Babbling elicits simplified caregiver speech: Findings from natural interaction and simulation

Steven L. Elmlinger
Psychology Department
Cornell University
Ithaca, New York - USA
sle64@cornell.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*—What is the function of babbling in language learning? Our recent findings suggest that infants' immature vocalizations may elicit simplified linguistic responses from their caregivers. The contributions of parental speech to infant development are well established; individual differences in the number of words in infants' ambient language environment predict communicative and cognitive development. It is unclear whether the number or the diversity of words in infants' environments is more critical for understanding infant language development. We present a new solution that observes the relation between the total number of words (tokens) and the diversity of words in infants' environments. Comparing speech corpora containing different numbers of tokens is challenging because the number of tokens strongly influences measures of corpus word diversity. However, here we offer a method for minimizing the effects of corpus size by deriving control samples of words and comparing them to test samples. We find that parents' speech in response to infants' babbling is simpler in that it contains fewer word types; our estimates based on sampling also suggest simplification of word diversity with larger numbers of tokens. Thus, infants, via their immature vocalizations, elicit speech from caregivers that is easier to learn.**

Keywords—Parent-child interaction, prelinguistic vocalizations, conversational turn-taking, language environment, simulation

I. Introduction

Across many species, parents' responses to their offspring's immature vocalizations play a crucial role in the development of communication. Vocal learning in humans [1], songbirds [2], and marmosets [3] relies on adults' coordination of their vocalizations around those of their offspring to create contingent social feedback that facilitates learning of more advanced vocal patterns. Human infants develop vocal communication gradually, forming expectations that their immature babbling will elicit social input [4,5]. Individual differences in communicative development are related to the social environment, as no two parents talk to their children in exactly the same way. Variability in the statistical structure of parent's speech to prelinguistic infants predicts later language outcomes [6]. While our understanding of parents' speech and responsivness to infants is well-informed by research, the statistical structure of parents' speech in response to infant babbling has received little attention.

Recent efforts have revealed that infant babbling catalyzes the production of simplified, more easily learnable language from their caregivers [7]. Parents' speech in response to infants' babbling is both lexically and syntactically simplified. Parent utterances that were contingent on babbling contained fewer unique word types, shorter utterance lengths and higher rates of single-word utterances. Whether this simplification effect is consistant across infant development, larger sample sizes (i.e., larger corpora) and longer time periods is unknown. Sophisticated efforts to disentangle sample size effects from true effects of diversity of words among different data sets are ongoing [8]. A deep understanding of just how simplified parents' responses to babbling are and how linguistic simplification may scale with sample size is the goal of the current work.

Analytical, theoretical, and interpretation problems have emerged from the recent surge in large-scale data collection, but new techniques are emerging to address them. The main reason research concerning word diversity poses a challenge to researchers is because word frequency rank distributions do not form a normal distribution. Instead, word ranks are characterized by distributions wherein few words are very

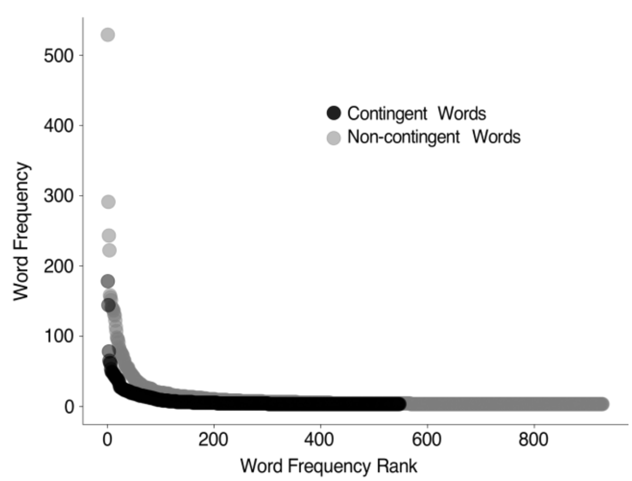

Figure 1. A word-rank by word-frequency plot of contingent vs non-contingent words.

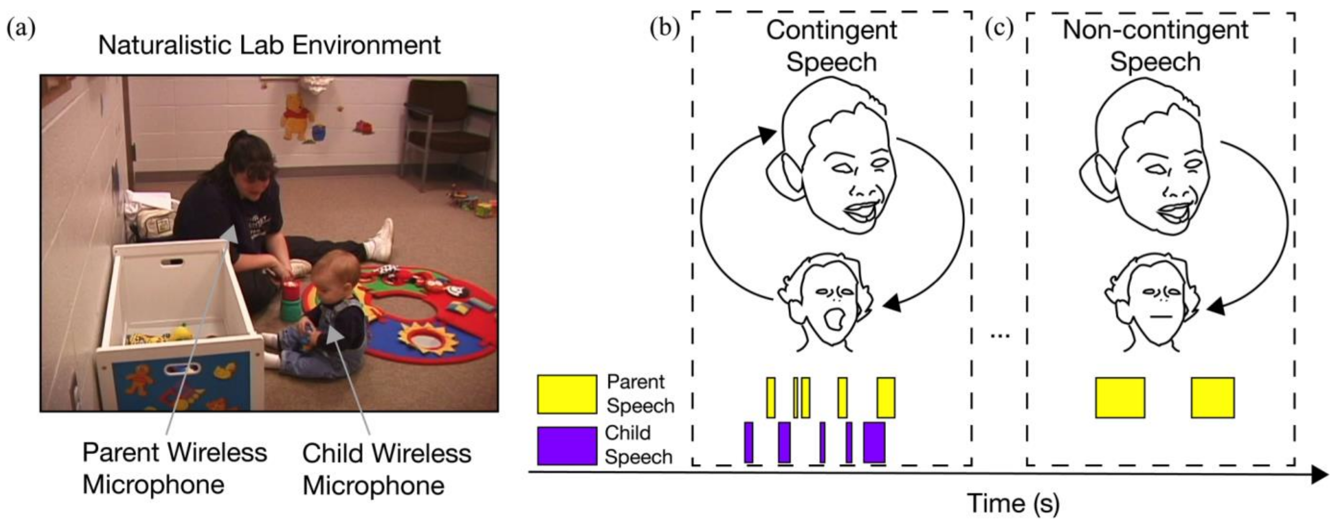

Figure 2. (a) Parents and infants outfitted with wireless microphones played with toys in the lab. Example speech streams from recordings, which allowed for categorical coding of parents' speech as contingent on infants' vocalizations (b) or not (c). Contingent parent speech was operationalized as speech which occurred within 2 seconds of a non-cry infant vocalization.

frequent and many words are rare (see Figure 1 for word frequency rank distributions of Elmlinger et al., 2019; also see [9]). These heavy-tailed distributions charactazize the words infants hear, the regularity of objects infants see in everday learning environments, and which object labels children learn first [10]. Summarizing these distributions with a single value is not straightforward because assumptions of central tendency do not apply [11]. Recent insights have demonstrated the predictive power in deriving ratios of the number of unique words (types) to the total number of words (tokens), that is, the type-token ratios (TTR), of target language corpora [12]. However, two main problems, in our view, remain underdetermined with regard to TTR measures. First, it is well known that TTR measures cannot be compared if they were derived from language corpora, or subsets of words, with differing sample sizes, as TTR scales with sample size non-linearly [13]. Second, any one measure of type-token relationships will obscure the nature of word distributions produced over different periods of time [8]. For example, the number of word types observed in a single corpus reveals little about the number of word types in a corpus containing a different number of words. Montag and colleagues have begun to address these issues with success. We aim to add a new technique to shed additional light on the analytical issues in parallel with our main research question: what is the relationship between number of words and diversity of words that infants' babbling elicits from their caregivers?

With larger scale data analyses on the rise, insights can be gleaned from real-time analyses which are simulated to scale to those of longer time-scales. The linguistic environment of infants in everyday learning contexts operates at multiple timescales. Learning happens in real-time from caregivers' contingent responses to infants babbling [14]. Learning also happens gradually over long time scales over successive vocal turn-taking bouts [15]. Here, we connect the two timescales by analyzing parent utterance data from recordings of free-play parent-infant interactions and simulate type-token relationships of parents' speech in response to babbling, as compared to parents' speech which was not uttered in response to babbling.

II. Methods

A. Participants

Thirty mother-infant dyads participated (infant mean age = 9.20, range: 9.4 -10.14). Participants were recruited from birth announcements in local newspapers and through advertisements. Mothers received a t-shirt or a bib as a gift for their participation. Participants were part of a larger corpus from a previously published study [1,7].

B. Apparatus

Sessions were recorded in a naturalistic environment (a 12 ft. x 18 ft. playroom) with toys, a toy box, and posters of animals. Infants were free to roam around the room and explore. Interactions were video recorded via three remote-controlled digital video cameras. To obtain detailed audio recordings, each infant wore denim overalls concealing a wireless microphone (Telex FLM-22; Telex Communications, Inc., Burnsville, MN) and transmitter (Telex USR-100). Caregivers wore a wireless lapel microphone (Telex FMR-150) with a transmitter concealed in a pouch at their waist (Telex USR-100) (Figure 2a). Infant vocalizations and caregiver speech were recorded on distinct audio channels.

C. Procedure

Participants came to the lab for 30-min play sessions. The sessions were unstructured, with parents asked to play as they normally would at home.

D. Speech Transcription

Parents' speech during play was transcribed in full. Caregiver utterances were segmented if they were bounded by silence longer than 2 sec and/or if they exhibited terminal pitch contours [16]. Utterances from the parents were categorized as contingent if they occurred within 2 sec of the offset infants' vocalizations (Figure 2b); all other parent utterances were categorized as non-contingent (see Figure 2c). Caregiver responses to crying, fussing, and vegetative vocalizations (e.g., coughs) were excluded from our analysis.

E. Sampling procedure

Type-token ratios are highly determined by sample size and thus do not lend themselves to straightforward interpretation when relying on a single sample [8]. We address this limitation by capturing the number of unique word types at multiple sample (token) sizes. To observe the aggregate number of unique word types as we increased the number of tokens, we conducted random sampling from contingent and non-contingent words at various sizes from the two corpora separately.

Test Sample. We constructed aggregate samples by taking incrementally larger random samples (from 100 – 3000 words) that increased in increments of 100 words. This sampling was conducted with replacement, so each sample was selected from the total set of all words in each corpus. The procedure was repeated 100 times for each sample size. The number of unique word types was then counted for each sample.

Control Sample. The corpus of contingent words (n = 6,199) is much smaller than our corpus of non-contingent words (n = 19,548), so we sought to understand the effects corpus size has on our aggregate test sample. To accomplish this, we randomly selected new words at equivalent token sizes (6,199 words and 19,548 words) observed in our original contingent and non-contingent corpora from the total number of words in the corpus (summing contingent and non-contingent). This produced new corpora that tested the effects of sample size because any effect of the contingent or non-contingent words would be equally present in both samples due to randomization. Our control sample was designed to assess whether having fewer total words in our contingent corpora could account for the smaller number of word tokens observed in our contingent test sample. By constructing control samples of the same token size as our contingent and non-contingent samples, we can isolate the effects of smaller corpora size on word type counts.

III. Results

Below we present two datasets that highlight how parents reduce their lexical diversity when responding to infants' babbling. We first compared parents' contingent and non-contingent speech to determine whether they differed in their number of word types as a function of increasing word token sizes (type-token ratio curves – TTR). As opposed to observing a single TTR value which is highly influenced by sample size, we observe TTR over a range of sample sizes to investigate potential differences in contingent vs non-contingent word type changes as a function of total words (tokens). Our approach draws both from well-known and more recent advanced attempts at characterizing the interdependence of token and type size at scale [8,17,18]. In effect, our technique generates a hypothetical subject which would hear random samples of speech across all of our subjects at a timescale much longer than we could implement in the lab. If we observe differences

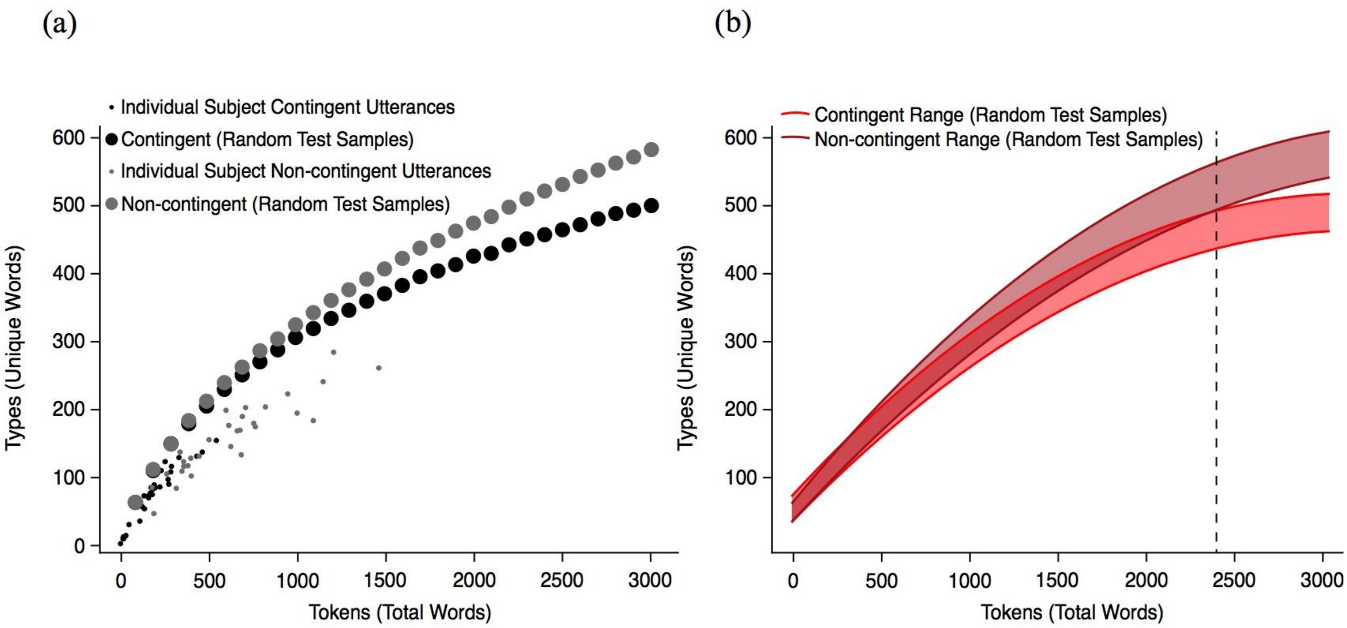

Figure 3. (a) Mean number of unique words (types) as a function of the total number of words (tokens) in the samples taken from words uttered contingently and non-contingently on infants' vocalizations and the individual type-token counts per subject. (b) Range of unique words (types) as a function of the total number of words (tokens) in the samples taken from words uttered contingently and non-contingently on infants' vocalizations.

Manuscript 67 submitted to 2019 Joint IEEE 9th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob). Received March 4, 2019.

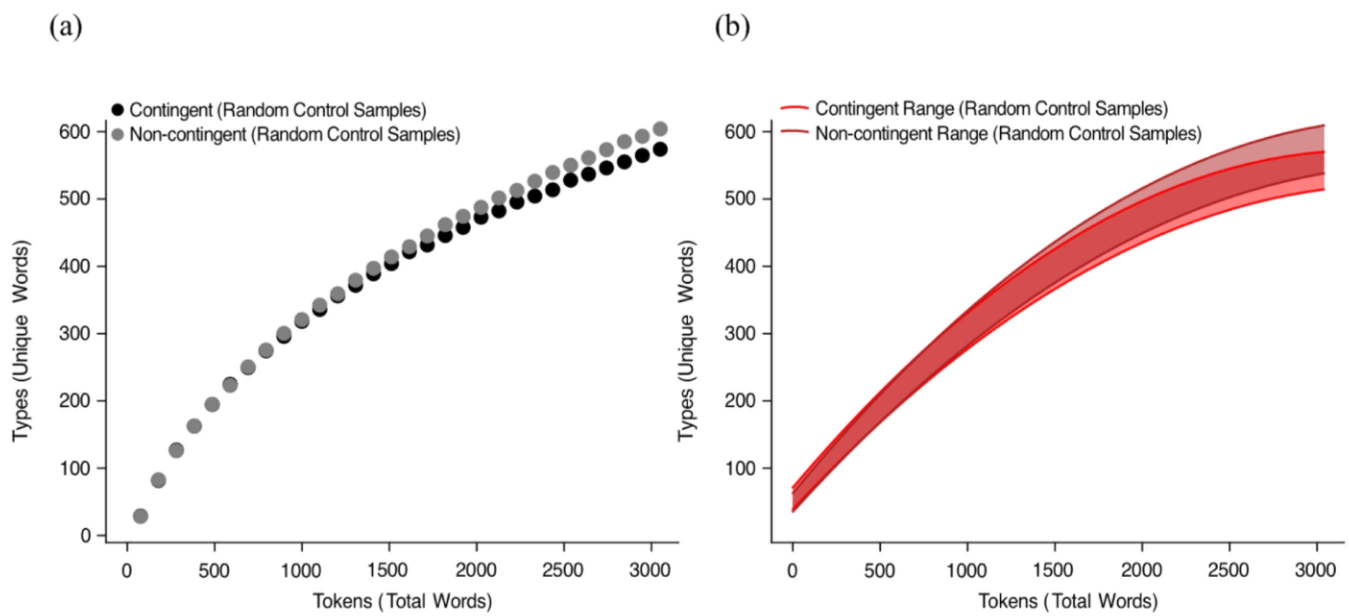

Figure 4. (a) Mean number of unique words (types) as a function of the total number of words (tokens) in the samples taken from control corpora of contingent-sized and non-contingent-sized random samples of words. (b) Range of unique words (types) as a function of the total number of words (tokens) in the samples taken from control corpora of contingent-sized and non-contingent-sized random samples of words.

between the TTR values for contingent and non-contingent speech at a variety of sample sizes, this indicates that parents' utterances differ in diversity of words as a function of whether they are responding to infant vocalizations. We then asked whether the difference between contingent and non-contingent word types increased as a function of increasing token size. By taking the difference between control and test simulation results, we examined whether or not the difference between the number of contingent and non-contingent word types remains constant over increasing word token sizes.

A. *Contingent speech is less lexically diverse than non-contingent speech*

Test Sample Results. Figure 3a shows the mean number of unique words (types) as a function of the total number of words (tokens) from the random test samples of contingent and non-contingent words. When we compared the type counts in paired contingent and non-contingent samples at token size 100, 53 of the 100 pairs of random samples contained greater number of unique types in contingent than in non-contingent speech. When we compared type counts at token size 400, only 38 of the 100 pairs of random samples contained greater number of unique types in contingent than in non-contingent speech. In comparisons of all token sizes greater than 400, there were more unique word types in samples from non-contingent utterances; moreover, in all token sizes greater than 1600, 100 out of 100 paired random samples contained a greater number of unique types in non-contingent speech. For all sample sizes greater than 2400, the ranges of unique word types are completely nonoverlapping (Figure 3b). These estimations suggest the possibility for contingent utterances to be equally repetitive as non-contingent utterances at smaller sample sizes (at token sizes < 400) and for contingent utterances to be more repetitive than non-contingent at larger sample sizes. This simplification of word types when responding to infants' babbling is consistent with accounts of adult-adult dialog where simplified initial responses to conversation partners spoken contingently on the partner's speech afford opportunities for speakers to decide how to frame their next thought, spoken non-contingently, in longer, more complex utterances [19].

Individual subject type and token counts of contingent and non-contingent speech can also be seen in figure 3. The type and token counts here are much lower than the means derived from random sampling because the pragmatic coherence of an individual caregiver's speech required her speech to be much more repetitive than would be found in a random selection of a similar number of words from all participants.

Control Sample Results. Figure 4a shows the mean number of unique words (types) as a function of the total number of words (tokens) from the random control samples of contingent and non-contingent words. Again, because the corpus of contingent speech is much smaller than the corpus of non-contingent speech, we derived the control sample to test the effects of corpus size on resulting type-token ratio curves. If the difference we observe between the curves in Figure 3 are due to corpus size, we would observe similar differences in samples drawn from similarly-sized corpora in the control sample. When we compared the type counts in paired contingent and non-contingent samples at token size 100, 58 of the 100 pairs of random samples contained greater number of unique types in contingent than in non-contingent speech. When we compared type counts at token size 900, only 47 of the 100 pairs of random samples contained greater number of unique types in contingent than in non-contingent speech. In comparing all token sizes greater than 900, there were more unique word types in samples from non-contingent utterances. For all token sizes, the ranges of unique word types overlap (Figure 4b). These estimations suggest that, given randomly sampled words of corpus sizes used in the test simulation, we do not observe divergence in the type-token ratio curves. We interpret this to mean that the divergence we see in the test simulation is due to actual word diversity differences and not differences in sample size.

B. Contingent speech becomes increasingly less lexically diverse than non-contingent speech with increasing token size

One possibility is that, although there is no overlap in the ranges sampled, the trajectory in types vs. tokens in contingent test vs. control random samples are indistinguishable. In other words, simply because the test sample demonstrates non-overlapping ranges and the control sample demonstrates completely overlapping ranges does not clarify whether the contingent samples of test vs. control random samples are different from one another over varying token sizes. Here we take the difference score between test and control samples of contingent and non-contingent type counts across all token sizes to observe the magnitude of difference between contingent samples from the test simulation and contingent-sized-random samples from the control simulation. We treat this difference as the number of words expected to differ due to the effect of parents responding to infants' babbling. As seen in figure 5, the difference between control and test contingent samples emerges at token size 500 and continues to grow at each increasing step in token size.

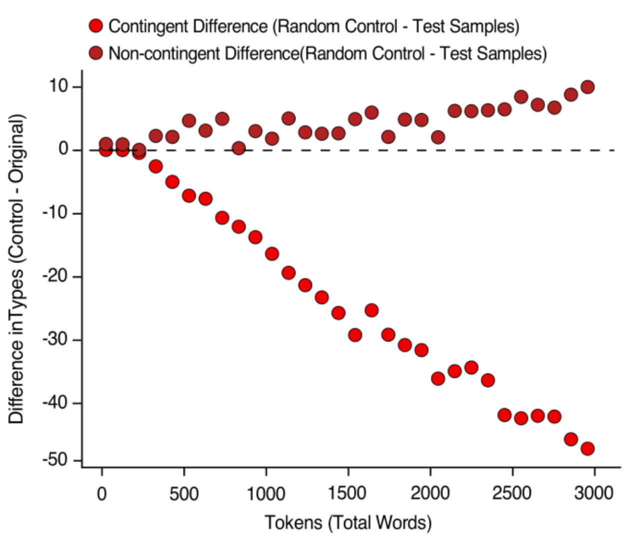

Figure 5. Difference in unique words (types) between control and test samples as a function of the total number of words (tokens) in the samples taken from contingent and non-contingent words.

IV. Discussion

The contributions of our simulated environments are threefold. First, the distributional differences between parents' contingent and non-contingent speech reported in the current analyses suggest a new form of influence that infants wield over their learning environment. Infants' immature vocalizations may create language learning opportunities by eliciting responses from parents that contain simplified, more learnable information. Second, our simulations estimate that the difference between contingent and non-contingent distributions of words would not remain constant at increasing word counts. It estimates that, at scale, contingent talk to infants will remain simplified with increasing words while non-contingent talk to infants will continue increasing in diversity, thus expanding the complexity difference between the two distributions as a function of more talk. Third, because more talk is positively but nonlinearly related to more unique words, less talk in a sample need not be the source of lexical simplification of that talk, as demonstrated by our simulations. It is not the case that contingent talk is simply in need of 'catching up' with non-contingent talk in terms of linguistic complexity. From our simulation, it is estimated that no amount of time would allow contingent talk to become as linguistically complex as non-contingent talk. Contingent talk follows a different type-token curve, or a different trajectory of linguistic complexity as a function of linguistic activity.

Previous highlights of the limitations of using type-token ratio measures should not discourage the use of type-token curves. Instead, we hope to highlight practical uses of these curves that encourage new techniques to query the nature of word distributions in children's language environments. To be sure, it is correct to say that type-token ratios are confounded by sample size. However, as pointed out in our simulation, it is equally correct to say that it is possible to query and compare distributions of words even if the sizes of the datasets are unequal. New insights can be revealed through the use of type-token ratio curves, in conjunction with size-matched control curves to demonstrate what influence sample size has over the nature of the curve.

Our simulation estimates that over long timescales, parents' responses to infants' babbling will generally contain more repetition of words. The effects of repetition on language development are mixed at longer timescales (e.g., 6 months to a year); more repetition has been associated with both lower vocabulary later in development [20] and with larger vocabulary [6]. Over shorter time-scales, more repetition in contingent parent speech predicts less advanced infant vocalizations [7]. The effects of repetition on language development likely depend on the type of repetition that is used. Partial repetition in the form of variation sets, in which adjacent utterances contain some words in common, facilitates statistical learning in adults and could promote language learning in infants [21].

In our view, immature vocalizations create learning opportunities by eliciting social responses that contain simplified, learnable information. These findings have important implications for current large-scale data collection and intervention studies on language development. Sophisticated and useful data on changes in linguistic structure as a function of contingent timing can be gleaned from home recording efforts that are currently focused on turn-taking and other forms of parent-infant interaction [22,23]. Several interventions for at-risk families currently focus on increasing the number of words parents say (e.g. Providence Talks; http://www.providencetalks.org) or turn-taking interactions with infants [24] but have not, to the best of our knowledge, focused on the relevance of learning distributional and temporal properties of parents' speech to infants.

Our results suggest that infants, via their immature vocalizing, play an important role in shaping their own language environment. Accurate prediction of environmental changes, an underlying learning mechanism in computational models of vocal learning, may also support infant learning in social contexts [25,26]. Models based on curiosity choose to learn from data over which they can minimize the error of their own predictions at the highest rate. By vocalizing, infants have the opportunity to observe the effects of their vocalizations on parents. Over their first year, infants quickly come to associate their immature vocalizations with responses from their parents [5]. Eliciting mature speech sounds from caregivers may become the target of infants' curiosity and subsequent guidance of vocal development. For more advanced understanding of early infant learning, future experimental, large-scale observational, and computational research should incorporate the affects infants have on the temporal and distributional properties of parents' speech.

ACKNOWLEDGMENTS

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

REFERENCES

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

[2] 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.

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

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

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

[9] 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.

[10] 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.

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

[12] A. Weisleder, and A. Fernald, "Talking to children matters early language experience strengthens processing and builds vocabulary," Psych. Sci., vol 24(11), pp. 2143-2152, 2013.

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

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

[15] 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.

[16] 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.

[17] D. Malvern, B. Richards, N. Chipere, and P. Durán, Lexical Diversity and Language Development. New York: Palgrave Macmillan, 2004.

[18] 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.

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

[20] 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.

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

[22] R. Romeo, J. Leonard, S. Robinson, M. West, A. Mackey, M. Rowe,and J. Gabrieli, "Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function," Psych. Sci., vol 29(5), pp. 700-710, 2018.

[23] 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, 2016, p. 128.

[24] 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.

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

[26] 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