants' vocalizations, then they were considered contingent utterances (Figure 2b). Responses which occurred after a two second time frame were considered non-contingent [23] (see Figure 2c). The mean and range F_0 of caregivers' contingent and non-contingent speech are consistent with previous descriptions of naturally produced infant-directed speech [8] (see Elmlinger et al., 2019 for more details). All caregiver utterances were directed at their infant. We excluded caregivers' production of sound effects and their responses to infant vegetative vocalizations such as coughs, cries, and fusses from the analyses.

E. Sampling procedure

The values derived by TTR track closely with the size of the sample and therefore cannot be interpreted through an isolated sample run [11]. To circumvent this constraint, we capture several measures of word diversity along a continuum of total word sizes. To understand the changes in the counts of unique word types as number of total words increases, we pool all of the speakers in our corpora together and randomly sample from contingent and non-contingent corpora in increasing increments of size from the respective corpora separately. Because we are mainly interested in changes in type-token relationships over a range of token sizes, we allow individual caregivers to vary naturally in the number of words they contribute to the pooled corpora.

Test Sample. We built samples through computing iteratively larger random samples starting at 100 words up to 3000 words, incrementing in steps of 100 words (Table 1). We sampled with replacement to ensure that every sample was drawn from the entirety of the word distributions in both corpora. We conducted the sampling 100 times for each sample size. We then counted the number of unique word types for each sample.

Control Sample Techniques. The contingent word corpus contains far fewer total words (n = 6,199) than the non-contingent word corpus (n = 19,548) (Table 1). In previous work we utilized size-matched random control samples to better understand how this difference in sample size may contribute to the TTR curves generated from the two

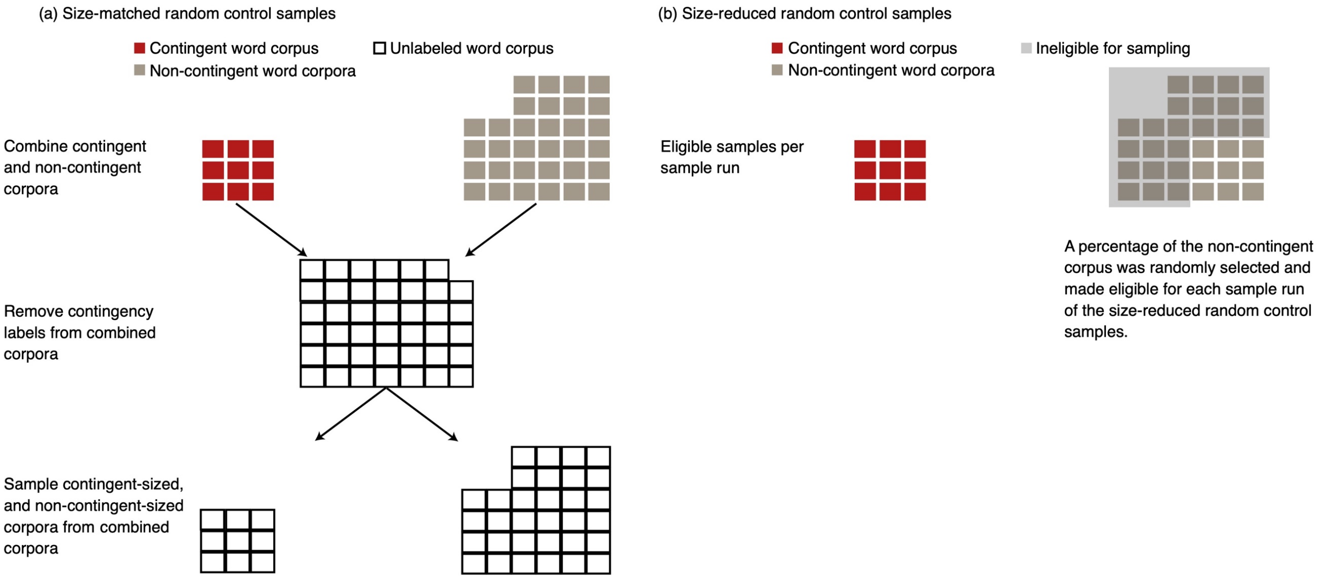
(a) Size-matched random control samples
🟥 Contingent word corpus
□ Unlabeled word corpus
▨ Non-contingent word corpora
Combine contingent and non-contingent corpora
Remove contingency labels from combined corpora
Sample contingent-sized, and non-contingent-sized corpora from combined corpora
(b) Size-reduced random control samples
🟥 Contingent word corpus
▨ Non-contingent word corpora
□ Ineligible for sampling
Eligible samples per sample run
A percentage of the non-contingent corpus was randomly selected and made eligible for each sample run of the size-reduced random control samples.

Figure 3. Control sample techniques. (a) Graphical depiction of size-matched random control samples where a single distribution is created from mixing contingent and non-contingent corpora into one unlabeled distribution of words [20]. To test whether we sample from the same underlying distribution when we generate TTR curves from original contingent and non-contingent corpora, we create new contingent-sized and non-contingent-sized corpora from a single distribution and derive control TTR curves to compare to the original TTR curves. If the curves differ from original to control samples, then this is evidence for the original samples not being drawn from the same distribution of words. (b) The extent of word diversity in the non-contingent corpora is investigated by plotting TTR curves of the word distribution after deeming a proportion of the corpora ineligible for sampling.

corpora [20]. In this approach, a single distribution is created from mixing contingent and non-contingent corpora into one unlabeled distribution of words (Figure 3a). The goal of this technique was to test whether we sampled from the same underlying distribution when we generated TTR curves from original (test) contingent and non-contingent corpora. We created new contingent-sized and non-contingent-sized corpora drawn from the unlabeled distribution and derived control TTR curves to compare to the test TTR curves. The new control corpora produced from this technique test the effects of sample size because the effects of contingent or non-contingent words themselves would be inherent in both samples as a result of randomization. This approach was tailored to provide evidence for or against our test samples having been drawn from the same distribution of words. If the control TTR curves differ from the test curves, then this is evidence for the test samples not being drawn from the same distribution of words. In previous work we found that the curves generated from the size-matched random control samples indeed yielded curves which differed from the curves generated from the test samples [20].

The extent of word diversity in the non-contingent corpora is investigated by plotting TTR curves of the word distribution after deeming a proportion of the corpora ineligible for sampling (Figure 3b). The goal of size-reduced random control sampling is to observe how the TTR curves of non-contingent words changes as we reduce the number of eligible words which generate the curve. We observed the curve generated when deeming 90, 60 and 30 percent of the word corpora eligible for sampling. Crucially, reducing the non-contingent word corpus to 30 percent of its original size matches approximately to the size of the contingent word corpus. Curve comparisons between the non-contingent 30 percent eligibility corpora and contingent corpora establishes whether reducing non-contingent corpora's eligible size to match that of the contingent corpora's size produces similar TTR curves.

TABLE I. STUDY SAMPLE SIZES

	Original corpora	*All sampling techniques*
Contingent	6,199	100 - 3,000
Non-contingent	19,548	100 - 3,000

III. Results

We report on two pieces of evidence which demonstrate that caregivers simplify their speech which is coordinated around their infants' vocalizations. The comparison of primary importance is whether caregiver contingent and non-contingent speech diverge in their TTR curves (counts of the number of unique word types as a function of increasing word token sizes). The approach we follow creates data which approximates the speech an infant would hear if they randomly selected samples of speech across all of our subjects. In addition, the simulations allow observations of hypothetical data at a larger scale of time than could be implemented within our laboratory (assuming that more time leads to more caregiver speech). If we see divergences between TTR curves of contingent and non-contingent speech as we increase sample size, this provides evidence for caregivers differentiating the complexity of their talk as they organize it around their infants' vocalizing.

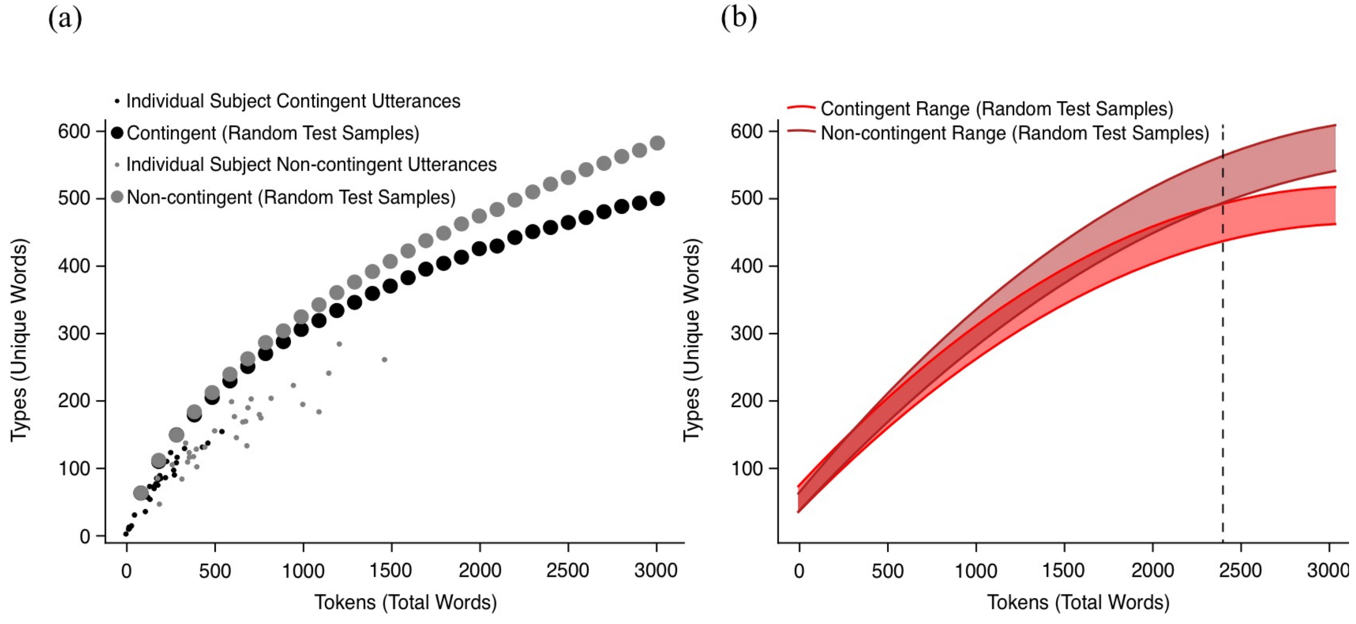

Figure 4. (A) Average counts of word types (number of unique words) as related to the total count of words (tokens) taken from random samples of words which were contingent and non-contingent on infants' vocalizations and the individual type-token values for individual subjects. (B) Maximum and minimum counts of word types as related to the total count of words (tokens) taken from random samples of words which were contingent and non-contingent on infants' vocalizations.

Finally, we conducted null hypothesis testing with Monte Carlo simulations to verify whether the difference in the number of tokens in contingent and non-contingent conversation is due to chance.

Test Sample Results. Figure 4a is a plot of the average count of the number of word types as a function of counts of total word tokens from the test samples of contingent and non-contingent words. In comparing the number of word types from contingent and non-contingent samples we found that 53 of the 100 pairs of random samples had more contingent unique word types than non-contingent speech at token size 100. With tokens at size 400 we found that 38 of the 100 pairs of samples had more unique word types in contingent than in non-contingent speech. Tokens at size 400 and above, all exhibited comparisons which showed greater unique word types in non-contingent speech; furthermore, tokens at size 1600 and above all showed that all 100 paired random samples included a greater number of non-contingent word types. The ranges of unique word types were completely nonoverlapping at all tokens at size 2400 and above (Figure 4b). Our estimates suggest that contingent and non-contingent speech may be similarly diverse when token size is small (e.g., less than 400) but at high token sizes contingent speech may be less diverse than non-contingent speech. The complexity gap we observe between the word diversity of caregiver speech which is coordinated around infants' vocalizations and non-contingent speech is harmonious with findings from adult conversations. In adult turn taking conversation, initial responses to conspecific's speech turn typically consist of much more simple speech content than speech which comes non-contingently and speakers have time to decide how to frame their next thought [24].

To compare our sampling results to the raw subject data, in the bottom left corner of Figure 4 we plotted the respective subjects' type and token counts of speech from their contingent and non-contingent utterances (small dots). It is clear that the counts of both types and tokens are much lower than the curves generated by the random sampling. The reason for this difference lies in the coherence of contiguous speech content of any given speaker compared to the incoherence obtained from random sampling. As caregivers speak, in order to form coherent speech, they must repeat words at a much higher rate than would be observed in a random sampling of an equivalent size of words from a pooled distribution of all the caregivers' speech. The usefulness of this raw data, however, is limited because they are confounded by sample size, a problem that is circumvented by our sampling approach.

Control Sample Results. Figure 5 depicts the average count of unique word types as a function of the word tokens sampled from contingent words and non-contingent words at 90%, 60% and 30% eligibility (see Figure 3b for a visual depiction of our size-reduced random control sample technique). Because contingent speech contains fewer total words to sample from in general, we derived non-contingent TTR curves which test the effects of incrementally decreasing the size of the eligible sample pool from which the curves are derived to test the effects of corpus size on TTR curve outcomes. The extent of differences from the contingent curve to the 30% eligibility non-contingent curve suggest the effect size of the difference between the word distributions of contingent and non-contingent speech. It is important to note that the contingent word corpus is .31 times the size of the non-contingent corpus, so 30% eligibility sampling is conservative in its estimate of non-contingent's TTR curve at a size comparable to the contingent corpora. Comparing the number of word types in pairs of contingent and non-contingent samples at token size 100 results in 54 out of 100 pairs of random samples which had more contingent unique word types than non-contingent. Making the same comparison at token size 300 results in 9 out of 100 pairs of random samples which had more contingent unique word types than non-contingent. At every token size, the range for contingent and non-contingent unique word types overlap. By this estimation, when restricting the eligibility of the non-contingent corpora to more closely resemble the size of contingent corpora, the divergence between the two TTR curves is vastly reduced. We interpret this to mean that when ignoring a potential source of variation, such as amount of talk, it is possible that contingent and non-contingent talk only differs by a small amount. However, in our view, size-matched control samples, which utilize all of the data points available in a given experiment, offer distinct advantages which the size-reduced control samples lack (Figure 3a). In previous simulations, we calculated differences between TTR curves generated from test samples and size-matched random control samples. This was useful because we could estimate effects due to sample size differences alone and compare those effects against those found in our test samples [20]. Further details on the results of size-matched random control samples can be found in our previous study [20].

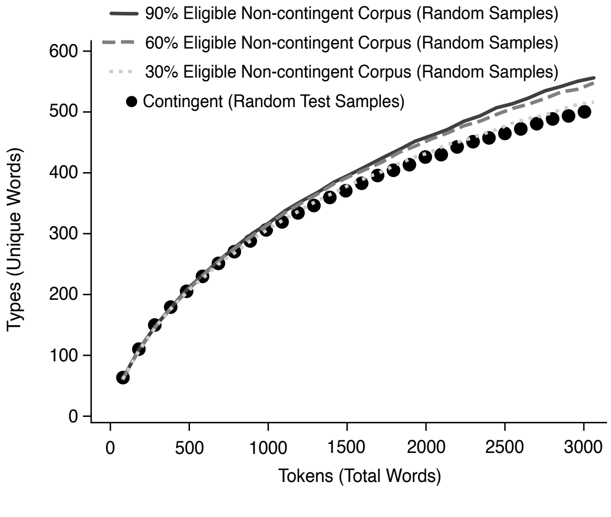

Figure 5. Average counts of word types (number of unique words) as related to the total count of words (tokens) taken from random samples of words which were contingent and non-contingent on infants' vocalizations. Non-contingent samples were pulled from corpora where 90%, 60% and 30% of the words were eligible for sampling.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCDS.2021.3095766, IEEE Transactions on Cognitive and Developmental Systems

IV. Statistical Monte Carlo Tests

The data presented above show differences in lexical diversity between contingent and non-contingent speech. It would be helpful if we can replicate those results in an independent simulation to verify our interpretation of the sampling results. The goal of the simulation below is to estimate the probability that the sampling results can happen even though there is no difference in lexical diversity between contingent and non-contingent speech. It is generally accepted that word frequency and word frequency rank follow a statistical trend known as Zipf's law. Therefore, we can simulate caregivers' speech in caregiver-infant conversation by sampling words from a dictionary of caregiver speech, while adhering to Zipf's law. If we assume that the dictionary sizes for contingent and non-contingent speech are same, we can count the number of unique tokens in the simulated dialogues. If we repeat this many times and count the occurrences when the number of unique tokens is same with the experiment, we can estimate the probability that the experiment result can happen even though there is no lexical diversity between contingent and non-contingent speech.

The distribution of parents' talk which is contingent on infants' vocalizations is characterized by a TTR curve which is shaped differently than the curve produced by parents' talk which is non-contingent. The importance of this finding is the demonstration that parent talk has special properties as a function of its timing within a turn-taking context. A key point of emphasis is that even over long timescales, the frequency of word types which characterize different kinds of talk does not relate to the amount of talk under scrutiny in a deterministic manner. Indeed, in a sample from Montag and colleagues (2015) which compared a corpus of words found in several picture books (total word counts were approximately 70,000) against words in the CHILDES corpus (total word counts were about 6.5 million), it is clear from simulation that the picture book (smaller token count) has a higher diversity of words [22,25]. However, in our sample, the contingent speech (smaller token count), has a lower diversity of words. We now turn to computational experiments which replicate the results of this paper with functions that relate a range of presumptive caregiver vocabulary sizes to observe how contingent unique words accumulate as new words are uttered.

Models and Assumptions. We used Monte Carlo simulations to test the generalizability and boundary conditions of our findings. The simulations tested the effects of a range of randomly selected caregiver vocabulary sizes on contingent and non-contingent lexical diversity. First, we assume that the vocabulary size of caregivers follows a Gaussian distribution. We denote the distribution of vocabulary size for contingent conversation as $\mathcal{N}_C(\mu_C, \sigma_C)$ and $\mathcal{N}_{NC}(\mu_{NC}, \sigma_{NC})$ for non-contingent conversation. After we randomly sample vocabulary sizes for contingent and non-contingent conversation, we set the probability of selecting each word in this vocabulary following a near-Zipfian distribution as shown in Eq 1. If we rank each word according to its frequency of occurrence, the frequency of the word with rank r, $Freq(w_r)$ is proportional to the inverse of its rank [26,27]. The probability of selecting word w_r, $P[w_r]$ can be derived by normalizing the probabilities of all words as shown in Eq. 2 [28]. Mandelbrot introduced the parameters α and β to improve the fit of the frequency distribution of actual languages across contexts and sample sizes [13]. However, the parameters of a near-Zipfian distribution are also different among participants, and we model this with the assumption that α and β themselves follow Gaussian distributions (where α is the y-intercept of the distribution and β is the slope). Finally, a conversation is built by randomly sampling words from this vocabulary which is characterized by a near-Zipfian distribution.

In this simulation, we estimate the near-Zipfian parameters from the conversation data and use it to estimate the vocabulary size distribution. Using these parameters, we calculate the probability of a null-hypothesis that the experimental results might happen by chance across a range of simulation results. Figure 6 shows the overall process of this approach.

$$Freq(w_r) \propto \frac{1}{(r+\beta)^{\alpha}} \qquad \textit{Eq. 1}$$

$$P[w_r] = \frac{\frac{1}{(r+\beta)^{\alpha}}}{\sum_{j=1}^{n} \frac{1}{(r+\beta)^{\alpha}}} \text{for i} = 1, 2, \ldots, n \qquad \textit{Eq. 2}$$

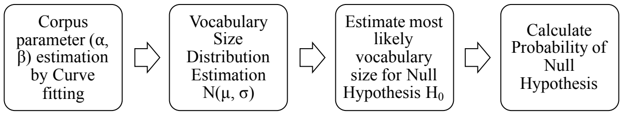
Corpus parameter (α, β) estimation by Curve fitting
Vocabulary Size Distribution Estimation N(μ, σ)
Estimate most likely vocabulary size for Null Hypothesis H_0
Calculate Probability of Null Hypothesis

Figure 6: Overall process of null-hypothesis testing using Monte Carlo Simulation.

Step 1: Estimating Zipfian parameters. We begin by estimating the near-Zipfian parameters from the experimental results because it is independent from other factors and affects all subsequent estimations. Zipf's law is an empirical observation and the parameters for distribution are different across individual languages and contexts. Piantadosi, 2014, estimated that α is 1.13, and β is 2.73 for a general English-speaking adult corpus but also showed that these parameters change according to the time and category.

We used curve fitting methods with least square loss, where we estimated α and β for all participants for contingent and non-contingent conversation. The mean α obtained was 0.88 with a standard deviation of 0.035. β has sample mean of 2.21 and standard deviation of 0.378. We used a Gaussian distribution for α and β with these mean and standard deviation in the following simulation.

Step 2: Estimating the vocabulary size distribution. We estimated the Gaussian distributions $\mathcal{N}_C$ and $\mathcal{N}_{NC}$ of the vocabulary sizes. The main assumption in this process is that the number of unique word types in the conversation of a certain length is predicted by a caregivers' vocabulary size. We set μ_U and σ_U as the mean and the standard deviation of the number of unique types in the conversation, and μ_V and σ_V as parameters for the vocabulary size which follows the Gaussian distribution. To find optimal μ_V and σ_V , we estimate the parameters iteratively. First, we fix σ_V and then find the μ_V, which will most likely produce the actual experiment data, μ_U and σ_U. After finding the optimal μ_V, we fix it and find the optimal σ_V using the same methods. Then we use the optimal σ_V value in the next iteration and search for an optimal μ_V again. We repeat this simulation until the μ_V and σ_V converge to the specified threshold.

$$\boldsymbol{P[H_0] = P[Z_C]P[Z_{NC}]} \qquad \textbf{\textit{Eq. 3}}$$

Given μ_V and σ_V, we utilize simulation to estimate the expected μ_U and σ_U. We sample a vocabulary size estimate from the $\mathcal{N}(\mu_V, \sigma_V)$. Then we sample α and β from Gaussian distribution we estimated in Step 1 and use it to build a vocabulary with near-Zipfian distribution. Then we construct a random conversation by sampling from this vocabulary. We generated 30 contingent and non-contingent conversation pairs using the conversation length of the actual participants. We calculated the expected number of unique word types by repeating this process.

When we use the mean (84.3) and standard deviation (41.2) of the number of unique types in the contingent conversation, we can estimate the $\mu_C\ as$ 206.0 and σ_C as 90.2. Similarly, we estimated the non-contingent vocabulary distribution parameter μ_{NC} as 275.8 and σ_{NC} as 101.3 by using the mean (181.1) and the standard deviation (57.4) of non-contingent conversation.

This computational simulation approach provides complimentary evidence to the findings in the Control Sample Results section if there is a low probability that the vocabulary size of contingent and non-contingent conversation are the same (null hypothesis). If this is the case, the null hypothesis is rejected. In this section, we will calculate the probability that the result obtained occurred when the mean of the vocabulary size is the same across contingent and non-contingent conversation.

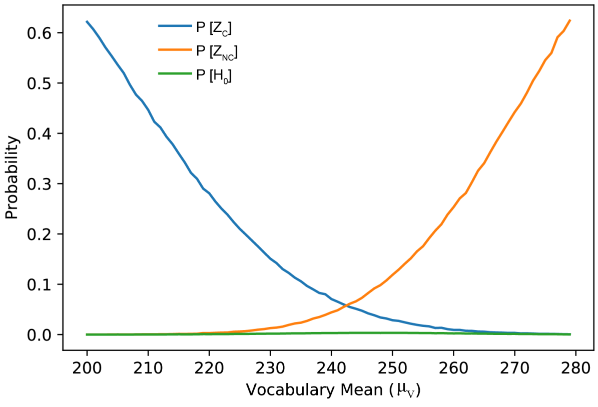

Figure 7: P[Z_C], P[Z_NC], and P[H_0] as the mean of the vocabulary size changes. We can see that we get maximum P[H_0] around when the vocabulary size is about 246. We used σ_0 = 101.3 for this simulation.

Null Hypothesis testing. Our null hypothesis H_0 is that the mean of the vocabulary size distribution for the contingent and non-contingent conversation is same. We denote this vocabulary size distribution as $\mathcal{N}_0(\mu_0, \sigma_0)$.

We define the event Z_C as obtaining less than 2528 unique types from 30 conversations with the length from the actual experiment. The event Z_{NC} is obtaining more than 5434 unique types from the 30 contingent conversations.

What is the probability that Z_C and Z_{NC} were drawn from the same vocabulary distribution $\mathcal{N}_0(\mu_0, \sigma_0)$ by chance? Because the two events are independent from each other, we can get $P[H_0]$ by multiplying the probability of the events by each other.

The σ_0 value affects the $P[H_0]$ such that the larger σ_0 value is, the higher $P[H_0]$ is estimated. For the conservative estimation of the Null-Hypothesis test, we use the higher value between $\sigma_C\ and\ \sigma_{NC}$ which was 101.3.

Similarly, $P[H_0]$ is affected by the μ_0. Smaller values of μ_0 make the event Z_C more likely. Higher μ_0 values make the event Z_{NC} more likely. When we search for the μ_0 that maximizes $P[H_0]$ by monte carlo simulation, we get $\mu_0 =$ 246 (Figure 7).

When we calculate the $P[H_0]$ probability that Z_C and Z_{NC} happened when the vocabulary size was from $\mathcal{N}_0(\mu_0 = 246, \sigma_0 = 101.3)$, we obtain $P[H_0]$ = 0.35%. Therefore, we can reject the null-hypothesis with 1% significance and conclude that the vocabulary size for the contingent context is smaller than the non-contingent context (Figure 8).

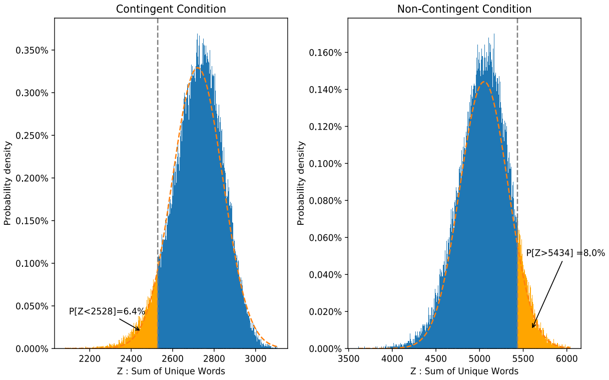

Figure 8: The histogram of Z_C and Z_NC when the vocabulary size is sampled from N(246; 101.3). The orange colored region represents the trials where the results were as extreme as the outcomes in the actual experiment.

V. Discussion

Our simulated environments make three main contributions. First, the content of caregivers' contingent and non-contingent speech are different, showing that immature infant behavior functions to influence their learning environment. Infants' prelinguistic vocalizing may promote language learning because it facilitates parental behavior that contains simplified, more easily learned information. In our view, the coordination of caregivers' speech content around infant immature vocalizations is an emergent property of early vocal turn-taking between caregivers and infants which facilitates the development of communication and early language. Second, our estimates demonstrate that the divergence of word diversity between contingent and non-contingent speech was exacerbated at larger word counts. At larger scales, our estimates suggest that contingent talk to infants remains simplified while non-contingent talk continues to increase in complexity. Third, our simulations demonstrate that the simplification of talk is not necessarily caused by smaller samples of talk and that non-contingent and contingent talk are two distinct distributions of words in the infants' early learning environment. The lexical complexity of contingent talk will not 'catch up' with that of non-contingent talk. This is because the contingent and non-contingent talk follow different curves describing their lexical complexity as a function of lexical activity.

A. Methodological contribution

The limitations inherent in simple TTR metrics need not dissuade researchers from utilizing TTR curves. On the contrary, by demonstrating theoretically motivated applications of these curves, we hope to promote additional novel approaches to better understand the early ambient language learning environment of prelinguistic infants. It has been well documented that TTR are tightly linked to sample size. Our simulations point out, however, that there are techniques to minimize the effects of sample size. Size-reduced random control samples may not yield as many insights as size-matched random control samples. Through size-matched control samples, we can test whether TTR curve differences persist when we artificially create a single distribution of words and resample new corpora at matching sizes to the original. By forcing TTR curve generation from a single word distribution, we observe the true effects of sample size differences from a single word distribution. Results obtained by reducing the size of one corpus to compare the resulting curve against a similarly-sized corpus may be difficult to interpret. The main disadvantage of size-reduced sampling is that it does not provide evidence pertaining to whether the original corpora were drawn from the same distribution of words. Crucially, distributions of words can be compared to one another even when the size of the datasets are not equivalent. TTR curves can give rise to unique insights when used alongside size-matched control curves that illuminate whether sample size determines the nature of the distribution. A limitation of the present work is that our estimations do not incorporate any information about change in caregivers' lexical diversity over the course of infant development. Together with our previous research, we provide evidence that a) individual caregivers contingent lexical diversity is simplified and b) this simplification phenomenon exists above and beyond what is expected due to sample size differences in contingent and non-contingent speech [10].

However, visual inspection of the estimates becomes more challenging as vocabulary size differences become smaller. Prior research utilized the size-reduced random sampling technique to investigate the difference between word diversity in a picture book corpus and the CHILDES corpus [25]. When we employ the same technique, the observed differences between contingent and non-contingent conversation are small. Similarly, the size-matched random control sampling would show less of a difference between the lexical diversity of contingent and non-contingent talk as the vocabulary size difference becomes narrower. Monte Carlo simulation studies provide another framework for estimating the difference in vocabulary size between contingent and non-contingent talk. Monte Carlo simulation studies rely on many assumptions and corresponding parameters. Therefore, the results should not be accepted as a conclusion, rather as additional supporting evidence. Among many parameters, our conjecture is that the standard deviation of the vocabulary size between individuals will be the key factor affecting the simulation result. Our Monte Carlo simulation estimates the chances of obtaining our results when vocabulary sizes for contingent and non-contingent speech are the same. As a concrete example, we used 101.3 as the standard deviation because it was the largest number from the fit in the experimental data. In this case, the $P[H_0]$ was 0.35% meaning that it is unlikely that vocabulary sizes for the contingent and non-contingent speech are the same. However, when we use larger standard deviation such as 160, the $P[H_0]$ was 6.5% which will not pass a 5% significance level for rejection of this null hypothesis. A more reliable estimate of the standard deviation of individual vocabulary sizes would give our simulation more predictive power. Estimating individual vocabulary size, however, is still a challenging problem especially given the diversity of the social contexts.

B. Developmental contribution

Functions of simplified caregiver speech. Over long timescales, our simulation estimates that caregivers' speech which is organized around infants' vocalizations will generally contain a higher amount of repeated lexical items. The evidence on the influence repetition has on language development is mixed over longer time scales (6 to 12 months); repetitive language input has been linked to both lower [29,16] and greater [6] vocabulary sizes later in development. The extent to which caregivers organize less repetitive speech around infants' vocalizations predicts infants' vocal maturity [10]. Caregivers naturally use similar words and phrases in contiguous utterances, these quasi-repetitive adjacent utterances are called variation sets [31]. For example, a caregiver might say "Is that a spoon? Where's your spoon? Get your spoon!". Variation sets facilitate learning in adults and similarly guide early linguistic learning in infants [30,31]. Repetition is not isolated to infants' language environment. For example, during meal times infants' encounter iconic objects that are labeled reliably by caregivers, creating recurrent visual and linguistic cues coupled closely in time. Caregiver speech, organized around daily routines, is part of a constellation of cues which facilitate word learning from everyday activities [32,33,34]. The earliest words infants learn could arise out of prelinguistic turn-taking interactions, such as when an infant points at an object and vocalizes, then the parent utters the object's label. Indeed, infants in experimental settings learn word-object mappings when caregivers respond to infants' object-directed vocalizations with object-label utterances [35].

Caregivers' contiguous speech is highly coherent and therefore repetitive, but over long time scales, continuous speech will change topic or dive deeper into a single topic and therefore elicit a greater diversity of words. However, because it takes speakers time to plan more elaborate and diverse utterances in production, contingent speech will always be relatively simplified compared to non-contingent speech. We do not hypothesize that the more parents simplify their speech, the better the opportunities for infants' learning. Evidence suggests that infants who hear more lexical diversity spoken contingently on their vocalizations have a greater capacity for producing syllables that include consonants [10]. Thus, increased contingent caregiver lexical diversity may facilitate infants' vocal learning.

Evidence suggests that stable speech sounds might be better targets for infants to base refinement of their vocal repertoires on. Simplified caregiver speech might be useful because it does not overtax infants' limited working memory at the moments they are ready to learn. In addition to the perception of mature adult speech, infants' own vocal productions can serve as a source of stability in the input. Recent findings suggest that infants who have a stable phonological pattern in their vocal repertoire are better at segmenting the speech stream in perception [36]. One interpretation of these findings is that when stable representations in infants' phonological memory are present (from either within or without), the processing load for similar phonological structures is eased. It is possible that early stability in infants' production is useful for similar reasons in perception. When caregivers produce similar sound forms, infants could discover underlying structure by examining the input's redundancy.

Why is the speech content of caregivers' responses to infants' vocalizations simplified? We propose two testable hypotheses which can guide future work on this question. The *altriciality effects hypothesis* suggests that the source of change between contingent and non-contingent lexical diversity is infants' vocal altriciality. Human infants are altricial – they depend on caregivers for survival over an extended period of development. Characteristics of immaturity (e.g., neotenic appearance) may serve as cues that facilitate caregiving behavior [37]. Infants' early vocalizations may serve as salient cues of immaturity. Even when infants are capable of engaging in sophisticated social behavior (e.g., smiling and pointing) their prelinguistic vocalizations are still immature and do not resemble words or language. Such immaturity may drive the observed simplification of adults' contingent speech, and explain why non-contingent speech is not simplified.

An alternative hypothesis is that caregivers respond to infants' vocalizations with simplified speech because of the reduced processing time allotted them. The *processing time hypothesis* suggests that the source of change between contingent and non-contingent lexical diversity is the small increment of time between infants' vocalization and caregivers' speech in response. Two pieces of evidence would speak to these hypotheses. First, adults' lexical diversity should be measured in adult-adult turn-taking as a function of contingency. If adults' speech complexity differs across responses to adult speech and infant vocal turns, this would provide evidence in favor of the altriciality effects hypothesis. Second, if the reduced processing time hypothesis holds, then the latency between infants' vocalizations and caregiver speech responses should correlate positively with caregivers' speech complexity. If there is no correlation between latency and speech complexity, this would provide evidence in favor of the altriciality effects hypothesis. These analyses are beyond the scope of the present paper, but future work in our lab will shed light on these hypotheses.

The length of infant vocalizations may have also influenced the nature of caregiver responses. As infants age, they begin to string together multiple prelinguistic vocalizations closely in time, structuring vocal bouts into syllable sequences [38]. In the current study we considered parents' responses to non-cry vocalizations to investigate the influence of these vocalizations on caregiver speech. However, it is possible that the complexity of caregivers' speech changes in response to infant vocalizations that extend further in time. In the future, there is a need to study the influence of infants' vocalizations which span multiple syllables to investigate whether infant vocal sequences change caregiver responding. It is possible that one pathway for infants to increase the lexical diversity of their caregivers' speech is by organizing their vocalizations with sequential structure within conversational turn-taking contexts.

Our findings have important implications for data collection at large scales and language development intervention studies. Home recording efforts can reveal the extent to which there are changes in linguistic structure within parent-child vocal turn-taking bouts over time [39]. The main focus in several interventions for at-risk families surrounds the number of words produced by caregivers (e.g., Providence Talks; http://www.providencetalks.org) or turn-taking with infants [40]. Evidence suggests that interventions are effective at promoting early language development when caregivers' increase their lexical activity that is organized around infants [41]. Future research should seek to better illuminate how caregivers might attempt to continually adapt their contingent talk to match their infants' current communicative capacities.

Our findings point to the role of infants' immature vocalizations in shaping infants' own language learning environment. Computational models of vocal learning utilize mechanisms of accurate prediction of environmental changes; such a mechanism may also support infants in contexts of social learning [42, 43]. Theoretical frameworks have postulated that discrepancies between predictions and observed outcomes may elicit a learner's curiosity. Models centered around curiosity select to learn from information from which they can diminish the error of their own predictions at maximum rates. When infants vocalize they create opportunities to learn the effects of their own vocalizations on their caregivers' behavior. During their first 12 months of life, infants rapidly learn that their own prelinguistic vocalizing elicits responses from their caregivers [5]. Eliciting mature speech sounds from caregivers may become the target of infants' curiosity and subsequently guide their vocal development. For a more advanced understanding of early infant learning, future large-scale observational, computational and experimental research should investigate the effects infants have on the temporal and distributional properties of parents' speech.

ACKNOWLEDGMENTS

Jessica Montag and Felix Thoemmes provided helpful critical discussion. We thank the families who participated in the study. Sofia Carrillo, Shelly Zhang, Kexin Zheng and SoYoung Kwon transcribed parent speech and coded infant vocalizations.

REFERENCES

[1] S. Carouso-Peck, and M. Goldstein. "Female social feedback reveals non-imitative mechanisms of Vocal Learning in Zebra Finches," Curr. Bio., vol 29, pp. 1-6, 2019.

[2] Y. Gultekin, and S. Hage, "Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys," Sci Advances, vol 4, eaar4012, 2018.

[3] M. Goldstein, and J. Schwade, "Social feedback to infants' babbling facilitates rapid phonological learning," Psych. Sci., vol 19(5), pp. 515–523, 2008.

[4] P. Kuhl, F. Tsao, and H. Lui, "Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning," PNAS, vol 100, pp. 9096–9101, 2003.

[5] M. Goldstein, J. Schwade, and M. Bornstein, "The value of vocalizing: five-month-old infants associate their own noncry vocalizations with responses from caregivers," Child Dev., vol 80(3), pp. 636-644, 2009.

[6] R. Newman, M. Rowe and N. Ratner, "Input and uptake at 7 months predicts toddler vocabulary: the role of child-directed speech and infant processing skills in language development," J. of Child Lang., vol 43, pp. 1158-1173, 2016.

[7] C. S. Tamis-LeMonda, M. H. Bornstein, and L. Baumwell, "Maternal Responsiveness and Children's Achievement of Language Milestones," Child Development, vol. 72, no. 3, pp. 748–767, 2001

[8] E. D. Thiessen, E. A. Hill, and J. R. Saffran, "Infant-Directed Speech Facilitates Word Segmentation," Infancy, vol. 7, no. 1, pp. 53–71, 2005.

[9] L. Smith, S. Jayaraman, E. Clerkin, and C. Yu, "The developing infant creates a curriculum for statistical learning," Trends in Cognitive Sciences, vol. 22(4), pp. 324-336, 2018.

[10] S. Elmlinger, J. Schwade, and M. Goldstein, "The ecology of prelinguistic vocal learning: parents simplify the structure of their speech in response to babbling," J. of Child Lang., 2019.

[11] J. Montag, M. Jones, and L. Smith, "Quantity and diversity: Simulating early word learning environments," Cog Sci., vol 42, pp 375-412, 2018.

[12] E. Bergelson, M. Casillas, M. Soderstrom, A. Seidl, A. S. Warlaumont, and A. Amatuni, "What Do North American Babies Hear? A large-scale cross-corpus analysis," Developmental Science, vol. 22, no. 1, p. e12724, 2018.

[13] S. Piantadosi, "Zipf's word frequency law in natural language: A critical review and future directions," Psych. Bull. & Rev., vol 21(5), pp. 1112-1130, 2014.

[14] E. Clerkin, E. Hart, J. Rehg, C. Yu, L. Smith, "Real-world visual statistics and infants' first-learned object names." Phil. Trans. R. Soc. B 327, 20160055.

[15] W. Johnson, "Studies in language behavior: A program of research," Psych. Monographs, vol 56(2), pp. 1-15, 1944.

[16] J. Schwab, M. Rowe, N Cabrera, C. Lew-Williams, "Fathers' repetition of words is coupled with children's vocabularies." J of Exp. Child Psych., vol 166, 437-450, 2018.

[17] B. Richards, "Type/Token ratios: what do they really tell us?" J. of Child Lang., vol 14(2), pp. 201-209, 1987.

[18] M. Goldstein, A. King, and M. West, "Social interaction shapes babbling: Testing parallels between birdsong and speech," PNAS vol 100(13), pp. 8030-8035, 2003.

[19] A. Warlaumont, J. Richards, J. Gilkerson, and D. Oller, "A social feedback loop for speech development and its reduction in autism," Psych. Sci., vol 25(7), pp. 1314-1325, 2014.

[20] S. Elmlinger, J. Schwade, and M. Goldstein, "Babbling elicits simplified caregiver speech: Findings from natural interaction and simulation," ICDL-EpiRob, pp. 1-6. 2019.

[21] C. Venker, D. Bolt, A. Meyer, H. Sindberg, S. Weismer, and H. Tager-Flusberg, "Parent telegraphic speech use and spoken language in preschoolers with ASD," J. of Speech, Lang. and Hear. Res., vol 58(6), pp. 1733-1746, 2015.

[22] B. MacWhinney, "The childes project: The database", vol. 2. Psychology Press, 2000.

[23] L. Van Egeren, M. Barratt, and M. Roach, "Mother-infant responsiveness: Timing, mutual regulation, and interactional context." Dev Psych., voc 37.5, pp. 684, 2001.

[24] J. Du Bois, "Towards a dialogic syntax," Cog. Ling., vol 25(3), pp. 359-410, 2014.

[25] J. Montag, M. Jones, and L. Smith, "The words children hear: Picture books and the statistics for language learning." Psych. Sci., vol 26(9), pp. 1489-1496, 2015.

[26] G. K. Zipf, "The psychobiology of language (rout-ledge, london)," 1936.

[27] G. K. Zipf, "Human behavior and the principle of least effort." 1949.

[28] B. Mandelbrot, "An informational theory of the statistical structure of language," Communication theory, 1953.

[29] E. Newport, H. Gleitman, and L. Gleitman, "Mother, I'd rather do it myself: some effects and non-effects of maternal speech style." in Talking to Children: Language input and acquisition, Cambridge: Cambridge University Press, 1977, pp. 109–149.

[30] J. Schwab, C. Lew-Williams, "Repetition across successive sentences facilitates young children's word learning." Developmental Psychology, vol. 52, p. 879, 2016.

[31] L. Onnis, H. Waterfall, and S. Edelman, "Learn locally, act globally: Learning language from variation set cues," Cognition, vol 109, pp. 423-430, 2008.

[32] M. Casillas, P. Brown, S. Levinson, "Early language experience in a Tseltal Mayan village," Child Dev., vol. 90, pp. 1819-1835, 2020.

[33] C. Tamis-LeMonda, S. Custode, Y. Kuchiro, K. Escobar and T. Lo, "Routine language: Speech directed to infants during home activities." Child Dev., vol. 90, pp. 2135-2152, 2019.

[34] C. Yu, and L. Smith, "Embodied attention and word learning by toddlers." Cognition, vol 125.2, pp. 244-262, 2012.

[35] M. Goldstein, J. Schwade, J. Briesch, S. Syal, "Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn." Infancy, vol. 15, pp. 362-391, 2010.

[36] M. Hoareau, H. Yeung, T. Nazzi, "Infants' statistical word segmentation in an artificial langauge is linked to both parental speech input and reported production abilities." Developmental Science, vol 22, p.e12803, 2019.

[37] L. Zebrowits, S. Brownlow, K. Olson, "Baby talk to the babyfaced." J Nonverbal Behav, vol 16, pp. 143-158, 1992.

[38] D. Lipkind, G. Marcus, D. Bemis, K. Sasahara, N. Jacoby, M. Takahsi, … & O. Tchernichovski, "Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants," Nature, vol 498, pp. 104-108. 2013.

[39] M. VanDam, A. Warlaumont, E. Bergelson, A. Cristia, M. Soderstrom, P. De Palma, and B. MacWhinney, "HomeBank: An online repository of daylong child-centered audio recordings," in Seminars in speech and language, NIH Public Access, p. 128, 2016.

[40] K. Leffel, and D. Suskind, "Parent-directed approaches to enrich the early language environments of children living in poverty," Sems. in Speech & Lang., vol 34, pp. 267–278, 2013.

[41] M. McGillion, J. Pine, J. Herbert, D. Matthews, "A randomised controlled trial to test the effect of promoting caregiver contingent talk on language development in infants from diverse socioeconomic status backgrounds," Journal of Child Psychology and Psychiatry, vol. 58, pp. 1122-1131, 2017.

[42] C. Kidd, B. Hayden, "The psychology and neuroscience of curiosity," Neuron, vol 88(3), pp. 449–460, 2015.

[43] C. Moulin-Frier, S. Nguyen, and P. Oudeyer, "Self-organization of early vocal development in infants and machines: the role of intrinsic motivation," Fntiers in Psych., vol 4, p. 1006, 2014.