Capturing Qualitative Variability in Early Overhearing Experiences

Ruthe Foushee[1], Grace Horton[1], and Mahesh Srinivasan[1]

[1]Department of Psychology, University of California, Berkeley, 2121 Berkeley Way West, Berkeley, CA 95720

Capturing Qualitative Variability in Early Overhearing Experiences

Inspired by qualitative studies typically limited to child-directed speech, we develop a coding scheme designed to characterize *all* utterances accessible to two English-learning children in terms of their relative utility for word-learning. We focus in particular on contributors to *referential transparency* as a well-established and meaningful dimension of language learnability in context. These include the spatial positions of the caregivers and children, the caregivers' use of gaze or gesture to illustrate their meaning, the child's visual access to the caregiver or to the referent of the utterance, and the caregiver's use of modified prosody. As a proof of concept, we apply this coding scheme to existing naturalistic video corpora for one English-learning child whose language development is well-documented. We find that both speech directed *to* the child and speech overhearable by her are highly variable along the qualitative dimensions we coded, and identify the heterogeneity of overheard speech as a source of noise in previous investigations. While irrelevant as a referential cue, our results suggest caregivers' prosodic modification may play a functional role in marking speech intended for the child — especially given the significant qualitative overlap between overhead and child-directed speech along other dimensions. In spite of the frequent similarity between overheard and child-directed speech, overheard utterances were significantly less associated with child attention. Taken together, our results shed light on how adults and children co-structure the early language environment, and promise to provide similar insights when applied to naturalistic video corpora for children across the world.

Introduction

Despite substantial research on (a) differences in the contributions of child-directed versus overheard speech to vocabulary size (Ramírez-Esparza et al., 2014; Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012; Weisleder & Fernald, 2013), and (b) the impact of qualitative differences in child-directed speech contexts on vocabulary acquisition (e.g., Cartmill et al., 2013; Hirsh-Pasek et al., 2015; Ramírez-Esparza et al., 2014), to our

knowledge there remains no systematic study within these literatures of qualitative differences in *overhearing* contexts as they relate to learning. This is important because overheard speech is common across the world (e.g., Casillas et al., 2019; Sperry et al., 2019), and because overheard speech is likely to be a more heterogeneous category than child-directed speech, such that understanding the range of roles it may play in young children's lives is critical and not straightforward.

In quantitative studies correlating amount of speech with vocabulary size, both child-directed and overheard speech are treated as monolithic. However, speech directed *to* the child is likely to be a much more coherent category than speech *around* her, which might be directed to variable audiences, at variable distances, and with variable relevance to the child. The two categories of speech (child-directed versus overheard) undoubtedly differ in their overall rates of features that we know children can use to solidify word–referent mappings. Mindful of this, our study takes inspiration from previous studies of input *quality*, where researchers unpack the influence of child-directed speech by hand-coding qualitative aspects of naturalistic audio or video recordings, often with the intention of relating that variability to metrics of children's development of language (e.g., Hirsh-Pasek et al., 2015; Ramírez-Esparza et al., 2017; Rowe et al., 2004; Rowe et al., 2016).

In studies using qualitative coding schemes, utterances with the same token count, and even same ratio between types (unique words) and tokens, might be found to differ along some social-contextual dimension that we know is relevant for learning. Previous work analyzing such qualitative diversity has found, for example, that amount of speech not only directed to the child, but specifically one-on-one and in the sing-songy register of so-called *parentese*, is predictive of vocabulary growth (Ramírez-Esparza et al., 2014), as is caregivers' tendency to use nouns when the noun's referent is highly salient or easily inferred from context (Cartmill et al., 2013). Notably, fine-grained coding schemes of this nature have historically been applied exclusively to speech that is child-directed, leaving a

gap in the extant literature. Here, we develop a coding scheme that will enable us to characterize the full range of linguistic inputs experienced by children across contexts, and to analyze their relative utility for language-learning. We initially apply this coding scheme to an existing naturalistic English-language video corpus, corresponding to a target child whose language development is well-documented. However, our system is designed to be used to capture the richness and latent structure within the early language environments of children across contexts, cultures, and languages.

Specifically, we use longitudinal samples of the language environment of a single child to ask five questions regarding overheard language quality:

(1) How does the *quality* — in terms of hypothesized utility for language learning — of overheard speech compare to the quality of child-directed speech?

(2) How does the qualitative *variability* of overheard speech compare to the qualitative variability of child-directed speech?

(3) In a naturalistic context, how *distinguished* are overheard and child-directed speech?

(4) How does the quality of child-directed and overheard speech change as the child matures?

(5) What aspects of speech quality are associated with child attention?

Here, we focus on *referential ambiguity* as a meaningful and well-studied dimension of individual utterances that is reliably associated with learning.

Method

Sample Selection

We selected videos (Databrary.org) and transcripts (https://phonbank.talkbank.org/browser/index.php?url=Eng-NA/Providence/) from the Providence corpus to explore qualitative differences in varieties of adult speech. The

Providence corpus was collected by Demuth and colleagues (2006) as part of a longitudinal study of phonological development, and documents the early language development of six children, approximately one year in age at the time of enrollment. Data collection for the children in the corpus began at the onset of children's first words, after which they were videotaped in their homes for one hour every two weeks, for up to three more years. Our only prerequisite in selecting videos to analyze was that there be at least two adults present during the recording, and that it include multiple adult–adult conversational turns. This ended up being highly constraining, as videos for five out of the six children largely recorded single adult–child dyads, effectively narrowing our sample from six children to one.

The recordings that we ultimately analyzed for this case study, then, represent hour-long samples of naturalistic speech from the home of a single child, Naima, across the first three years of her life. Naima is one of the most densely sampled children in the corpus: she and her family contributed an impressive 88 sessions total, spanning the time just before Naima's first birthday (00;11;27), to a couple months before her fourth (03;10;10). Of these videos, 12 met our criteria for inclusion.

Procedure

We used Datavyu (Team, 2014) to code the sample of videos. Transcripts of the relevant sessions were downloaded from the CHILDES database, and used to populate time-locked coding cells, organized by speaker. Two coders were responsible for coding eleven out of the twelve transcripts. Coders were responsible for alternating videos with respect to the age of the child, so that potential inconsistencies in coding were not confounded with child age. As the codes hinged on an understanding of the pragmatic context of the utterances, coders watched each video in full before coding the utterances. Coders also used this initial viewing to annotate coded dimensions that typically spanned multiple utterances, including the context of the interaction and the physical position of the child. In the critical coding pass, coders entered values for each of the qualitative dimensions described below, for all adult utterances in the recording. Ambiguity in the

application of the codes was rare by design, as any dimensions triggering disagreement in previous adult coders were dropped before finalizing the scheme. When uncertainty did arise, the primary coder(s) and first author reviewed the video to reach a decision. In cases where the dimension could not be coded (i.e., where the utterance was inaudible, or the speaker or child were out of frame), the code for that utterance was marked as 'na.'

Coding Scheme

Speech Audience

We used a combination of pragmatic cues to code the audience to whom each utterance was directed, including: (1) the content of the utterance, (2) the surrounding linguistic context, (3) the gaze of the speaker, (4) the focus of attention of the scene participants, and (5) the physical positions of the speakers, combined with the relative volume or force of the utterance. The audience of the utterance was coded as 'target child,' 'adult,' or 'phone.' Utterances receiving the latter two codes were classified as *overheard speech*. Our method of classifying overheard speech differs from most previous studies in that it is coded on a by-utterance basis, rather than generally across segments of speech (e.g., Weisleder & Fernald, 2013), and in that it includes adult phone conversations that take place when the child is within earshot (*contra* e.g., Shneidman et al., 2013).

We next coded a set of six qualitative features for each utterance individually. The coding scheme is based on evidence for qualitative dimensions of spoken language associated with heightened child attention at Naima's age, and/or cues that children can reliably use to resolve referential ambiguity and learn new words (e.g., Cartmill et al., 2013; Cooper & Aslin, 1990; Golinkoff et al., 2015; Golinkoff & Hirsh-Pasek, 2006).

Here & Now Reference

Coders indicated whether the utterance described or referred to the current environment. Utterances referring to the "here and now" are argued to make the task of word-learning easier (e.g., Ellis & Wells, 1977), particularly early in the course of

acquisition. "Here and now" coding reflects the intuition that if "coffee" is an unfamiliar word, it will be easier to learn when Naima's family is in the kitchen and her mother says, "Yum, Daddy's drinking coffee," than when Naima and her mother are in the living room when her father comes home, and Naima's mother says, "Daddy was shopping, he was looking for coffee." One prominent view in the literature emphasizes the information available in the syntax of an utterance (Gleitman et al., 2005). If a sentence is about the immediate context of the utterance, the child can use the relational structure implied by syntax to parse the scene and infer the meanings of new embedded words (Hoff & Naigles, 2002; Naigles, 1990). More generally, assuming that speech refers to the "here and now" is a sensible starting hypothesis for a learner, with the implication that learning will be enhanced when that assumption is met (Mervis, 1983; Shatz, 1978).

When coding utterances about the "here and now," coders further distinguished between utterances where the child was visibly attending to the relevant part of the scene, and utterances where she was not.

Referential Gesture

"Referential gesture" was coded as present when an utterance was accompanied by non-verbal cues to its reference (Baldwin et al., 1996; Booth et al., 2008; Brooks & Meltzoff, 2008; Frank et al., 2013; León, 1999; Slobin, 1985). Referential gesture occurred in a variety of forms: when Naima's mother leans in to pull a fine thread off Naima's tongue and says, "You had a hair in your mouth," and lifts the hair before Naima's eyes, but also when Naima's mother points at the laundry basket in conversation with Naima's father, or looks toward the fridge when discussing dinner plans, or even when she mimes sleeping when whispering about a nap. Thus, referential gesture coding considered a more expansive *locus of reference* (Ellis & Wells, 1977), and captured distinct information from the "here and now" code.

Child Gaze Toward Speaker

Caregiver's referential gestures might be lost on Naima if she were not attending to the speaker. Thus, we additionally coded Naima's visual attention to the speaker (Bakeman & Adamson, 2019; Grassmann et al., 2015).

Sing-song Prosody

This code captured whether the utterance had the cadence or exaggerated prosody typical of infant-directed speech (Fernald & Kuhl, 1987; Saint-Georges et al., 2013; Snow & Ferguson, 1977; Soderstrom, 2007), and best reflected pitch *variability*. Previous work suggests that this dimension attracts and maintains infants' attention, resulting in enhanced learning of associations between, e.g., visual and auditory stimuli (Cooper & Aslin, 1990; Kaplan et al., 1996; Ma et al., 2011), or of mappings between sound and meaning (Graf Estes & Hurley, 2013).

In addition to the above binary features, we analyzed three continuous measures of speech quality, auditory clarity, morphological complexity, and utterance length.

Auditory Clarity

We rated the auditory *clarity* of the utterance (e.g., Fernald & Simon, 1984), from 0 (inaudible) to 3 (clear). Of course, clarity for Naima may be different than for the coder viewing the tape. However, likely because the original study (Demuth et al., 2006) targeted phonological development, the camera placement was always designed to optimize the recording of Naima's productions. Therefore, the recording audio may provide a more accurate reflection of the child's own auditory experience than if our data had come from a study with a different intent. We note that some dimensions could still be coded, even for "inaudible" utterances, as in the case of inaudible speech on the phone.

Morphosyntactic Complexity

To capture trends in structural complexity, we used the pre-existing annotations of morpheme and token counts for each utterance to analyze *utterance length*, as well as to

compute a measure of "morphological complexity" that increased with the ratio of morphemes to tokens (MacWhinney, 2008). For example, the utterance, "Oh baby, sorry," with three morphemes and three tokens, receives a morphological complexity score of 1, while the utterance "Why're you growling," with five morphemes and three words, receives a score of 1.67.

Results & Discussion

Summary data for all dimensions can be found in Appendix . To assess the reliability of our coding scheme, an independent research assistant coded two especially dense thirty-minute segments of videos from the first and fourth quartiles of our age range. Agreement was typically high ('Here and Now': 97%, Referential Gesture: 100%, Sing-song Prosody: 70%, Speech Type: 97%, Child Gaze toward Speaker: 78%).

Distribution of Utterances

Both speaker and child were visible for 56% of all utterances, enabling complete coding for 3,801 utterances. In an additional 16% of utterances (1,075 total), the child, but not the speaker, was in frame, enabling coding of child position and gaze, but not referential gesture. We include all coded utterances in our analyses; scripts for accessing transcripts from the CHILDES database (childes-db; Sanchez et al., 2018), populating Datavyu coding spreadsheets, and all data analyses can be found at https://osf.io/hy5z2/.

Speech Context and Child Position

Utterances occurred most frequently in contexts coded as "play time" (71.2% of utterances), followed by "meal time" (22.6%), "bath time" (3.8%), and "bed time" (2.4%). An average of 2 contexts occurred in each video. The child was typically seated in a high chair (31.7% of utterances) or standing (26.7% of utterances). There was insufficient variability in early videos to support further analyses of the relation between the child's physical position and her language environment.

Speech Type

Despite selecting videos based on the presence of overheard speech, the majority (85.3%; 5,643 utterances) of utterances were child-directed. Overheard speech accounted for 12.4% of all utterances (773 utterances between adult caregivers, and 102 utterances over the phone). A remaining 2.3% (184 utterances) were uncodeable, all due to sufficiently poor audio quality that the original researchers had not been able to transcribe their content.

The infrequency of overheard speech may accurately reflect the statistics of the child's environment, or may reflect the original study's focus on the child's verbal *production*, leading Naima's mother to engage in more speech-eliciting behaviors during recordings than she might otherwise. Speech was also not equally distributed across caregivers: Naima's mother accounts for 71.2% of all utterances (64.0% or 4,310 utterances in child-directed speech and 8.3% or 558 utterances in overheard speech), while Naima's father accounts for 23.5% (19.8% or 1,333 child-directed, and 4.6% or 312 overheard). We collapse across utterances from both caregivers in our analyses, and structure the results below according to our primary research questions.

Is overheard speech less *learnable*?

Below, we compare child-directed and overheard speech along the qualitative dimensions designed to capture the *referential transparency* of each utterance in context, as well as caregivers' structural simplification of their speech.

All coded features of caregivers' utterances in context were reliably different between speech directed to Naima and speech that Naima could overhear (all ps $< .001$; see Table 2). Interestingly, overheard utterances were not uncommonly about the "here and now" ($M = 0.28$), though very infrequently combined with a referential gesture that the child could use to identify that this was the case ($M = 0.02$). The absence of referential gesture may partly explain why Naima rarely gazed toward the referent in overheard

speech, even when the utterance was about the here and now (Figure 1). Alternatively, the disparity between speech types in how regularly Naima looks at the co-present referent might indicate that Naima's parents talk about "here and now" objects *because* Naima is looking at them. Overheard speech was also typically rated high for clarity ($M = 2.64$), suggesting that even when parents were speaking with one another or on the phone, they maintained proximity to Naima. This is consistent with Naima's tendency to look at the overheard speaker ($M = 0.44$), which we might expect to be reduced if the speaker were further away.

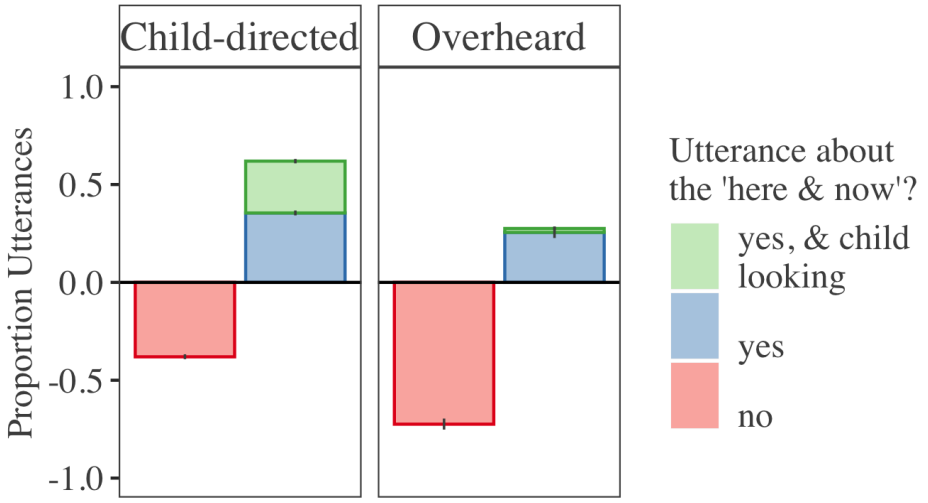

Figure 1
Semantic Accessibility in Child-directed and Overheard Speech.

Child-directed and overheard speech were also reliably distinguished along structural dimensions, although variably so. Child-directed utterances were consistently shorter ($M_{\text{tokens}} = 4.59$) than overheard utterances ($M_{\text{tokens}} = 6.48$). However, the two often overlapped along our measure of morphological complexity, suggesting that while caregivers tended toward shorter utterances, they did not refrain from inflecting the words they used.

Differences in what we coded as "sing-song prosody" are the most dramatic in our data. "Sing-song prosody" characterizes almost all child-directed utterances ($M = 0.98$), and 11% of overheard utterances. The frequency of exaggerated pitch variation in speech

directed *to* Naima is not surprising, as the videos start when Naima is still an infant (Cristia, 2013; Soderstrom, 2007; Spinelli et al., 2017). However, Naima's parents' prosodic modification when addressing *each other* is unexpected, and may speak to the competing demands they experience as caregivers of a small child. For example, Naima's videos reveal various motivations for one of her caregivers to be in sustained physical proximity to her (e.g., to feed her in her high chair, or to prevent her from climbing precarious furniture, breaking something, or dissolving into tears). This means that, for much of the day, if Naima's parents also need to have a conversation, Naima will be present for it (and potentially experiencing a reduction in the attention she so recently enjoyed). Thus, adults in such contexts may be driven to "multi-task" in their speech production, using word meanings and syntax to transmit their *messages* to their partners, and using melodic prosody to signal their continued care and awareness to their infants. Consistent with this interpretation of caregivers' verbal behavior, "sing-song" adult-directed utterances (e.g., "Daddy what's today's date, is it the twenty-first?") in our data were equivalent to unmodulated adult-directed utterances in terms of length ($M_{\text{tokens}} = 5.59$) and morphological complexity ($M = 1.28$). We return to caregivers' instrumental use of prosody as a signal in the General Discussion.

Is overheard speech more *variable*?

We predicted that overheard utterances would comprise a more heterogeneous category than child-directed utterances. We explore this prediction in two ways. Figure 2 plots the standard deviations for each qualitative variable by speech type, controlling for age. Panels where the point on the righthand side is higher than the point on the lefthand side suggest greater variability along that particular dimension within the set of overheard utterances.

For a better sense of the reliability of this difference — especially in light of the difference in the size of the two datasets — Figure 3 plots the frequency distributions of each binary feature in 1,000 bootstrapped samples of each dataset, and Figure 4 does the

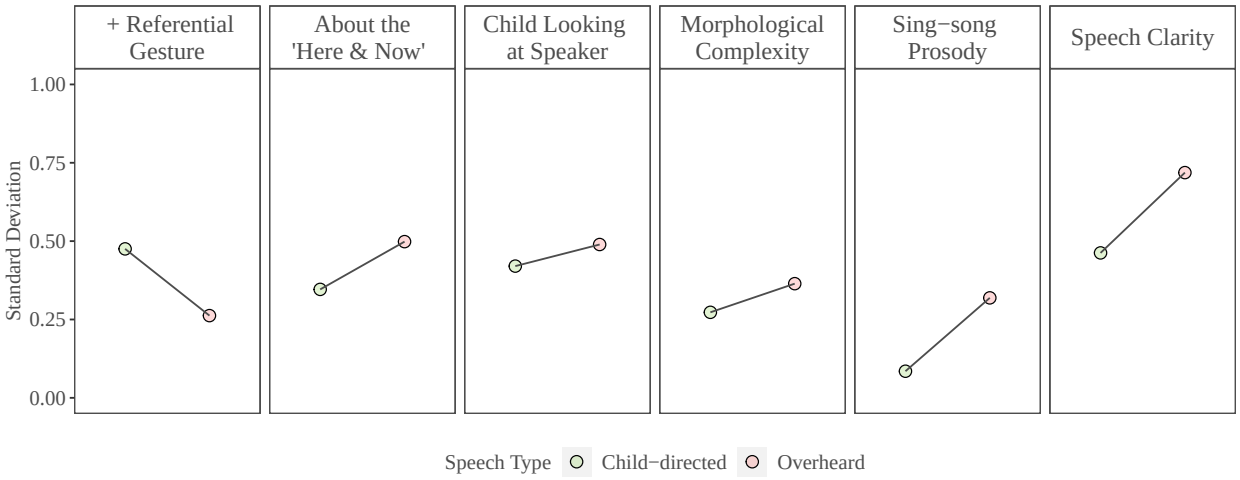

Figure 2
Standard Deviations for Qualities in CDS and OHS.

same along the continuous dimensions we coded. The width of each distribution gives a sense of the reliability of our frequency estimate, based on our dataset, while its horizontal position gives a sense of the overall rate or value range of that feature. Together, these analyses suggest that overheard and child-directed speech are reliably differentiated in their prosodic modification, tendency to describe the current environment, and correspondence with the current target of the child's visual attention — in terms of both an utterance's referent and its speaker. However, they are more frequently similar in their co-occurrence with referential gesture, clarity, and utterance length. Importantly, Figure 3 suggests that neither speech type is entirely predictable in its degree of referential transparency.

Are child-directed and overheard speech reliably distinguished?

Our third hypothesis concerned how distinguishable overheard speech is from child-directed speech in a naturalistic context. Again, we tested this in two ways. We first fit a logit model to the data, using our coded variables to predict the type of the speech (coded as *overheard speech* = 0, *child-directed speech* = 1) to which each utterance belonged. The model included age, "here and now" reference (a categorical variable with three levels: "no," "yes, but not looking at the referent" and "yes *and* the child's gaze is on

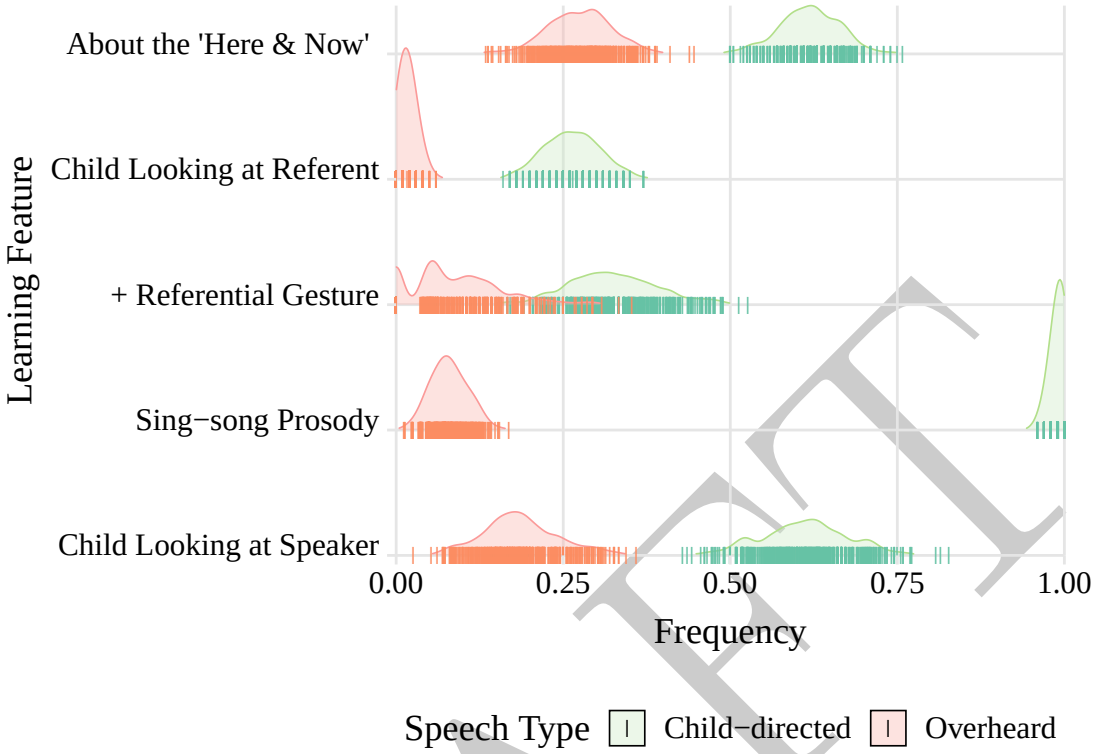

Figure 3
Binary Feature Frequency in Resampled Distributions of Utterances.

the referent"), referential gesture (0, 1), whether the child was looking at the speaker (0, 1), "sing-song prosody" (0 = absent, 1 = present), clarity (rating 0–3), and morphological complexity (computed values 0–3).

Odds ratios and 95% confidence intervals for this model are shown in Table 3. Whether the speech was about the "here and now" was a significant predictor of child-directed status (OR = 1.80 [0.79, 4.10]; $\chi^2(2) = 19$, $p < .001$), especially when the child was currently looking at the referent (OR = 8.88 [3.09, 27.00]). The child's concurrent gaze toward the speaker was also associated with child-directed speech status (OR = 2.42 [1.16, 5.10]; $\chi^2(1) = 6$, $p = .018$). Of all variables measured, prosody was the most reliable predictor of child-directed speech status (OR = 369.76 [190.68, 769.50]; $\chi^2(1) = 521$, $p = .001$). Finally, neither referential gesture ($\chi^2(1) = 1$, $p = .250$), speech

clarity ($\chi^2(1) = 0$, $p = .740$), utterance length ($\chi^2(1) = 1$, $p = .385$), morphological complexity ($\chi^2(1) = 1$, $p = .359$), nor age ($\chi^2(1) = 4$, $p = .058$) were reliable predictors of whether the utterance was child-directed.

To further evaluate the distinguishability of child-directed and overheard speech, we conducted a linear discriminant analysis using the MASS library in R (Ripley et al., 2013), with a uniform prior on whether each data point was child-directed or overheard. We used only the coded speech variables that were not contingent on the child's own attention or behavior (that is, we included acoustic, semantic, and morphosyntactic variables, but not whether the child was looking at the speaker or referent). The loadings for each variable in the single linear discriminant function appear in the first column of Table 4. Echoing previous results, "sing-song prosody" was almost entirely responsible for distinguishing child-directed from overheard speech, with reference to the "here and now" serving as a very distant second. Referential gesture, speech clarity, utterance length, and morphological complexity did little to contribute to the between-group variance captured by the function.

Interestingly, child-directed speech was better identified than overheard speech,

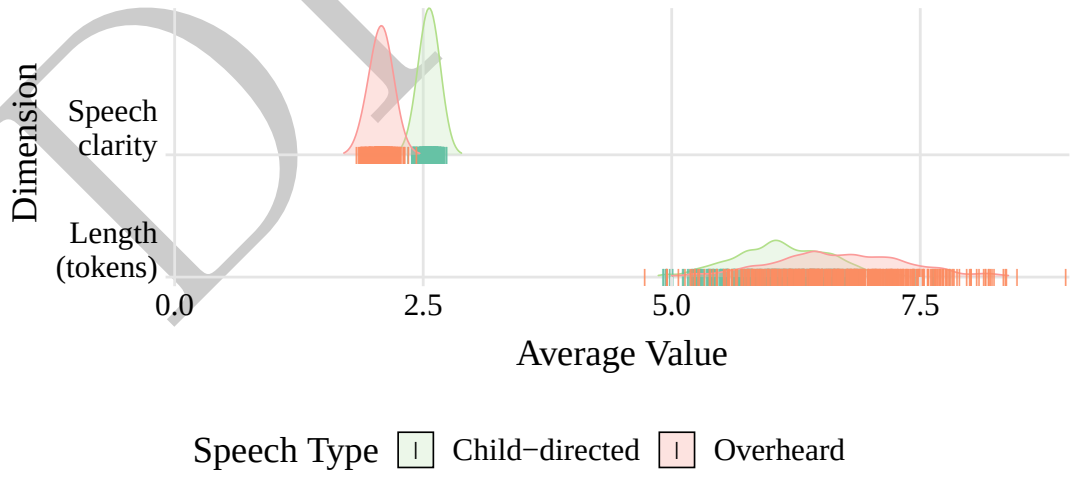

Figure 4
Average Continuous Feature Values in Resampled Distributions of Utterances.

which bears directly on our hypothesis of greater within-class variance for speech that could be *overheard* by the child relative to speech directed *to* her. To assess the accuracy of the linear discriminant, we withheld 25% of the raw data as a test set. The function accurately classified 89% of the overheard utterances in our test set, and 99% of child-directed utterances, with an overall error rate of less than 1% (0.91%). Removing "sing-song prosody" from the function (loadings shown in column (2) of Table 4) and repeating the procedure with the same training and test data illustrates the critical contribution of prosody to distinguishing speech intended for the child in this household and age range. Without information about prosody, only 66% of overheard and 77% of child-directed utterances were accurately classified, with an increased error rate of 24%.

Does overheard speech change with child age?

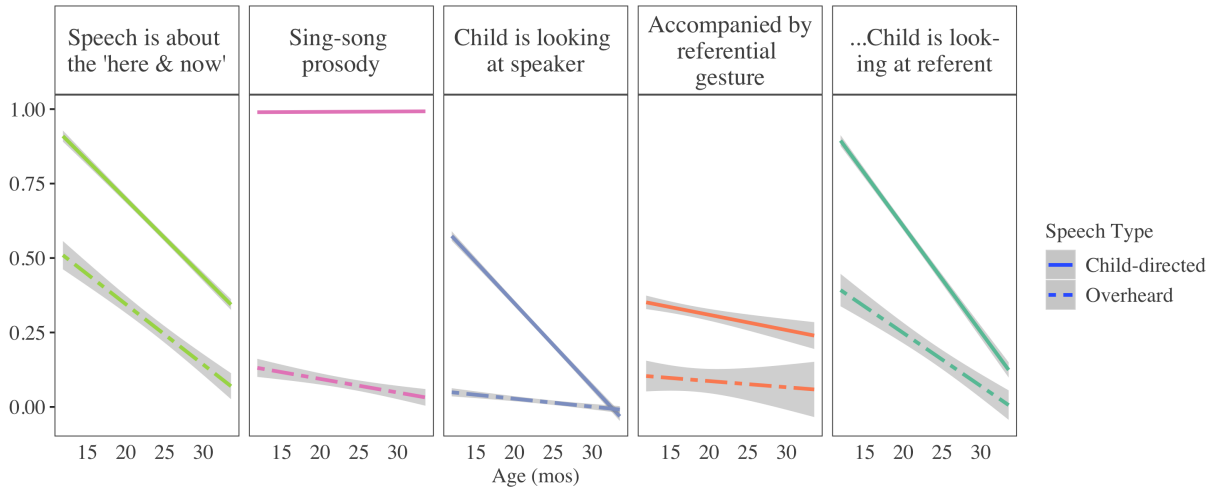

Figure 5
Feature Frequency across Child Age.

In child-directed speech, referential cues typically associated with caregiver modification like speech about the "here and now" ($r = -.5$, $[-.53, -.47]$, $p = .01$), utterance length ($r = .07$, $[.05, .10]$, $p < .001$) and referential gesture ($r = -.05$, $[-.1, -.01]$, $p = .05$) were correlated with child age, as was the likelihood that the child was looking at the speaker as they were talking ($r = -.57$, $[-.6, -.54]$, $p = .01$). Remarkably, qualitative *overheard* features also showed correlations with age. As in child-directed

speech, caregiver talk about the "here and now" was negatively correlated with age ($r = -.43$, $[-.55, -.29]$, $p = .01$), along with the child's tendency to be looking at a speaker as they talked ($r = -.33$, $[-.46, -.18]$, $p = .01$). In contrast to the interpretable pattern of increasing utterance length in child-directed speech, in overheard speech, utterance length was negatively correlated with age ($r = -.14$, $[-.19, -.08]$). We speculate that this may reflect caregivers conducting fewer full-fledged conversations in Naima's vicinity, and exchanging more brief, functional utterances, more frequently interrupted by their now-verbal daughter. Finally, referential gesture in overheard speech was not correlated with child age, and in neither speech type was caregivers' prosody or morphological complexity related to the age of the child. This is surprising, as previous work suggests that caregivers' exaggerated prosody decreases as the child matures (Bornstein et al., 1992; Cooper & Aslin, 1990), while morphological complexity increases (Ervin-Tripp, 1978; Huttenlocher et al., 2007; Sherrod et al., 1977). We speculate that our result might be a reflection of (a) Naima's age in the study, and/or (b) Naima's caregivers' awareness that they were being recorded, which might have caused them to exaggerate the child-directed features of their speech. To further explore correlations among contextual features of the learning environment and age, please see Appendix .

Finally, we fit models to the overheard and child-directed data for each binary speech quality, with age as the sole predictor. Exponentiated coefficients and confidence intervals for the effect of age are shown in Table 5.

Does overheard speech attract children's attention?

To better understand relations between speech qualities and child attention, we created a new, "child attention" variable that indexed whether the child was looking at either the speaker or the referent of an utterance. We fit another logit model to the by-utterance data, including all other speech qualities as predictors (see exponentiated coefficients and 95% confidence intervals in Table 6). "Sing-song prosody" was highly predictive of child attention (OR = 5.87 [4.34, 7.99], $\chi^2(1) = 146$, $p = .001$), as was "here

and now" reference (OR = 3.83 [3.14, 4.68], $\chi^2(1) = 173$, $p = .001$). Speech clarity was also associated with child attention (OR = .71 [1.48, 1.98], $\chi^2(1) = 54$, $p < .001$), again possibly speaking to the role of proximity in eliciting or following the child's attention. Interestingly, age was negatively related to child attention (OR = 0.90 $[0.89, 0.91]$, $\chi^2(1) = 352$, $p = .001$). This may likewise reflect increased independence and distance from her caregivers, or even increased capacity to distribute her attention, such that she can comprehend her caregivers' meaning without needing to look at them or the scene. Indeed, if Naima's prior gaze meant that she was seeking the referent of her caregivers' utterance, she will need to do so less with greater word knowledge.

Our structural variables were the only measures *not* reliably associated with Naima's visual attention (utterance length: OR = 1.01 [0.99, 1.04], $\chi^2(1) = 1$, $p = .23$; morphological complexity: OR = 1.17 $[0.84, 1.64]$, $\chi^2(1) = 1$, $p = .36$). At face value, this result might appear to cast doubt on our premise of complexity as a key driver of child attention and learning. However, we suspect it might say more about the sensitivity of this measure of complexity. If nothing else, that our calculation of "morphological complexity" showed no relation to Naima's age in child-directed speech — especially during this critical period of linguistic development — suggests that it may be ill-suited to capture the meaningful variation in language structure that we would expect to influence attention.

General Discussion

Theories of early learning suggest that children's attention is motivated by an ongoing sense that they are making sense of incoming data (e.g., Balcomb & Gerken, 2008; Gerken et al., 2011; Houston-Price & Nakai, 2004; Hunter & Ames, 1988; Hunter et al., 1983). Our study analyzed the degree to which different sources of spoken language in a child's daily environment might support that sense, focusing especially on support for learning new words. We homed in on *referential transparency* as a demonstrably important dimension of language learning contexts that could be coded from video, and took a case study approach, capitalizing on longitudinal video recordings documenting the language

environment of a single child — Naima, from the Providence corpus (Demuth et al., 2006). We analyzed over six thousand utterances spanning the first two years of Naima's life, when cues to words' meanings are argued to be especially critical for language development (e.g., Cartmill et al., 2013).

Our study is rare in considering the quality or learnability of *all* of the speech in the language learner's environment, including adult conversations that take place when the child is nearby, and even caregiver phone calls ('halfalogues'; Emberson et al., 2010). By applying the same qualitative coding scheme to caregiver utterances coded as 'child-directed' versus 'overheard,' we find that greater referential transparency characterizes the set of utterances spoken to Naima directly. Child-directed utterances were more frequently about Naima's immediate context, rather than the past, future, or another place, and more frequently coincided with her current focus of attention. Child-directed utterances were also more frequently accompanied by physical behaviors like pointing and pantomime, which Naima could use to infer her caregivers' communicative intentions.

That the child-directed speech in our study appears highly supportive of word-learning is concordant with findings in other samples that the amount of child-directed speech that children receive during this period predicts their later vocabularies (e.g., Huttenlocher et al., 2010; Ramírez-Esparza et al., 2017; Rowe, 2012; Shneidman et al., 2013; Shneidman and Goldin-Meadow, 2012; Weisleder and Fernald, 2013; see Hoff, 2006 for a review). One popular empirical approach uses correlations between measurements of the language in children's homes and children's language development to make inferences about the 'effectiveness' of particular forms of language data and/or linguistic interaction. These studies typically limit their analyses to language addressed directly to children, rather than consider the range of language sources that young children experience over the course of a day. The rare studies that also analyze speech addressed to others consistently find no statistical relation between the amount of

overheard speech regularly available to a child, and that child's level of language development (Ramírez-Esparza et al., 2017; Shneidman et al., 2013; Shneidman & Goldin-Meadow, 2012; Weisleder & Fernald, 2013), inviting researchers to conclude that children "do not readily make use of overheard input when learning words in naturalistic situations" (Shneidman et al., 2013, p. 7). Our study was partly motivated by a potential measurement issue in these investigations: namely, that the heterogeneity of overheard speech might introduce a significant amount of noise into correlational measures of learning. Not only are overheard utterances likely to be highly diverse, but we reason that child-directed utterances are likely to represent a significantly more homogeneous category in precisely the contexts where child-directed speech is typical — and typically marked.

We see this study as confirming the hypothesis that overheard speech represents a less coherent category than speech directed to children. However, our data also show significant *within*-category variability for child-directed speech. These data are consistent with claims by language development researchers that input *quantity* (i.e., the total number of words the child hears) predicts language outcomes by virtue of input *quality* (i.e., dimensions of learnability like those we code here; Cartmill et al., 2013; Hart and Risley, 2003; Rowe et al., 2017). The idea is roughly that greater 'quantity' means a greater number of samples from the frequency distributions in Figures 3 and 4, and with those samples, greater opportunities for individual high-quality learning episodes. Our fine-grained coding of the learning opportunities afforded by the overheard speech within a single child's home suggests that high-quality exposures to new words are not limited to child-directed utterances; however, they may be less likely to co-occur with the child's current focus of attention when overheard. This observation suggests new avenues of research: for example, how might children's attention be conditioned by the relative frequency and quality of child-directed versus overheard speech in their environments?

That both child-directed and overheard speech were highly variable suggests a

functional role for prosodic modification in discriminating two language sources that might be less naturally distinguished than previously thought. The decisive role of prosody in identifying speech as intended for the child was borne out in our analyses, where classification error by a linear discriminant function skyrocketed when information from prosody was removed (Section). Caregivers' prosodic modification is especially interesting in light of our study's focus on *referential transparency*. In contrast to, for example, eye gaze or pointing (gestures we explicitly coded as "referential"), variable pitch does not in itself provide a disambiguating cue to reference. That is, while your mother's gesture to your father's coffee cup might help you infer the meaning of /kafi/, her melodic pronunciation does not. Learners may make use of caregivers' non–adult-like prosody not to decrypt language itself, as has been suggested in previous literatures; instead, prosody may be understood as a learned — and self-reinforcing — cue, as our data suggest that infants' attention has a higher probability of being rewarded when an utterance is intended for them. Here again, Figures 3 and 4 provide a useful illustration of this point: prosody may mark an utterance as coming from the *green* distributions, which offer greater promise for learning — motivating simultaneously children's selective attention to child-directed speech and inattention to ambient overheard speech. This perspective is consistent with evidence that infants whose caregivers do not habitually acoustically exaggerate their speech show weaker or absent preferences and learning benefits from hearing exaggerated infant-directed speech in the lab (see e.g., Cristia, 2013; Soderstrom, 2007, for reviews). It is also reminiscent of evolutionary accounts of infant-directed song as a way for caregivers to signal attentional investment to their infants from afar (Mehr & Krasnow, 2017).

Nonetheless, we note that these results also come with a caveat, as our assessment of caregivers' "sing-song prosody" was highly impressionistic, potentially leading the code to reflect something like "child-directed register," rather than prosodic modulation, *per se*. In support of this hypothesis is the relatively low agreement between our initial coding and an independent reliability coder (70%) — though the fact that both parties also identified

"sing-song prosody" in utterances coded as "overheard" suggests that they were not basing their acoustic assessment entirely on intended audience. Our coding of caregivers' prosody was also notably independent of Naima's age, despite the well-documented observation that caregivers typically reduce their acoustic exaggeration as children mature (e.g., Henning et al., 2005; Smith & Trainor, 2008). While it is possible that Naima's caregivers persisted in exaggerated 'baby talk' for the entirety of our study, it is also possible that they gradually reduced their exaggeration, but that this continuous trend was obscured by our binary "sing-song" code. To address these concerns, ongoing work further grounds our scheme in objective proxies for theoretically important variables. For example, to capture caregivers' prosodic modification, we use variability in the first formant of clips of speech, quantified via acoustic analysis software, rather than subjective coding of the "sing-song" quality of caregivers' utterances (Cristia, 2013).

Conclusion

Even in the absence of information about individual children's language outcomes, qualitative coding schemes like ours provide valuable vocabularies with which to describe early language environments, which are in turn useful for generating hypotheses and making contact with more humanistic fields like anthropology. Relative to child-directed speech, the unknowns of overheard speech are remarkably basic: how variable are a child's overhearing experiences over the course of a day, and how does both the scale and quality of that variation compare across ages, versus across households, versus across cultures? Evidence for claims about cross-cultural differences in linguistic and child-rearing practices have typically taken the form of ethnographies (e.g., de León, 1998; Heath, 1983; Ochs, 1982; Schieffelin, 1990; Ward, 1971), which provide rich descriptions of community customs and beliefs, but make systematic comparisons between contexts difficult. We cannot build theories about the mechanisms underlying language development without a sense of how universal versus idiosyncratic the language environments that developmental scientists typically study are (Frank et al., 2017; Lieven, 1994; Ochs, 1990). It is difficult to

understand how children transition to acquiring language in classroom contexts without understanding how the overheard input there — between the teacher and another student, or among nearby peers — compares to the overheard input before schooling. Likewise, we can make better hypotheses about how young children's attention is organized if we can find patterns among features of non–child-directed contexts, and understand how those environments vary in their support for child participation, observation, and apprenticeship (Rogoff et al., 2003). Language development research that continues to be focused on the impacts of child-directed speech may be missing nuances in how different environments are organized to support children's entry into the adult speech community (Leon, 1998; Ochs, 1990; Vogt et al., 2015). In providing a common vocabulary with which to describe diverse milieu, we aim to bring the psychological and anthropological literatures into contact, such that theories of language development can be tested against the full range of children's linguistic lives.

Table 1
Examples of Qualitative Overlap in Child-Directed and Overheard Speech

ccc

Speech Type	+ Qualitative Features
Child-directed	MOT: Mmmm we're eating our supper!" CHILD: Mmmm" MOT: Here it is! Here
Overheard	MOT: I packed you a towel and diaper and all that. FAT: Oh, good. MOT: I mean I'm

Table 2
Mean Values and Differences in Means (Child-directed − Overheard Speech)

Dimension	Child-directed	Overheard	Difference[†]
About the 'Here & Now'	0.34 (0.32, 0.37)	0.28 (0.25, 0.30)	0.34^{***}
Child Looking at Referent	0.26 (0.25, 0.28)	0.02 (0.01, 0.02)	0.25^{***}
Child Looking at Speaker	0.86 (0.85, 0.88)	0.44 (0.36, 0.52)	0.43^{***}
Referential Gesture	0.77 (0.75, 0.79)	0.09 (0.05, 0.13)	0.23^{***}
Sing-song Prosody	0.99 (0.99, 1.00)	0.11 (0.07, 0.17)	0.88^{***}
Speech Clarity	2.83 (2.81, 2.85)	2.64 (2.52, 2.75)	0.19^{***}
Morphological Complexity	1.21 (1.20, 1.22)	1.30 (1.24, 1.35)	-0.09^{***}
Utterance Length	4.59 (4.45, 4.74)	6.48 (5.72, 7.26)	1.90^{***}

[†] Observed difference in means (Child-directed − Overhead)
*** $p < 0.001$, via exact permutation test comparing observed difference in means to empirical null distribution.

Table 3
Logit Model Predicting Child-directed versus Overheard Utterance Status

	Dependent variable:	
	Child-Directed {0, 1}	
Constant	0.03	(0.002, 0.40)
Here & Now (Child Looking at Referent)	8.88^{***}	(3.09, 27.00)
Here & Now (Not Looking at Referent)	1.80	(0.79, 4.10)
Referential Gesture	1.77	(0.68, 5.00)
Child Looking at Speaker	2.42^{*}	(1.16, 5.10)
Sing-song Prosody	369.76^{***}	(190.68, 769.50)
Speech Clarity	1.11	(0.59, 2.00)
Morphological Complexity	0.59	(0.21, 1.80)
Utterance Length	0.97	(0.89, 1.00)
Age	1.06	(0.10, 1.10)
Observations		2,225
Log Likelihood		−167
Akaike Inf. Crit.		355

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 4
Linear Discriminant Functions Classifying Utterances as Child-directed or Overheard

Variable	+ Prosody Loading	− Prosody Loading
About the 'Here & Now'	0.19	1.57
Referential Gesture	0.07	0.86
Sing-Song Prosody	7.91	—
Speech Clarity	0.001	0.25
Morphological Complexity	−0.06	−0.98
Utterance Length	−0.002	−0.08

Table 5
Logit Models Predicting Binary Features from Child Age

	Child-directed			Overheard		
	Constant	Age		Constant	Age	
About the 'Here & Now'	38.95	0.88	(0.87, 0.88)	3.72	0.90	(0.88, 0.92)
Child Looking at Referent	11.49	0.84	(0.83, 0.85)	1.19	0.78	(0.64, 0.88)
Child Looking at Speaker	54.47	0.84	(0.83, 0.85)	2.93	0.88	(0.86, 0.91)
Referential Gesture	0.73	0.98	(0.96, 0.99)	0.16	0.97	(0.90, 1.04)
Sing-song Prosody	79.4	1.02	(0.98, 1.05)	0.33	0.94	(0.91, 0.97)

Table 6
Logit Model Predicting Child Attention from Qualitative Dimensions of Speech

	Dependent variable:
	Child Attention {0, 1}
Constant	0.33** (0.17, 0.64)
About the 'Here & Now'	3.83*** (3.14, 4.68)
Sing-song Prosody	5.87*** (4.34, 7.99)
Speech Clarity	1.71*** (1.48, 1.98)
Utterance Length	1.01 (0.99, 1.04)
Morphological Complexity	1.17 (0.84, 1.64)
Age	0.90*** (0.89, 0.91)
Observations	3,605
Log Likelihood	−1,485
Akaike Inf. Crit.	2,985

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
† Composite variable combining codes for child gaze toward the speaker and toward the referent.

References

Bakeman, R., & Adamson, L. B. (2019). Coordinating attention to people and objects in mother-infant and peer-infant interaction. *55*(4), 1278–1289.

Balcomb, F. K., & Gerken, L. A. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*, *11*(5), 750–760. https://doi.org/10.1111/j.1467-7687.2008.00725.x

Baldwin, D. A., Markman, E. M., Bill, B., Desjardins, R. N., Irwin, J. M., & Tidball, G. (1996). Infants' reliance on a social criterion for establishing word-object relations. *Child Development*, *67*(6), 3135–3153.

Booth, A. E., McGregor, K. K., & Rohlfing, K. J. (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. *Language Learning and Development*, *4*(3), 179–202.

Bornstein, M. H., Tal, J., Rahn, C., Galperin, C. Z., Pecheux, M.-G., Lamour, M., Toda, S., Azuma, H., Ogino, M., & Tamis-LeMonda, C. S. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, france, japan, and the united states. *Developmental Psychology*, *28*(4), 593.

Brooks, R., & Meltzoff, A. N. (2008). Infant gaze following and pointing predict accelerated vocabulary growth through two years of age: A longitudinal, growth curve modeling study. *Journal of child language*, *35*(1), 207.

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 of the United States of America*, *110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Casillas, M., Brown, P., & Levinson, S. C. (2019). Early Language Experience in a Tseltal Mayan Village. *Child Development*. https://doi.org/10.1111/cdev.13349

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child development*, *61*(5), 1584–1595.

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed Speech. *Linguistics and Language Compass*. https://doi.org/10.1111/lnc3.12015

de León, L. (1998). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, *8*(2).

Demuth, K., Culbertson, J., & Alter, J. (2006). Word-minimality, epenthesis and coda licensing in the early acquisition of English. *Language and Speech*, *49*(2), 137–174. https://doi.org/10.1177/00238309060490020201

Ellis, R., & Wells, G. (1977). Enabling Factors in Adult-Child Discourse, 46–62.

Emberson, L. L., Lupyan, G., Goldstein, M. H., & Spivey, M. J. (2010). Overheard cell-phone conversations. *Psychological Science*, *21*(10), 1383–1388. https://doi.org/10.1177/0956797610382126

Ervin-Tripp, S. (1978). Some features of early child-adult dialogues. *Language in Society*, *7*(3), 357–373.

Fernald, A., & Kuhl, P. K. (1987). Acoustic determinants of infant preference for motherese speech. *Infant behavior and development*, *10*(3), 279–293.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental psychology*, *20*(1), 104.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., Hamlin, J. K., Hannon, E. E., Kline, M., Levelt, C., et al. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435.

Frank, M. C., Tenenbaum, J. B., & Fernald, A. (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. *Language Learning and Development*, *9*(1), 1–24.

Gerken, L. A., Balcomb, F. K., & Minton, J. L. (2011). Infants avoid 'labouring in vain' by attending more to learnable than unlearnable linguistic patterns. *Developmental Science*, *14*(5), 972–979. https://doi.org/10.1111/j.1467-7687.2011.01046.x

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). *Hard Words* (tech. rep. No. 1).

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344. https://doi.org/10.1177/0963721415595345

Golinkoff, R. M., & Hirsh-Pasek, K. (2006). Baby wordsmith: From associationist to social sophisticate. *Current directions in psychological science*, *15*(1), 30–33.

Graf Estes, K., & Hurley, K. (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, *18*(5), 797–824. https://doi.org/10.1111/infa.12006

Grassmann, S., Schulze, C., & Tomasello, M. (2015). Children's level of word knowledge predicts their exclusion of familiar objects as referents of novel words. *Frontiers in Psychology*, *6*, 1200. https://doi.org/10.3389/fpsyg.2015.01200

Hart, B., & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27*(1), 4–9.

Heath, S. B. (1983). *Ways with words: Language, life and work in communities and classrooms*. cambridge university Press.

Henning, A., Striano, T., & Lieven, E. V. M. (2005). Maternal speech to infants at 1 and 3 months of age. *Infant Behavior and Development*, *28*(4), 519–536. https://doi.org/10.1016/j.infbeh.2005.06.001

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

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418–433. https://doi.org/10.1111/1467-8624.00415

Houston-Price, C., & Nakai, S. (2004). Distinguishing novelty and familiarity effects in infant preference procedures. *Infant and Child Development: An International Journal of Research and Practice*, *13*(4), 341–348.

Hunter, M. A., & Ames, E. W. (1988). A multifactor model of infant preferences for novel and familiar stimuli. *Advances in infancy research*.

Hunter, M. A., Ames, E. W., & Koopman, R. (1983). Effects of stimulus complexity and familiarization time on infant preferences for novel and familiar stimuli. *Developmental Psychology*, *19*(3), 338.

Huttenlocher, J., Vasilyeva, M., Waterfall, H. R., Vevea, J. L., & Hedges, L. V. (2007). The varieties of speech to young children. *Developmental Psychology*, *43*(5), 1062–1083.

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.

Kaplan, P. S., Jung, P. C., Ryther, J. S., & Zarlengo-Strouse, P. (1996). Infant-directed versus adult-directed speech as signals for faces. *Developmental Psychology*, *32*(5), 880.

Leon, L. D. (1998). The Emergent Participant: Interactive Patterns in the Socialization of Tzotzil (Mayan) Infants. *Journal of Linguistic Anthropology*, *8*(2), 131–161. https://doi.org/10.1525/jlin.1998.8.2.131

León, L. D. (1999). Verbs in Tzotzil (Mayan) early syntactic development. *International Journal of Bilingualism*, *3*(2), 219–239.

Lieven, E. V. (1994). Crosslinguistic and crosscultural aspects of language addressed to children.

Ma, W., Golinkoff, R. M., Houston, D. M., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*(3), 185–201.

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. *Corpora in Language Acquisition Research: History, Methods, Perspectives*, *6*, 165–197. http://purl.org/net/MacWhinney-08.pdf

Mehr, S. A., & Krasnow, M. M. (2017). Parent-offspring conflict and the evolution of infant-directed song. *Evolution and Human Behavior*, *38*(5), 674–684.

Mervis, C. B. (1983). Acquisition of a lexicon. *Contemporary Educational Psychology*, *8*(3), 210–236. https://doi.org/10.1016/0361-476X(83)90015-2

Naigles, L. (1990). Children Use Syntax To Learn Verb Meanings. *Journal of Child Language*, *17*(2), 357–374. https://doi.org/10.1017/S0305000900013817

Ochs, E. (1982). Talking to children in Western Samoa. *Language in Society*, *11*, 77–104.

Ochs, E. (1990). Cultural universals in the acquisition of language. *Papers and Reports on Child Language Development*, *29*, 1–19.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*(6), 880–891. https://doi.org/10.1111/desc.12172

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontiers in Psychology*, *8*(JUN), 1008. https://doi.org/10.3389/fpsyg.2017.01008

Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D., & Ripley, M. B. (2013). Package 'mass'. *Cran R.*, *538*.

Rogoff, B. et al. (2003). *The cultural nature of human development*. Oxford university press.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774.

Rowe, M. L., Coker, D., & Pan, B. A. (2004). A comparison of fathers' and mothers' talk to toddlers in low-income families. *Social development*, *13*(2), 278–291.

Rowe, M. L., Denmark, N., Harden, B. J., & Stapleton, L. M. (2016). The role of parent education and parenting knowledge in children's language and literacy skills among white, black, and latino families. *Infant and Child Development*, *25*(2), 198–220.

Rowe, M. L., Leech, K. A., & Cabrera, N. (2017). Going beyond input quantity: Wh-questions matter for toddlers' language and cognitive development. *Cognitive science*, *41*, 162–179.

Saint-Georges, C., Chetouani, M., Cassel, R., Apicella, F., Mahdhaoui, A., Muratori, F., Laznik, M. C., & Cohen, D. (2013). Motherese in Interaction: At the Cross-Road of Emotion and Cognition? (A Systematic Review). *PLoS ONE*, *8*(10), 1–17. https://doi.org/10.1371/journal.pone.0078103

Sanchez, A., Meylan, S., Braginsky, M., MacDonald, K. E., Yurovsky, D., & Frank, M. C. (2018). *childes-db: a flexible and reproducible interface to the Child Language Data Exchange System*. psyarxiv.com/93mwx

Schieffelin, B. (1990). *The give and take of everyday life: Language socialization of Kaluli children*. Cambridge: Cambridge University Press

1990. The give and take of everyday life: language socialization of Kaluli children. Cambridge: Cambridge University Press.

Shatz, M. (1978). On the development of communicative understandings: An early strategy for interpreting and responding to messages. *Cognitive psychology*, *10*(3), 271–301.

Sherrod, K. B., Friedman, S., Crawley, S., Drake, D., & Devieux, J. (1977). Maternal language to prelinguistic infants: Syntactic aspects. *Child Development*, 1662–1665.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672–686. https://doi.org/10.1017/S0305000912000141

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Slobin, D. I. (1985). Crosslinguistic evidence for the language-making capacity. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition: Volume 1: The data*. Psychology Press.

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy*, *13*(4), 410–420. https://doi.org/10.1080/15250000802188719

Snow, C. E., & Ferguson, C. A. (1977). *Talking to children*. Cambridge University Press.

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

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019). Reexamining the Verbal Environments of Children From Different Socioeconomic Backgrounds. *Child Development*, *90*(4), 1303–1318. https://doi.org/10.1111/cdev.13072

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review*, *44*, 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Team, D. (2014). *Datavyu: A Video Coding Tool. Databrary Project, New York University*. Databrary Project. New York University.

Vogt, P., Mastin, J. D., & Schots, D. M. (2015). Communicative intentions of child-directed speech in three different learning environments: Observations from the Netherlands,

and rural and urban Mozambique. *First Language*, *35*(4-5), 341–358. https://doi.org/10.1177/0142723715596647

Ward, M. (1971). *Them children: A study in language*. New York: Holt, Rinehart, Winston.

Weisleder, A., & Fernald, A. (2013). Talking to Children Matters: Early Language Experience Strengthens Processing and Builds Vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145 doi: 10.1177/0956797613488145

Summary of Primary Coded Variables

Variable	N	Mean	SD	Min	Max
Age	6,733	23.00	8.40	30.00	34.00
Tokens	6,73	4.90	4.00	7	55
Morphemes	6,514	6.20	5.00	8	65
Here & Now	6,532	0.57	0.49	0	1
Referential Gesture	2,685	0.31	0.46	0	1
Looking @ Speaker	3,655	0.56	0.50	0	1
Sing-Song Prosody	6,545	0.86	0.34	0	1
Speech Clarity	6,711	2.50	0.76	0	3
Morphological Complexity	6,514	1.20	0.28	1.30	3.00
Child-Directed	6,733	0.84	0.37	1	1
Looking @ Referent	6,733	0.22	0.42	0	1
Play Context	6,733	0.75	0.43	0	1
Child Standing	6,733	0.25	0.43	0	1
Child Held	6,733	0.012	0.11	0	1

Correlations in the Language Environment

Table 7

Correlations Among Qualitative Variables in Speech Directed to Naima

	Age (mos)	Play	Stand-ing	Tokens	Morphs	Morph. Comp.	Here & Now	Ref. Gesture	Sing-song	Look @ Speaker
Age (mos)										
Play	-0.09***									
Standing	-0.16***	0.43***								
No. Tokens	0.14***	0.04	0.01							
No. Morphs	0.12***	0.04	0.00	0.97***						
Morph. Comp.	-0.02	0.05*	-0.03	0.11***	0.30***					
Here & Now	-0.50***	0.10***	0.13***	-0.13***	-0.12***	0.05*				
Gesture	-0.05*	-0.05*	0.04	0.05*	0.05*	0.02	0.24***			
Prosody	0.03	-0.04	-0.05*	-0.01	-0.01	-0.01	-0.02	0.01		
Look@Spkr	-0.57***	0.09***	0.04	-0.10***	-0.08***	0.06**	0.38***	0.04	0.01	
Clarity	-0.54***	0.08***	0.08***	-0.03	-0.02	0.07**	0.40***	0.04	-0.03	0.43***

Table 8

Correlations Among Qualitative Variables in Speech Around Naima

	Age (mos)	Play Context	Child Standing	Tokens	Morphs	Morph. Comp.	Here & Now	Ref. Gesture	Sing-song	Look Speaker
Age (mos)										
Play Context	-0.09***									
Child Standing	-0.16***	0.43***								
No. Tokens	0.14***	0.04	0.01							
No. Morphemes	0.12***	0.04	0.00	0.97***						
Morph. Complexity	-0.02	0.05*	-0.03	0.11***	0.30***					
Here & Now	-0.50***	0.10***	0.13***	-0.13***	-0.12***	0.05*				
Referential Gesture	-0.05*	-0.05*	0.04	0.05*	0.05*	0.02	0.24***			
Sing-song Prosody	0.03	-0.04	-0.05*	-0.01	-0.01	-0.01	-0.02	0.01		
Looking @ Speaker	-0.57***	0.09***	0.04	-0.10***	-0.08***	0.06***	0.38***	0.04	0.01	
Speech Clarity	-0.54***	0.08***	0.08***	-0.03	-0.02	0.07***	0.40***	0.04	-0.03	0.43***

Qualitative Aspects of Overhearing Context Codeable from Video

sccc

Category	Type	Code	*Description*
semantic	−/+	here_now	Is the speech about the 'here and now,' or decon
	0–3	adultness	*(opposite of 'babiness')*
visual	−/+	speaker	Is the child looking at the speaker?
attention	−/+	referent	Is the child looking at speaker is referring
	0–3	clutter	How cluttered is the scene?
referential	−/+	gaze	Is the speaker looking at what they're talkin
	−/+	gesture	Is the speaker gesturing, demonstrating, or p
audience	=	target child	To whom is the utterance directed?
	=	other child(ren)	
	=	adult(s)	
	=	phone	
audio	−/+	sing-song	Is the speaker using exaggerated child-directed
	0–3	auditory clarity	How clear is the utterance?
	0–3	proximity	How near is the speaker?
	−/+	dialogue	Does the child have access to addressee backc
	0–3	noise	*(auditory equivalent of clutter)* How much competition is there for

source	=	live	Where is the speech coming from?
	=	tv	
	=	tablet	
	=	radio	
	=	phone	
child	=	supine	How is the child positioned?
position	=	pront	
	=	crawling	
	=	sitting_low	
	=	sitting_high	
	=	held_hip	
	=	held_front	
	=	held_back	
	=	standing	

Code Types:

[+/−]	Binary Feature
[=]	Variable with Mutually Exclusive Values
[0–3]	Subjective Rating Scale