Early language experience in a Papuan village

Marisa Casillas[1], Penelope Brown[1], & Stephen C. Levinson[1]

[1] Max Planck Institute for Psycholinguistics

Author Note

Correspondence concerning this article should be addressed to Marisa Casillas, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: Marisa.Casillas@mpi.nl

Abstract

To be completed later.

Keywords: Child-directed speech, linguistic input, non-WEIRD, vocal maturity, interaction, Papuan

Word count: XXXXX (XXXX not including references)

Early language experience in a Papuan village

Introduction

In their first few years of life, children hear an extraordinary amount of language. Tracking the distribution and characteristics of this linguistic input over multiple interactional contexts, across developmental time, and between different families is a difficult task. Traditionally, developmental language science has relied on short video recordings of caregiver-child interaction, at home or in the lab, to get a grasp on what kinds of language children typically hear. This approach has been fruitful in teasing out individual and group-based differences in interactional behaviors (REFS). However, over the last decade or so, a new method for tracking child language experience has gained rapid popularity: daylong recordings. Daylong recordings are typically made from a single audio recorder worn by the target child at home, unleashing participants from the limits of a single-camera and allowing them to freely navigate their environment for multiple hours at a time. Unfortunately, however, daylong recordings often require immense resources in order to extract meaningful lingusitic information from the audio signal.

Daylong recordings may therefore appear at first blush to have little value in settings where researchers can instead invest their time in ethnographic microanalysis with selective, short recordings that have high emic validity and considerable semantic depth. In particular, researchers investigating language development outside of their own cultural context may struggle in deciding with which approach is best; identifying "typical" or "representative" behaviors to record and measure requires intensive familiarization with participating families and the community at large, but hasty collection and analysis of daylong data risks mischaracterizing language use and language learning in that community. In the present study we investigate the differing perspectives offered by intensive, close study of short recordings collected during ethnographic study and broad, panoramic recordings of the language landscape using daylong methods. We contrast the use of these two approaches—hereafter the Close-Study approach and the Panoramic approach—on a single language community: Rossel Island, Papua New Guinea.

The Close-Study approach

Short, multimodal recordings (e.g., audio plus video data, motion tracking, or eye movements), give rich insight into the moment-to-moment characteristics of interaction. The increased context provided by multi-modal recordings helps discern the meaning of each communicative behavior documented, including young children's sometimes-jumbled linguistic productions (REFS). Such recordings can be made in nearly any context, take little time to collect, and are consistent with other, more costly and time-consuming recording styles (REFS). When richly annotated and paired with intensive ethnographic study, these recordings become potent samples of language development in the studied community that can be used again and again for a wide variety of meaningul analyses.

Ethnographic work is essential for appropriately situating recording collection, chosen behaviors for analysis, and data interpretation within the realm of normal and relevant behaviors for the studied community. In practice, this approach means that decisions on what to study and precisely how to study it are informed by knowledge of daily tasks, typical household relations and responsibilities, attitudes about child rearing, considerations about when children qualify as co-interactants, and what behaviors are expected of children and caregivers in the first years of life. In a situation where the researcher is a member of the community under study (e.g., middle-class US researchers investigating language development in middle-class US families), assumptions about what to study and how are implicitly enriched by this knowledge. However, when the researcher is a visitor to the community, selecting the right measures and finding ways to compare them to child development outcomes in other sites is an serious challenge.

The drawbacks of the Close-Study approach are few but significant. First, the time and financial investment needed to gain familiarity with a community and to add detailed, comprehensive annotation and transcription to the gathered recordings limit the feasible sample size of most studies; language development in a handful of focal children may provide many insights, but may take decades of dedicated work to explore in depth. Second, while researchers using this method can dilligently track a variety interactional contexts, the anchoring effect of a single video camera or audio recorder on the child (and caregivers) makes it difficult to capture daily activities that involve a lot of free motion (e.g., talking while running with other children in a field) or access to parts of the household or activities that are typically not accessible to others, even researchers on close terms with the recorded family (e.g., pre-sleep routines in bedding areas). There may be meaningful and frequent sources of linguistic information during these hard-to-capture activities. Finally, unless a microphone is worn by the child, whispered speech, speech to self, and other quiet but hearable events are difficult to capture from a third-person recording perspective.

The Panoramic approach

Improved recording hardware and advances in speech technology in the last 20 years have allowed us to peek into children's broader language landscapes. These recordings give a bird's eye view into the ebb and flow of everyday language activity, inclusive of both animated chatter while running with siblings and quiet self-directed talk when sitting alone. This broadened view is uniquely suited to estimating the total linguistic input children encounter, and the typical axes on which this input rate varies (e.g., specific speakers, times of day, etc.). Accurate measures of linguistic input are critical for investigating how much experience is needed to acquire a given linguistic or communicative phenomenon (REFS). Starting up daylong recordings is quick and straightforward—the main hurdle is getting the child to wear the vest/shirt in which the recorder is placed—and researchers have had success implementing these recordings in a wide range of cultural contexts (REFS; but see Defina REFS). The most popular daylong recording system is the LENA, which comes with a recording device that captures up to 16 hours of audio at a time and comes with software for automatically analyzing basic properties of the speech signal (REFS). The LENA system is expensive, but is not the only route to daylong data; several groups have succesfully experimented with daylong recordings using other devices (e.g., Olympus, Zoom, USB recorder) paired with manual or open-source automated annotation (REFS). Once an efficient pipeline for annotation is established, daylong recordings can also be used to collect comparable recordings from large, representative samples of a given language community (REFS).

The Panoramic approach has several significant drawbacks (REFS), particularly for research questions that involve linguistic analysis. Here we focus on those drawbacks that prevail even when we assume that the researcher has some resources to add manual or automated linguistic annotation. First, the resulting recording collections are typically too large for comprehensive transcription or annotation, with no easy way to scan for the specific phenomena of interest. Researchers must therefore employ some strategic sub-sampling technique in order to annotate the data, even though best practices for doing so are not yet well established (REFS). Second, even once clips are sampled from the daylong recording, adding relevant annotations to them can take nearly as long as a Close-Study approach, but with reduced likelihood of capturing interesting or relevant caregiving and language use behaviors. Third, a whole day of recording is a lot of data, but may not be enough to achieve a stable estimate of average linguistic input (REFS Anderson & Fausey, in prep). A fourth drawback is that properly collecting, processing, and archiving daylong data is not easily achieved; the fact that participants are likely to habituate to the recorder is fantastic for documenting ecologically valid language use, but raises urgent questions about participant privacy standards (REFS Ethics). Fourth, at time of writing, there are few options for capturing visual information across the day (but see REFS), limiting this method primarily to acoustic phenomena (REFS). Even if researchers add manual annotation to these audio files, they typically do so without the benefit of the visual context; a difficulty compounded by the diversity of activities and interlocutors captured over the recording.

Differing perspectives on the child language environment

Which approach should one choose when describing children's language environments? The Close Study approach takes the general stance that richer data is better data, with the primary problem being that the researcher can't know how well their zoomed-in perspective generalizes to the rest of the population. The Panoramic approach takes the general stance that more data is better data, with the primary problem being that the researcher can't know if they are measuring the right phenomena, particularly when importing pre-conceived notions about learning into culturally unfamiliar contexts. The ideal solution, of course, is to thoroughly annotate and analyze large, representative samples of data, but doing so would require many years of well-funded multi-researcher commitment—a risky prospect for a basic descriptive question.

One alternative approach is to add complementary data to a community where one approach has already been taken. For example, extensive ethnographic research among multiple indigenous Mayan communities of Southern Mexico and Guatemala has forged a consistent view of childrearing and child-directed speech: adult caregivers shape infants' and young children's worlds such that the children learn to attend to what is going on around them rather than expecting to be the center of attention (REFS). These findings lay out a broader ideology of caregiving, including component attitudes (e.g., orientation toward infants as *not* conversational partners) that can, in turn, be used to make predictions about Mayan children's linguistic input. Importantly, however, it is not clear how these attitudes play out on the scale of day-long averages; preferences for when and how to talk to children are balanced by the many other demands of everyday life. On this view, we may feel certain that the Panoramic view indeed captures the transmission of critical linguistic and cultural knowledge, but we can't point to where it happens. That said, a handful of findings up until now suggest a promising, though imperfect link between the attitudes and ideologies described in Close Study work and the broader behavioral patterns from Panoramic work in those same communities.

In the case of Mayan child language environments, findings using a larger-sample or Panoramic-type approach have been fairly consistent with the caregiving practices described in previous Close Study work. Shneidman (REFS) used short videos of interaction to conduct a quantitative, longitudinal study of the Yucatec children's typical speech experiences. She indeed found that infants were rarely spoken to, but that the prevalence of speech directed to children increased enormously with age, mostly due to an influx of speech from other children (REFS). That said, the input rate from adults predicted children's later vocabulary size more than their total input rate (REFS). Casillas and colleagues (REFS) used daylong recordings with children in a Tseltal Mayan community, again finding that infants and young children were spoken to rarely. However, they found no increase in speech input with age, and the majority of speech came from adult women, even when children were old enough to independently follow their older siblings and cousins around. The studies collectively suggest that, consistent with Close Study work in these and similar communities, (female) adult speech is relatively rare, but is a prominent and predictive source of linguistic input in Mayan children's language development.

Studies in a North American context, in which North American researchers can more reliably depend on their own intuitions about language learning, have tried to pinpoint the differences in close-up and zoomed-out views of the child language environment: short recordings display much denser input, with some changes in the types of language used, compared to longer recordings (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). For example, Bergelson and colleagues (Bergelson et al., 2019) analyzed the noun use encountered by 44 6- and 7-month-old children in the US in both hour-long at-home videos and comparable sub-samples of daylong recordings. The video and daylong data were markedly different in linguistic input rate; nouns were used 2–4 times more often in the videos. The authors also found some differences in input type: nouns were more likely to come embedded in questions in the videos, but the daylong data featured more noun types and noun input from more speakers (see (Bergelson et al., 2019) for the full range of differences). Other than these differences, the overall profile of input *type* was quite similar between the video data and the daylong recording sub-samples (e.g., similar relative use of speech act types). Other work using varying durations of video (i.e., short-structured vs. longer-unstructured video) with US child-caregiver pairs also found lower estimates for the rate of linguistic input in longer recordings, but found that children's relative rank was stable across recording contexts (Tamis-LeMonda et al., 2017).

Based on these findings from both the Mayan and US context, one might infer that the language use captured by Panoramic recordings is driven, at least in part, by the same factors driving language patterns highlighted in Close Study work. However, these preliminary results also hint at divergences between what caregivers do under ideal or performative circumstances and what they do when juggling childcare with the diverse activities and interlocutors encountered during everyday life. In trying to understand how children's language environments impact their language learning, researchers seek meaningful variation in children's linguistic experience; it may be that, with panoramic data, much of the variation children encounter has less to do with their caregivers' ideological stance toward talking to young children and more to do with who else is around and what other tasks are at hand.

Whether this circumstantial variation has greater or equal predictive validity to variation in caregiver ideology across a range of linguistic skills is an open question in need of further research. For example, it is difficult at present to determine the extent to which Mayan children hear less directed input because of the childrearing practices traditional to these communities or because of other features of their lifestyle (e.g., subsistence farming effects on who is present, number of other children present, etc.; see also Shneidman REFS). Our comparison group, US families, differ greatly from these Mayan communities in the circumstances of everyday life (e.g., work patterns, number of co-residents, child sleeping routines; REFS). Disentangling the sources of differences in the quantity of linguistic input children experience issue requires us to collect Close Study and Panoramic findings in a third community; one with a (broadly) similar lifestyle to that of the Mayans, but with different ideas about how to talk to young children.

The current study

In this study we present analyses of daylong recordings from a small-scale indigenous group, on Rossel Island, Papua New Guinea (PNG), in which prior ethnographic work has painted a clear picture of early caregiver-child interaction: child-centric, face-to-face interaction from the first days of infancy. Based on the prior ethnographic work, detailed below, we made four predictions about children's speech environments. First, we predicted that children on Rossel Island would hear frequent child-directed speech from a wide variety of caregiver types throughout the day. Second, given that infants are frequently passed between caregivers, we expected to see weaker effects of the subsistence farming schedule on Rossel children's input than has been found in other societies (Casillas et al., forthcoming). Third, as children get older, we expected to see a large increase in the proportion of child-directed speech coming from other children (see also Shneidman REFS). Fourth, we expected a large quantity of other-directed speech around them, given the large number of family numbers typically present. Based on prior work with daylong recordings with both Western and non-Western small-scale populations, we additionally expected (a) no age-related increase in child-directed speech (Scaff, Casillas, Bergelson, REFS), (b) an age-related decrease in other-directed speech (Casillas, Bergelson, REFS), and (c) that children's input would be non-uniformly distributed over the day (Abney, Smith, & Yu, 2017; Blasi, Schikowski, Moran, Pfeiler, & Stoll, in preparation) such that interactional peaks present a much denser view of their input (Casillas et al., forthcoming), similar to that observed in more Close-Study approaches.

In what follows we will review the ethnographic work done with this community previously, describe our methods for following up on these findings with daylong recordings, present the current findings, and discuss the differences that arose. This study was completed as part of a larger comparative project focusing on children's speech environments and linguistic development at two sites: the Tseltal Mayan community mentioned above (Casillas et al., forthcoming) and this Rossel Island community. Therefore all methods for annotation and analysis in this study parallel those reported elsewhere for Tseltal Mayan children's speech environments (Casillas, Brown, & Levinson, forthcoming).

Method

Corpus

The participants in this study live in a collection of small hamlets on north-eastern Rossel Island, approximately 250 nautical miles off the southern tip of mainland Papua New Guinea. The traditional language of Rossel Island is Yélî Dnye, a presumed Papuan isolate, which features a phonological inventory and set of grammatical features that are unlike any other in the (predominantly Austronesian) languages of the region. Rosselers are skilled farmers, cultivating taro, sweet potato, manioc, yam, coconut, and more for their daily subsistence, with protein coming from fishing and (occasionally) slaughtering pigs or local animals. Most children on Rossel Island grow up speaking Yélî Dnye monolingually at home, beginning to learn English as a second language once they begin school around age 7 or 8. Children grow up in patrilocal household clusters (i.e., their family and their father's brothers families), usually arranged such that there is some shared open space between households.

During their waking hours, infants are typically carried in a caregiver's arms as they go about daily activities. Infants, even very young ones, are frequently passed between different family members (male and female, young and elderly) throughout the day, returning to the mother to suckle when hungry. The arc of a typical day for an infant might include waking, being dressed and fed, then a mix of (a) spending time with nearby adults or older children as they walk around socializing and completing tasks with others and (b) more feeding, perhaps followed by short bouts of sleep in the late morning and afternoon, usually with the mother. Afternoon meals are cooked from around 15:00 onward, with another meal time and more socializing at home before resting for the night. Starting around age two or three, children also begin to spend a lot of their time in large, independent child playgroups involving up to 10 or more cousins at a time who freely travel near and around the village searching for nuts and fruits, bathing in nearby rivers, and engaging in group games (e.g., tag, pretend play, etc.).

Interaction with infants and young children on Rossel Island is initiated by women, men, girls, and boys alike in a face-to-face, contingency-seeking, and affect-laden style (Brown REFS). Children are considered a shared responsibility, but also a source of joy and entertainment for the wider network of caregivers in their community. In her prior ethnographic work, Brown details some ways in which interactants make bids for joint attention and act as if the infant can understand what is being said (REFS). Infants pick up on this pattern of caregiving, intiating interactions with others twice as frequently as Tseltal children, who are encouraged instead to be observers of the interactions going on around them (Brown 2011 REFS). At the same time, Brown (REFS) documents how Rossel caregivers encourage early independence in their children, observing their autonomy in choosing what to do, wear, eat, and say while finding other ways to promote pro-social behavior (e.g., praise; REFS). Overall, Rossel Island could be characterized as a child-centered language environment (Ochs & Schieffelin 1984; REFS but see Brown & Casillas REFS), in which children, even very young ones, are considered interactional and conversational partners whose interests are allowed to shape the topic and direction of conversation.

We were interested in using Panoramic methods to investigate the language environment of children acquiring Yélî Dnye because prior ethnographic work had suggested that child-directed speech is highly frequent in this community, from mothers and other adult caregivers, but also from other children. The typologically rare and complex structures characteristic of Yélî Dnye phonology and morphosyntax makes Rossel Island a theoretically intriguing place for extended language development study. For present purposes, and intial study of how much linguistic input children encounter, and how it varies within the home context, allows a first glimpse into the material available for children to learn from. At the same time, this work further our understanding of the different perspectives offered by the Close-Study and Panoramic approaches cross-culturally.

The data presented here come from Rossel Island subset of the Casillas HomeBank Corpus (Casillas, Brown, & Levinson, 2017), a collection of raw daylong recordings and supplementary data from over 100 children under age four growing up on Rossel Island and in the Tseltal Mayan community described elsewhere (Casillas et al. forthcoming). The Rossel Island subcorpus was collected in 2016 and includes daylong audio recordings and experimental data from 57 children born to XX mothers. On average, the target children in these recordings had X–X younger siblings (mean = X; median = X) and X–X older siblings (mean = X; median = X); most participating parents were on the younger end of parents in the community (mothers: mean = XX years; median = XX; range = XX–XX and fathers: mean = XX; median = XX; range = XX—XX). Based on our demographic data we estimate that mothers are typically XX years old when they give birth to their first child (median = XX; range = XX–XX) with an average inter-child interval of X years (median = X; range = X–X). Notably, however, we received several reports, including from nursing staff at the local health clinic, that mothers now are having children younger and closer together than in generations past. Household size, defined here as the number of people sharing kitchen and sleeping areas on a daily basis, ranged between X and XX (mean = X; meadian = X). Households are clustered into small hamlets which form a wider group of communal caregivers and playmates. The hamlets themselves are clustered together into broader patches of patrilocal residents. The average hamlet in our corpus comprises X households (median = X; range = X–X); assuming an average of X children under age seven (i.e., not schooling) and X adults per household, we estimate that there are between XX and XX children and between XX and XX adults present throughout the day, not including visitors, visits to neighboring hamlets or other nearby resident areas. Therefore, while XX% of the target children in our corpus are first born to their mothers, they are immediately incorporated into a much larger pool of young children whose care is divided among numerous caregivers. Among our participating families, most mothers had finished primary school (XX%; X years of education) or secondary school (XX%; X years of education), with a few having completed preparatory school (XX%; X years of education) or beyond (XX%; X years of education). Only XX% of mothers had less than a primary school education. Similarly, most fathers had finished primary school (XX%; X years of education) or secondary school (XX%; X years of education), with a few having completed preparatory school (XX%; X years of education) or beyond (XX%; X years of education), with only XX% having less than a primary school education. To our knowledge at the time of recording, all but two children were typically developing; one showed signs of significant language delay and one showed signs of multiple developmental delay (motor, language, intellectual), both children's delays were consistently observed in follow-up trips in 2018 and 2019.

Dates of birth for children were initially collected via parent report. We were able to verify the vast majority of birth dates using the records at the island health clinic. Because not all mothers give birth at the clinic and because dates are written by hand, some births are not recorded, are inaccurately recorded, or otherwise significantly diverge from what the parents report. In these cases we gathered information from as many sources as possible and followed up with the families, often using the dates of neighboring children born around the same time to home in on the correct date.

The data we present come from 7–9-hour recordings of a waking day at home for the child. Children wore the recording device, which was an elastic vest containing a small stereo audio recorder (Olympus WS-832 or WS-853) and a miniature camera that captured photos of the child's frontal view at a fixed interval (every 15 seconds; Narrative Clip 1). The camera was outfitted with a fisheye lens that, while distorting the images, allowed us to capture 180 degrees of children's frontal view. This technique allows us to use daylong recordings while also partially getting around the traditional Panoramic-method sacrifice of no visual context, thereby increasing ease and reliability of our transcrition and annotation. However, because the camera and recorder are separate devices, we had to synchronize them manually after the recordings were made. To do this, we used an external wristwatch to record the current time at start of recording on each device individually, with accuracy down to the second (photographed by the camera and spoken into the recorder). The camera timestamps each photo such that we can calculate the number of seconds that have elapsed between each one. These timestamps can be used with the cross-device time synchronization cue to create photo-linked audio files of each recording, which we then format as video files (see https://github.com/marisacasillas/Weave for post-processing scripts and more information). The informed consent process used with participants, as well as data collection and storage, were conducted in accordance with ethical guidelines approved by the Radboud University Social Sciences Ethics Committee.

Data selection and annotation

From the daylong recordings of XX Rossel children, we selected 10 representative children between ages 0;0 and 3;0 for transcription and analysis in the current study. The 10 children were selected to be spread between the target age range (0;0–3;0) while also representing a range of typical maternal education levels found in the community and being evenly split between male and female children (see also ACLEW REFS). For each child we then selected a series of non-overalapping sub-clips from the day for transription in the following order: nine randomly-selected 2.5-minute clips, five manually-selected "peak" turn-taking activity 1-minute clips, five manually-selected "peak" vocal activity 1-minute clips, and one manually-selected 5-minute expansion of the best one-minute clip, for a total of 37.5 minutes of transcribed audio for each child (6.25 audio hours in total). The criteria for manual clip selection are identical to those described for the parallel study on Tseltal by Casillas and colleagues (forthcoming).

We were limited to selecting sub-clips from 10 children for analysis because of the time-intensive nature of transcribing these naturalistic data; 1 minute of audio typically took approximately 60–70 minutes to be segmented into utterances, transcribed, annotated, and loosely translated into English (~400 hours total). Given that Yélî Dnye is nearly exclusively spoken on Rossel Island, where there is no electricity and unreliable access to mobile data, transcription could only be completed over the course of three 4–6 week visits by our research group to the island in 2016, 2018, and 2019.

We used the ACLEW Annotation Scheme (REFS) in ELAN (ELAN REFS) to transcribe and annotate all hearable speech—both near and distant—in the clips. We first segmented out the utterances and ascribed them to individual speakers (e.g., older brother, mother, aunt, etc.). We then annotated the vocal maturity of each utterance produced by the target child (non-canonical babble/canonical babble/single word/multi-word/unsure) and annotated the addressee of all speech from other speakers (addressed to the target child/one or more other children/one or more adults/a mix of adults and children/any animal/other/unsure). Transcription and annotation was done together by the first author and one of three community members (all native speakers of Yélî Dnye). The community-based research assistants personally knew all the families in the recordings, and were able to use their own experience, the discourse context, and information from the accompanying photos in reporting what was said and to whom speech was addressed for each utterance. Detailed manuals and self-guided training materials, including a "gold standard test" for this annotation scheme can be found at URL (REFS).

In what follows we first analyze the nine randomly selected 2.5-minute clips from each child to establish a baseline view of their speech environment, focusing on the effects of child age, time of day, household size, and number of speakers on the rate of target child-directed (TCDS) and other-directed speech (ODS) present. Next, we repeat these analyses, focusing instead only on the turn-taking clips to gain a view of the speech environment as it appears during the peak interactions for the day. This latter set of analyses may more closely mirror results from prior ethnographic work, which was designed to focus on typical, lively interactions with young children. Then as a first approximation of children's linguistic development, we map a coarse trajectory of children's use of babble, first words, and multi-word utterances. Finally, we wrap up by integrating our Panoramic-approach results with those from prior Close-Study work, relating these findings to the larger literature on child-directed speech and its role in language development.

Statistical models

We conducted all analyses in R, using the glmmTMB package to run generalized linear mixed-effects regressions on our dependent measures (M. E. Brooks et al., 2017; R Core Team, 2018). We used ggplot2 to generate all plots (Wickham, 2009). The dataset and scripts used in this study can be found at https://github.com/marisacasillas/Yeli-CLE. As in previous work on child speech environment measures (REFS Casillas et al. forthcoming, Bunce et al, in prep), TCDS and ODS minutes per hour are naturally restricted to non-negative (0–infinity) values, causing the distributional variance of those measures to become positively skewed. To address this issue we use negative binomial regressions, which can better fit non-negative, overdispersed data (M. E. Brooks et al., 2017; Smithson & Merkle, 2013). There were also many cases of zero minutes of TCDS across the clips—for example, this often occurred in the randomly sampled clips when the child was sleeping in a quiet area. To handle this additional distributional characteristic of the data, we added a zero-inflation model to TCDS analysis which, in addition to the count model of TCDS (e.g., testing effects of age on the input rate), creates a a binary model to evaluate the likelihood of TCDS being used at all. More conventional, gaussian linear mixed-effects regressions with logged dependent variables are available in the Supplementary Materials. The results of those alternative models are qualitatively similar to what we report here.

Results

The models included the following predictors: child age (months; centered and standardized), household size (number of people; centered and standardized), number of non-target-child speakers present in that clip (centered and standardized), and time of day at the start of the clip (factor: "morning" = before 11:00; "midday" = 11:00–13:00; "afternoon" = after 13:00). In addition, we included two-way interactions: (a) child age and the number of speakers present and (b) child age and time of day. We also added a random effect of child. For the zero-inflation model of TCDS, we included the number of speakers present. We limit our discussion here to significant effects in the models; full model results, including gaussian alternative models, are available in the Supplementary Materials.

```
## pdf
```

Table 1

Demographic overview of the 10 children whose recordings are sampled in the current study, including from left to right: child's age (years;months.days); child's sex (M/F); mother's age (years); level of maternal education (none/primary/secondary/preparatory/university); and the number of people living in the child's household.

Age	Sex	Mother's age	Level of maternal education	People in household
00;01.09	F	31	secondary	8
00;03.19	M	37	primary	9
00;04.13	M	24	preparatory	5
00;07.18	M	24	secondary	5
00;09.03	F	29	secondary	5
01;00.29	F	30	primary	9
01;05.02	M	25	secondary	6
01;08.03	F	33	primary	9
02;01.22	F	21	secondary	4
02;11.29	M	41	primary	8

```
##   2

## pdf
##   2

## pdf
##   2

## pdf
##   2
```

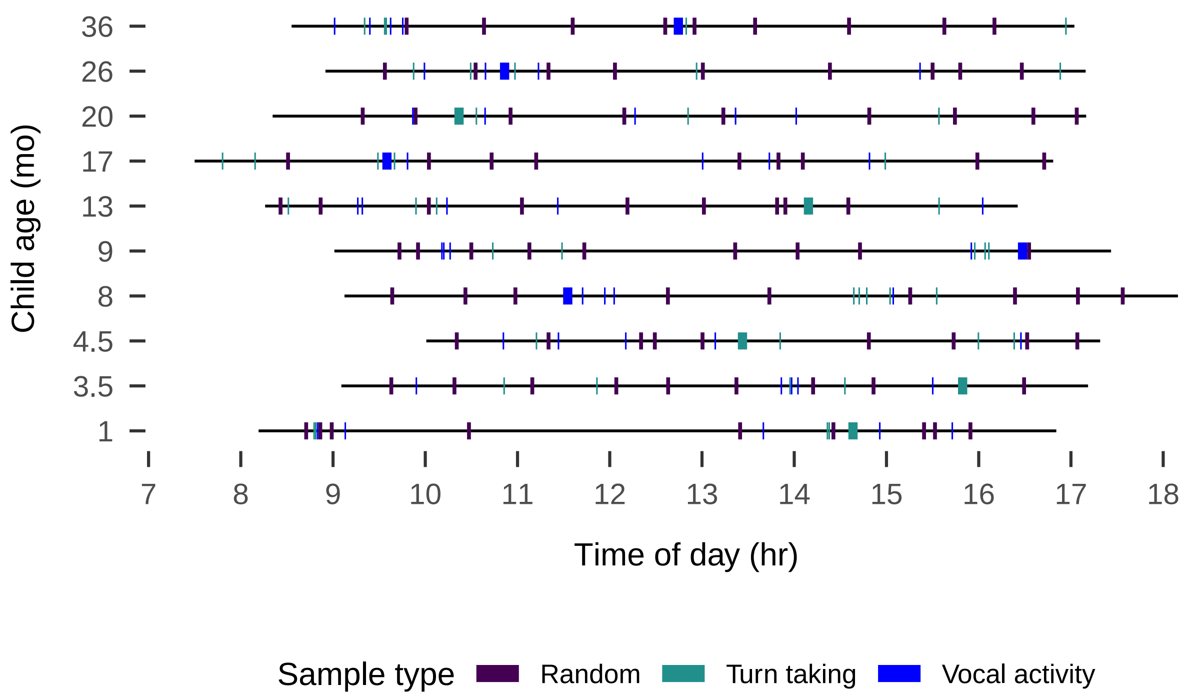

Figure 1. Recording duration (black line) and sampled clips (colored boxes) for each of the 10 recordings analyzed, sorted by child age in months.

Target-child-directed speech (TCDS)

In the random sample, these 10 children heard an average of 3.13 minutes of speech directly addressed to them per hour (median = 2.95; range = 1.58–6.26; Figure 3). For comparison, this is slightly less than reported values using a near-identical method of data collection, annotation, and analysis in a Tseltal Mayan community (3.6 minutes per hour for children under 3;0; Casillas et al., fortchoming) and comparable to what has been reported using a similar method in a Tsimane community (4.8 minutes per hour for children under 3;0 including all hearable speech; 1.6 minutes when excluding overlap and far-away speech; Scaff et al., in prep).

The zero-inflated negative binomial regression of TCDS minutes per hour (N = 90, log-likelihood = -195.26, overdispersion estimate = 3.37) suggested significant effects of child age, time of day, and their interaction on the rate at which children hear speech addressed

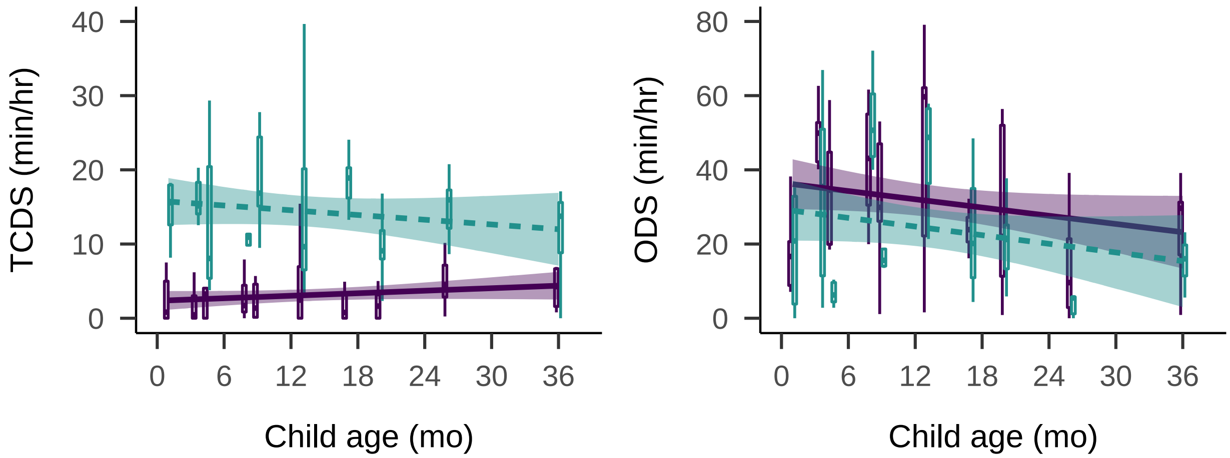

Figure 2. Estimates of TCDS min/hr (left) and ODS min/hr (right) across the sampled age range. Each box plot summarizes the data for one child from the randomly sampled clips (purple; solid) or the turn taking clips (green; dashed). Bands on the linear trends show 95% confidence intervals.

directly to them. First, the older children heard significantly more TCDS per hour (B = 0.73, SD = 0.23, z = 3.20, $p < 0.01$), with an average increase of 0.73 minutes per hour for every month of development. Overall, these children were also more likely to hear TCDS in the mornings (see Figure 4 for an overview of time-of-day findings), with significantly higher TCDS rates in the morning compared to both midday (B = 0.80, SD = 0.36, z = 2.23, p = 0.03) and the afternoon (B = 0.54, SD = 0.26, z = 2.10, p = 0.04), and no significant difference in TCDS rate between midday and the afternoon. However, the time-of-day pattern changed with child age. Older children were more likely than younger children to show a peak in TCDS during midday, with a decrease in TCDS between midday and the afternoon (B = -0.60, SD = 0.29, z = -2.04, p = 0.04) and marginally less TCDS in the morning than at midday (B = -0.59, SD = 0.30, z = -1.94, p = 0.05). There were no other significant effects in either the count or the zero-inflation model.

Children heard TCDS from a variety of different speakers. Overall, most TCDS came from adults (mean = 72.65%, median = 75.51%, range = 41.41–100%). On average, 77.23%

of the adult TCDS per clip came from women, and an average of 82.35% of the total adult TCDS minutes came from women. That said, an increasing quantity of TCDS came from other children (child-TCDS; "C-TCDS"); a Spearman's correlation showed a significant positive relationship between the average proportion of child TCDS in a clip and target child age (Spearman's $rho = 0.78$; $p = 0.01$).

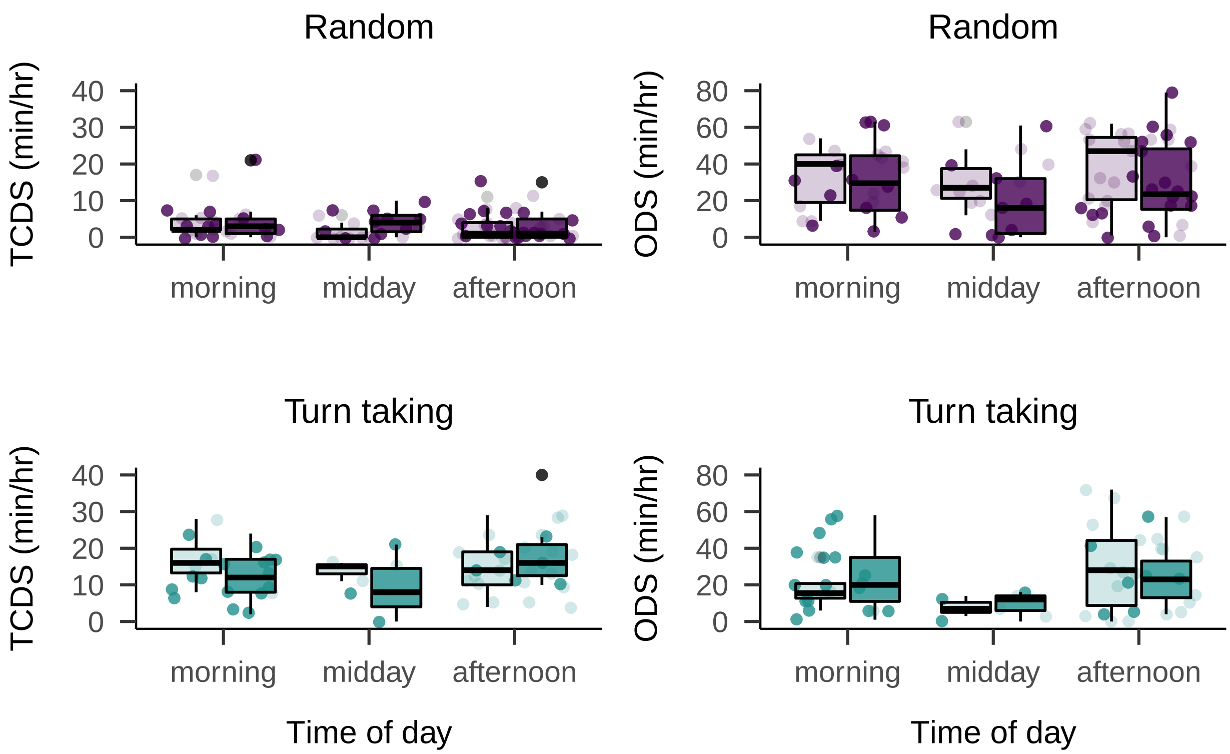

Figure 3. Estimates of TCDS min/hr (left panels) and ODS min/hr (right panels) across the recorded day in the random clips (top panels) and turn-taking (bottom panels) clips. Each box plot summarizes the data for children age 1;0 and younger (light) or age 1;0 and older (dark) at the given time of day.

Other-directed speech (ODS)

In the random sample, these children heard an average of 35.90 minutes of other-directed speech per hour (median = 32.37; range = 20.20–53.78): that is more than eleven times the average quantity of speech directed to them, with some children experiencing near-continuous background speech. For comparison, a prior estimate for Tseltal Mayan children using parallel methods to the present study found an average of 21 minutes of overhearable speech per hour (Casillas et al., forthcoming), and a recent study of North American children's daylong recordings found that adult-directed speech occurred at a rate of 7.3 minutes per hour (Bergelson et al., 2019).

The negative binomial regression of other-directed speech rate ($N = 90$, log-likelihood $= -370.87$, overdispersion estimate $= 9.14$) revealed effects of child age, number of speakers present, and time of day on the rate of ODS encountered. The rate of ODS significantly decreased with child age ($B = -0.57$, $SD = 0.17$, $z = -3.28$, $p < 0.01$) and significantly increased in the presence of more speakers ($B = 0.50$, $SD = 0.05$, $z = 10.07$, $p < 0.001$). Across clips, there were an average of 6.19 speakers present other than the target child in the randomly selected clips (median = 6; range = 1–19), an average of 59.99% of whom were adults. Comparing again to Tseltal and North American English, in which the average number of speakers present was 3.44 and XX respectively (REFS)(**???**), we can infer that the increased rate of ODS on Rossel Island is due in part to there simply being more speakers present. Time-of-day effects on ODS only came through in an interaction with child age. In particular, older children heard a pattern of ODS mirroring the general pattern of TCDS; significantly more ODS in the mornings compared to midday (midday-vs-morning: $B = 0.65$, $SD = 0.20$, $z = 3.23$, $p < 0.01$) and the afternoon (afternoon-vs-morning: $B = 0.37$, $SD = 0.15$, $z = 2.50$, $p = 0.01$). There were no other significant effects on ODS rate in the model.

In sum, the random baseline rates of TCDS and ODS in children's speech environments are influenced by child age (TCDS increases, ODS decreases), time of day (both generally peak in the morning), and their interaction (older children hear more TCDS and less ODS at midday). The rate of ODS is also impacted by the large number of speakers present in some clips. Correlational results suggest that children's TCDS comes increasingly from other children over the first three years. That said, the baseline rate of TCDS is low, on par with estimates in other small-scale farming communities (Casillas et al., forthcoming; Scaff et al., in prep); while the ODS rate is quite high relative to estimates in prior work.

TCDS and ODS during interactional peaks

If we instead investigate the rates of TCDS and ODS encountered by these children during their interactional peaks for the day, a different picture emerges (Figures 3 and 4 green/dashed summaries). In particular, the children heard much more TCDS in the turn-taking clips—14.45 min/hr; that is, more than four times the rate of TCDS in the random baseline (median = 15.07; range = 9.61–18.73). During these same clips, children heard a reduced rate of ODS: 25.27 min/hr (70.39% of the random-sample ODS rate; median = 19.59; range = 6.68–60.18). The negative binomial mixed-effects regression of TCDS (N = 55, log-likelihood = -183.25, overdispersion estimate = 2.91) revealed a significant decrease with child age (B = -0.63, SD = 0.27, z = -2.33, p = 0.02) and a significant interaction between child age and time of day; TCDS rate for older children during interactional peaks was marginally higher at morning compared to midday (midday-vs-morning: B = 0.53, SD = 0.28, z = 1.89, p = 0.06) and significantly higher in the afternoon than at midday (midday-vs-afternoon: B = 0.61, SD = 0.28, z = 2.17, p = 0.03).

As in the random sample, an increasing portion of TCDS during interactional peaks came from other children. While, overall, *more* of the TCDS in interactional peaks came from adults than in the random clips (mean = 82.68%, median = 88.04%, range = 50–100%), a Spearman's correlation showed an even stronger positive relationship between the average proportion of child TCDS in a clip and target child age (Spearman's $rho = 0.92$; $p = < 0.001$). Notably, women contributed proportionally less TCDS during interactional peaks than they did during the random clips: on average, women contributed 61.55% of the children's adult TCDS minutes in the turn-taking clips (compared to 82.35% in the random clips). In brief, interactional peaks include more directed speech from men and more directed speech from other children, with age.

The negative binomial mixed-effects regression of ODS (N = 55, log-likelihood = -202.60, overdispersion estimate = 4.66) only revealed a significant effect of number of speakers. As before, ODS rates were higher when more speakers were present (B = 0.56, SD = 0.08, z = 6.76, $p < 0.001$). There were no other significant effects on ODS rate in the turn-taking clips.

Overall, the results suggest that these children typically hear very little directly addressed speech, but that interactional peaks provide opportunities for dense input at multiple points during the day. While the majority of directed speech comes from women, an increasing portion of it comes from other children with age, and directed speech from men is more likely during interactional peaks. Directed and overhearable speech is most likely during the morning, before most of the household has dispersed for their work activities, similar to other work on subsistence farming households (Casillas et al., forthcoming). However, older children are more likely than younger children to show higher input rates at midday, perhaps due to their increased interactions with other children while adults attend to gardening and domestic tasks; we leave investigation of this idea to future work. Possibly because of the large number of speakers typically present, these children also experienced a high rate of overhearable speech, underscoring the availability of other-addressed speech as a resource for linguistic input in this context.

Vocal maturity

```
## pdf
##   2
```

```
## pdf
##   2
```

```
## pdf
##   2
```

Given the low overall rate of directed speech in these children's environments, we might expect that their early linguistic development, particularly the onset and use of single- and multi-word utterances, is delayed in comparison to children growing up in more CDS-rich environments. To briefly investigate this we plotted the proportion of all linguistic vocalizations (i.e., not laughter, crying, or unknown-type vocalizations; N = 4308 vocalizations) for each child that were (a) non-canonical babble, (b) canonical babble, (c) single-word utterances, or (d) multi-word utterances. With development, children are expected to traverse all four types of vocalization, primarily producing single- and multi-word utterances by age three.

In the onset of use for canonical babble, first words, and multi-word utterances, these Rossel children's vocalization data closely resemble expectations based on populations of children who hear more CDS Figure 5. That is, canonical babble appears in the second half of the first year, first words appear around the first birthday, and multi-word utterances appear a few months after that (Frank, Braginsky, Marchman, & Yurovsky, in preparation; Kuhl, 2004; Pine & Lieven, 1993; Slobin, 1970; Tomasello & Brooks, 1999; Warlaumont, Richards, Gilkerson, & Oller, 2014). Notably, these children also far exceed the usage rate of speech-like vocalizations associated with major developmental delay; while the canonical babbling ratio (CBR; proportion speech/speech-like vocalizations out of all vocalizations) associated with developmental delay is below 0.15 at age 0;10, all the Rossel children above 0;9 exceed this threshold, with a minimum CBR of 0.22 at age 0;9 (mean = 0.63; median = 0.68; range = 0.22– 0.86; see also REFS Cychosz et al.).

Over all annotated clips, children produced an average of 7.18 linguistic vocalizations per minute (median = 7.79; range = 4.57–8.95), which is less than might be expected in American infant-caregiver recordings (D. K. Oller, Eilers, Basinger, Steffens, & Urbano, 1995). However, this rate does align well with the frequency of child-initiated prompts estimated for Rossel interaction (Brown, 2011). The rate also matches estimates for Tseltal Mayan children, who hear a similar quantity of directed speech during this age range (REFS Casillas et al.).

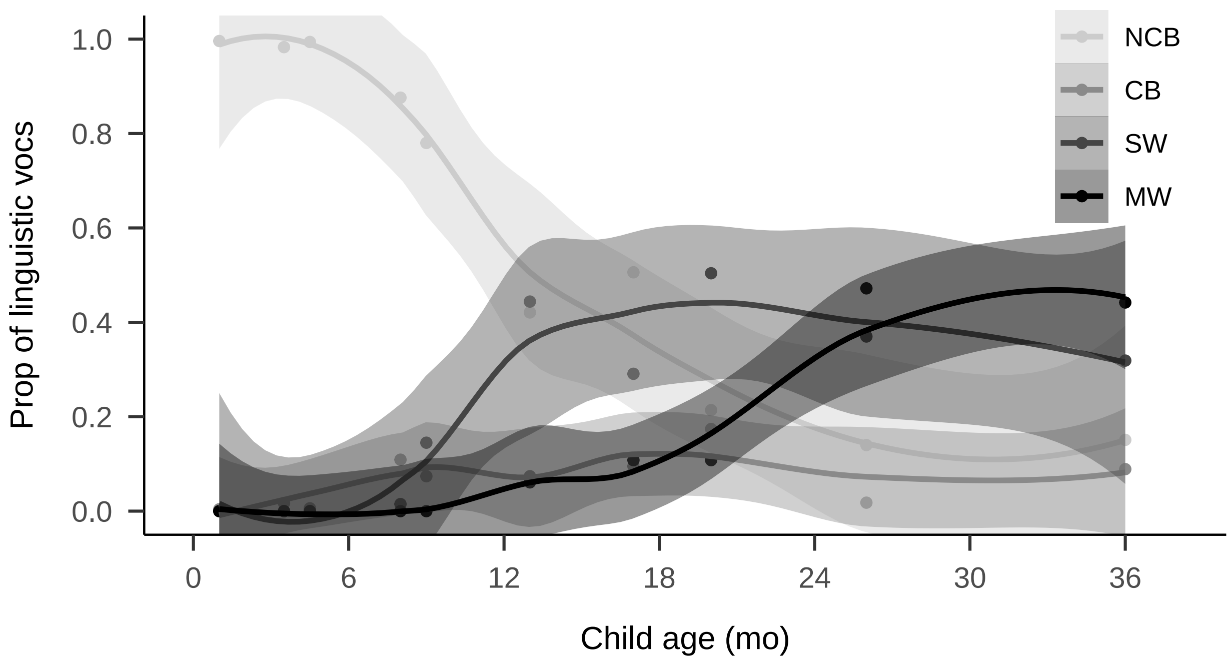

Figure 4. Proportion of vocalization types used by children across age (NCB = Non-canonical babble, CB = Canonical babble, SW = single word utterance, MW = multi-word utterance).

Discussion

We analyzed the speech environments of 10 Rossel children under age 3;0 to investigate: (a) how often children were spoken to directly, (b) how much other overhearable speech is available to them, (c) how these sources of linguistic input are shaped by child age and interactional context, and (d) whether this (relatively) low rate of directed input appears to impact their early production milestones.

Based on prior ethnographic work, we expected that these children would hear frequent child-directed speech from a wide variety of caregivers (REFS) and frequent speech directed to others (REFS). In fact children were rarely directly addressed. This surprisingly low baseline rate of TCDS is comparable—even slightly less—than that found in a Tseltal Mayan community where minimal TCDS is one means to socializing children into attending to their surroundings (REFS). On the other hand, there was ample overhearable speech directed to others; much more than has been reported for children in other communities at time of writing. We suspect that both the low relative rate of TCDS and the high incidence of ODS are partly attributable to the fact that multiple speakers are typically present in the recordings, as discussed further below.

Prior work using similar methods to those presented here, also led us to expect that the quantity of TCDS would be stable across the age range studied (Scaff, Casillas, Bergelson, REFS), and that an increasing proportion of it would come from other children (Shneidman, Brown, REFS). On the basis of these same studies, we expected the quantity of ODS to decrease across the age range studied (Casillas, Bergelson, REFS). Counter to expectations, we found a small but significant increase in TCDS rate with child age in the random clips and a small but significant decrease in TCDS rate with age in the turn-taking clips. The age-related baseline increase in TCDS may derive from more frequent participation in independent play with other children; in prior work, increased proportional input from other children was also associated with an increase in overall input rate (Shneidman refs). The age-related decrease in TCDS rate during peak interactional moments was not expected, but may be attributable to this change in interactional partners with age; if adults are more likely to be the source of TCDS during interactional peaks for younger children, they may also provide more voluminous speech during those peaks than other children do during interactional peaks later in development. Both of these explanations require follow-up work from a larger sample of children and, ideally, from a larger sample of their interactions throughout the day. As expected, we did see a decrease in ODS with age, consistent with prior work on both Western and non-Western samples (REFS).

Finally, while we anticipated that the children's input would be non-uniformly distributed over the recording day [Blasi et al. (in preparation); Abney et al. (2017); REFS Casillas], we also expected to see a somewhat even distribution of directed speech from morning to evening given that young children have been reported to pass between multiple caregivers during a typical day at home. We expected that this care-sharing practice might weaken the effect of farming activities in the late morning and early afternoon found in previous work with Tseltal Mayan subsistence farmers (Casillas et al. REFS). In fact, we found that children's rate of linguistic input was still significantly impacted by time of day, similar to prior work (Casillas et al. REFS). In paricular, most TCDS and ODS came during the morning, with older children more likely to hear TCDS at midday than younger children, possibly because this is when most adults are likely attending to gardening and domestic duties while children congregate in large play groups (REFS).

Talk to children: ideology vs. rate

- little speech to infants, possibly because of adult-adult talk w/ so many available speakers
- impact: daily activities and family members present
- cover more diverse speaker types during peaks
- consistency w/ ideology may be seen in the *how* and not in the *how often*

Independence and child-TCDS

- increase in CDS from other kids recalls Shneidman work w/ Yukatek kids, suggesting a possible role for speech from other kids in the first few years of life
- interestingly we didn't find this pattern w/ tseltal kids, possibly because in the Tseltal community extended independent play w/ other kids may start later, after age 3, where our sample ends

- ethnographic work on Rossel puts a heavy emphaseis on the early independence encouraged among young children, which could help explain why they appear more likely to participate in these child groups already before age 3

Limitations

- good for accurate measures of input
- we don't yet know how ideology impacts the shape of the input; we should look next to features of CDS in these contexts
- lacking semantically rich interpretations of lg use (needs the close study method)

Acknowledgements

This paper was written using the papaja library in RStudio (Aust & Barth, 2018).

References

Abney, D. H., Smith, L. B., & Yu, C. (2017). It's time: Quantifying the relevant time scales for joint attention. In G. Gunzelmann, A. Howes, T. Tenbrink, & E. Davelaar (Eds.), *Proceedings of the 39th Annual Meeting of the Cognitive Science Society* (pp. 1489–1494). London, UK.

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown.* Retrieved from https://github.com/crsh/papaja

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715. doi:10.1111/desc.12715

Blasi, D., Schikowski, R., Moran, S., Pfeiler, B., & Stoll, S. (in preparation). Human communication is structured efficiently for first language learners: Lexical spikes.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017). Modeling zero-inflated count data with glmmTMB. *bioRxiv*. doi:10.1101/132753

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs, & and B. B. Schieffelin (Eds.), *Handbook of Language Socialization* (pp. 29–55). Malden, MA: Wiley-Blackwell.

Casillas, M., Brown, P., & Levinson, S. C. (2017). Casillas HomeBank corpus. doi:10.21415/T51X12

Frank, M. C., Braginsky, M., Marchman, V. A., & Yurovsky, D. (in preparation). *Variability and consistency in early language learning: The Wordbank project.* Retrieved from https://langcog.github.io/wordbank-book/

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831. doi:10.1038/nrn1533

Oller, D. K., Eilers, R. E., Basinger, D., Steffens, M. L., & Urbano, R. (1995). Extreme poverty and the development of precursors to the speech capacity. *First Language*, *15*(44), 167–187.

Pine, J. M., & Lieven, E. V. M. (1993). Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language*, *20*(3), 551–571. doi:10.1017/S0305000900008473

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Slobin, D. I. (1970). Universals of grammatical development in children. In G. B. Flores d'Arcais & W. J. M. Levelt (Eds.), *Advances in Psycholinguistics* (pp. 174–186). Amsterdam, NL: North Holland Publishing.

Smithson, M., & Merkle, E. (2013). *Generalized linear models for categorical and continuous limited dependent variables*. New York: Chapman; Hall/CRC. doi:10.1201/b15694

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, *20*(6), e12456. doi:10.1111/desc.12456

Tomasello, M., & Brooks, P. J. (1999). Early syntactic development: A Construction Grammar approach. In M. Barrett (Ed.), *The Development of Language* (pp.

161–190). New York: Psychology Press.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2014). A social feedback loop for speech development and its reduction in Autism. *Psychological Science*, *25*(7), 1314–1324. doi:10.1177/0956797614531023

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from http://ggplot2.org