Child language experience in a Tseltal Mayan 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

We analyzed 9–11-hour at-home audio recordings from 10 Tseltal Mayan children between 0;2 and 3;0 to investigate how often they engaged in verbal interaction with others and whether their speech environment changed with age, time of day, household size, and number of speakers present. We found that Tseltal children are not often directly spoken to, that most directed speech comes from adults, and that directed speech does not increase with age. Most of children's directed speech came in the mornings or early evenings, particularly with younger children, and high interactional peaks tended to occur in bursts of turn taking that lasted approximately one minute. With some exceptions, these findings support previous characterizations of Mayan caregiver-child talk. An initial analysis of children's vocal development suggests that, despite relatively little directed speech, these children develop early language skills on a similar timescale to WEIRD children. Given these findings, we discuss multiple proposals for how Tseltal children might be efficient learners.

Keywords: Child-directed speech, Linguistic input, Non-WEIRD, Vocal maturity, Turn taking, Interaction, Mayan

Word count: X

Child language experience in a Tseltal Mayan village

Introduction

A great deal of work in developmental language science revolves around one central question: What linguistic evidence (i.e., what types and how much) is needed to support first language acquisition? In pursuing this topic, many researchers have fixed their sights on the quantity and characteristics of speech addressed to children; that is, speech designed for young recipients who may have limited attention and understanding (e.g., Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015; Hoff, 2006). In several languages, child-directed speech (CDS[1]) is linguistically accommodated to young listeners (A. Cristia, 2013; Soderstrom, 2007), interactionally rich (Bruner, 1983; Butterworth, 2003; Estigarribia & Clark, 2007; Masataka, 2003), and preferred by infants (Cooper & Aslin, 1990; ManyBabies Collaborative, 2017; Segal & Newman, 2015). In those same linguistic communities, these properties of CDS have been found to facilitate early word learning (Cartmill et al., 2013; e.g., Hirsh-Pasek et al., 2015; Hoff, 2003; Hurtado, Marchman, & Fernald, 2008; Rowe, 2008; Shneidman & Goldin-Meadow, 2012; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2012; Weisleder & Fernald, 2013). Yet ethnographic reports from a number of traditional, non-Western communities suggest that children easily acquire their community's language(s) even when the children are infrequently directly addressed (P. Brown, 2011). If so, large quantities of CDS may not be essential for learning language; just useful for facilitating certain aspects of language development. In this paper we investigate the language environment and early development of 10 Tseltal Mayan children growing up in a community where caregivers are reported to infrequently directly address speech to infants and young children (P. Brown, 1998b, 2011, 2014).

[1]Throughout this article, we use "child-directed speech" and "CDS" in the most literal sense: speech designed for and directed toward a child recipient.

Child-directed speech

Prior work on CDS in Western contexts has shown that the amount of CDS children hear influences their language development; more CDS is associated with larger and faster-growing receptive and productive vocabularies in young children (e.g., Hart & Risley, 1995; Hoff, 2003; Hurtado et al., 2008; Peter, Durrant, Bidgood, Pine, & Rowland, in preparation; Ramírez-Esparza, García-Sierra, & Kuhl, 2014, 2017; Shneidman & Goldin-Meadow, 2012; Shneidman et al., 2012; Weisleder & Fernald, 2013). CDS has also been linked to young children's speed of lexical retrieval (Hurtado et al., 2008; but see Peter et al., in preparation; Weisleder & Fernald, 2013) and syntactic development (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010). The conclusion drawn from much of this work is that speech directed to children is well designed for learning words—especially concrete nouns and verbs—because it is optimized for a child's attention in the moment it is spoken. Even outside of first-person interaction, infants and young children prefer listening to attention-grabbing CDS over adult-directed speech (ManyBabies Collaborative, 2017). There are, however, a few significant caveats to the body of work relating CDS quantity to language development.

First, while there is overwhelming evidence linking CDS quantity to vocabulary size, links to grammatical development are more scant (e.g., Brinchmann, Braeken, & Lyster, 2019; Frank, Braginsky, Marchman, & Yurovsky, in preparation; Huttenlocher et al., 2010). While the advantage of CDS for referential word learning is clear, it is less obvious how CDS facilitates syntactic learning. For example, utterance length (a proxy for syntactic complexity; Wasow, 1997) doesn't appear to increase with child age (Newport, Gleitman, & Gleitman, 1977), and parents may be less likely to directly correct their children's syntactic errors than their semantic ones (R. Brown, 1977; but see Chouinard & Clark, 2003)—even sometimes themselves producing ungrammatical utterances to make individual words salient to their young interlocutors (Aslin, Woodward, LaMendola, & Bever, 1996; see also Yurovsky, 2018). On the other hand, there is a wealth of evidence that syntactic knowledge is lexically specified (e.g., Arnold, Wasow, Asudeh, & Alrenga, 2004; Goldberg, 2003; Lieven, Pine, & Baldwin, 1997), and that, crosslinguistically, children's vocabulary size is one of the most robust predictors of their early syntactic development (Bates & Goodman, 1997; Frank et al., in preparation; Marchman, Martínez-Sussmann, & Dale, 2004). In short, what is good for the lexicon may also be good for syntax. For now, however, the link between CDS and other aspects of grammatical development still needs to be more thoroughly tested.

A second caveat is that most work on CDS quantity uses summary measures that average over the ebb and flow of interaction (e.g., proportion CDS). In both child and adult interactions, verbal behaviors are highly structured: while some occur at fairly regular intervals ("periodic"), others occur in shorter, more intense bouts separated by long periods of inactivity ("bursty" ; Abney, Dale, Louwerse, & Kello, 2018; Fusaroli, Razczaszek-Leonardi, & Tylén, 2014). For example, Abney and colleagues (2017) found that, across multiple time scales of daylong recordings, both infants' and adults' vocal behavior was clustered. Focusing on lexical development, Blasi and colleagues (in preparation) also found that nouns and verbs were used burstily in child-proximal speech across all six of the languages in their typologically diverse sample. Infrequent words were somewhat more bursty overall, leading them to propose that burstiness may play a key and universal role in acquiring otherwise-rare linguistic units. Experiment-based work also shows that two-year-olds learn novel words better from a massed presentation of object labels versus a distributed presentation (but see Ambridge, Theakston, Lieven, & Tomasello, 2006, and @childers2002two; Schwab & Lew-Williams, 2016). Structured temporal characteristics in children's language experience imply new roles for attention and memory in language development. Ideally, then, we should be investigating how CDS is distributed over children's daily experiences (Soderstrom & Wittebolle, 2013).

Finally, prior work has typically focused on Western (primarily North American) populations, limiting our ability to generalize these effects to children acquiring language worldwide (P. Brown & Gaskins, 2014; Henrich, Heine, & Norenzayan, 2010; Lieven, 1994; M.

Nielsen, Haun, Kärtner, & Legare, 2017). While we do gain valuable insight by looking at *within-population* variation (e.g., different socioeconomic or sub-cultures), we can more effectively find places where our assumptions break down by studying *new* populations. Linguistic anthropologists working in non-Western communities have long reported that caregiver interaction styles vary immensely from place to place, with some caregivers using little child-directed speech with young children (P. Brown & Gaskins, 2014; Gaskins, 2006; Lieven, 1994). Children in these communities reportedly acquire language with "typical"-looking benchmarks. For example, they start pointing and talking around the same time we would expect for Western middle-class infants (P. Brown, 2011, 2014; P. Brown & Gaskins, 2014; Liszkowski, Brown, Callaghan, Takada, & De Vos, 2012; but see Salomo & Liszkowski, 2013). These findings have had little impact on mainstream theories of word learning and language acquisition, partly due to a lack of directly comparable measures (P. Brown, 2014; P. Brown & Gaskins, 2014). If, however, children in these communities indeed acquire language without delay, despite infrequent CDS, we must reconsider what kind of linguistic evidence is necessary for children to learn language.

Language development in non-WEIRD communities

A growing number of researchers are using methods from developmental psycholinguistics to describe the language environments and linguistic development of children growing up in traditional and/or non-Western communities (Barrett et al., 2013; Demuth, Moloi, & Machobane, 2010; Fortier, Kellier, Fernández Flecha, & Frank, under review; see also Ganek, Smyth, Nixon, & Eriks-Brophy, 2018; Garcia, Roeser, & Höhle, 2018; Hernik & Broesch, 2018). We briefly highlight two recent efforts along these lines, but see Cristia and colleagues' (2017) and Mastin and Vogt's work (2016; 2015) for similar examples.

Scaff, Cristia, and colleagues (2017; in preparation) have used a number of methods to estimate how much speech children hear in a Tsimane forager-horticulturalist population in the Bolivian lowlands. From daylong audio recordings, they estimate that Tsimane children between 0;6 and 6;0 hear maximally ~5 minutes of directly addressed speech per hour, regardless of their age. For comparison, children from North American homes between ages 0;3 and 3;0 are estimated to hear ~11 minutes of CDS per hour in daylong recordings (Bergelson et al., 2018b). Tsimane children also hear ~10 minutes of other-directed speech per hour (e.g., talk between adults) compared to the ~7 minutes per hour heard by young North American children (Bergelson et al., 2018b). This difference may be attributable to the fact that the Tsimane live in extended family clusters of 3–4 households, so speakers are typically in close proximity to 5–8 other people (A. Cristia et al., 2017).

Shneidman and colleagues (2010; 2012) analyzed speech from one-hour at-home video recordings of children between ages 1;0 and 3;0 in two communities: Yucatec Mayan (Southern Mexico) and North American (a major U.S. city). Their analyses yielded four main findings: compared to the American children, (a) the Yucatec children heard many fewer utterances per hour, (b) a much smaller proportion of the utterances they heard were child-directed, (c) the proportion of utterances that were child-directed increased dramatically with age, matching U.S. children's by 3;0 months, and (d) most of the added CDS came from other children (e.g., older siblings and cousins). They also demonstrated that the lexical diversity of the CDS they hear at 24 months—particularly from adult speakers—predicted children's vocabulary knowledge at 35 months.

The current study

We examine the early language experience of 10 Tseltal Mayan children under age 3;0. Prior ethnographic work suggests that Tseltal caregivers do not frequently speak directly to their children until the children themselves begin to actively initiate verbal interactions (P. Brown, 2011, 2014). Nonetheless, Tseltal children develop language with no apparent delays. Tseltal Mayan language and culture has much in common with the Yucatec Mayan communities Shneidman reports on, allowing us to compare differences in child language environments between the two sites more directly than before.[2] We provide more details on this community and dataset in the Methods section.

We analyzed basic measures of Tseltal children's language environments including: (a) the quantity of speech directed to them, (b) the quantity of other-directed speech they could overhear from nearby speakers, (c) the rate of contingent responses to their vocalizations, (d) the rate of their own contingent responses to others' vocalizations, and (e) the duration of their interactional dyadic sequences. To link these findings to prior work on speech environment and development, we also rougly estimated the number of minutes per day they spent in "high turn-taking" interaction and outlined a basic trajectory for early vocal development (i.e., from non-canonical babbles to multi-word utterances).

Based on prior work, we predicted that Tseltal Mayan children are infrequently directly addressed, that the amount of CDS and contingent responses they hear increases with age, that most CDS comes from other children, and that, despite this, their early vocal development is on par with Western children. We additionally predicted that children's language environments would be bursty—that high-intensity interactions would be brief and sparsely distributed throughout the day, accounting for the majority of children's daily CDS.

Methods

Community

The children in our dataset come from a small-scale, subsistence farming community in the highlands of Chiapas in Southern Mexico. The vast majority of children grow up speaking Tseltal monolingually at home. The first few years of primary school are conducted mainly in Tseltal, but the remainder of primary school, secondary school, and any further education is conducted exclusively in Spanish. Nuclear families are often large (5+ children) and live in patrilineal clusters. Nearly all families grow staple crops such as corn and beans, but also cultivate bananas, chilies, squash, coffee, and more. Household and farming work is

[2]For a review of comparative work on language socialization in Mayan cultures, see Pye (2017).

divided among men, women, and older children. Women do much of the daily cleaning and food preparation, but also frequently work in the garden, haul water and/or firewood, and do other physical labor. A few community members (both men and women) earn incomes as teachers and shopkeepers but are still expected to regularly contribute to their family's household work.

More than forty years of ethnographic work by the second author has supported the idea that Tseltal children's language environments are non-child-centered and non-object-centered (P. Brown, 1998b, 2011, 2014). During their waking hours, Tseltal infants are typically tied to their mother's back while she goes about her work for the day. Infants receive very little direct speech until they themselves begin to initiate interactions, usually as they approach their first birthdays. Even then, interactional exchanges are often brief or non-verbal (e.g., object exchange routines) and take place within a multi-participant context (P. Brown, 2014). Rarely is attention given to words and their meanings, even when objects are central to the activity. Instead, interactions tend to focus on appropriate actions and responses, and young children are socialized to attend to the interactions taking place around them (León, 2000, see also 2011; Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003). That said, P. Brown (2014) points out that other language- and culture-specific properties may facilitate word and language learning, including conversational repetition, specific verbs, and the centrality of gesture and gaze in learning the spatial system.

Young children are often cared for by other family members, especially older siblings. Even when not on their mother's back, infants are rarely put on the ground. Therefore, children can't usually pick up the objects around them until they are old enough to walk. Toys are scarce and books are vanishingly rare, so the objects children do get their hands on tend to be natural or household objects (e.g., rocks, sticks, spoons, baskets, etc.). By age five, most children are competent speakers who engage daily in chores and the caregiving of their younger siblings. The Tseltal approach to caregiving is similar to that described for other Mayan communities (Gaskins, 1996, 1999; León, 1998, 2011; Pye, 1986; Rogoff et al., 1993, 2003; e.g., Shneidman & Goldin-Meadow, 2012).

Corpus

The current data come from the Casillas HomeBank Corpus (Casillas, Brown, & Levinson, 2017; VanDam et al., 2016), which includes daylong recordings and other developmental language data from more than 100 children under 4;0 across two indigenous, non-WEIRD communities: the Tseltal Mayan community described here and a Papua New Guinean community described elsewhere (P. Brown, 2011, 2014; P. Brown & Casillas, in press).

The Tseltal data, primarily collected in 2015, include recordings from 55 children born to 43 mothers. The families in our dataset typically only had 2–3 children (median = 2; range = 1–9), due to the fact that the participating families come from a young subsample of the community (mothers: mean = 26.3 years; median = 25; range = 16–43 and fathers: mean = 30; median = 27; range = 17—52). On average, mothers were 20 years old when they had their first child (median = 19; range = 12–27), with a following inter-child interval of 3 years (median = 2.8; range = 1–8.5).[3]. As a result, 28% of the participating families had two children under 4;0. To our knowledge at time of recording, all children were typically developing. We calculated the precise age of children based on the birthdates given by their caregivers, though these ages should be taken with a pinch of salt because documentation of birthdates is less rigorous than would be typically expected in Western post-industrial populations. Households size, defined in our dataset by the number of people sharing a kitchen or other primary living space, ranged between between 3 and 15 people (mean = 7.2; median = 7). Although 32.7% (18/55) of the target children are first-born, they were rarely the only child in their household. Most mothers had finished primary (37%) or secondary (30%) school, with a few more having completed preparatory (12%) or university (2%; 1 mother); the remainder (23%) had no schooling or did not complete primary school. All

[3]These estimates do not include miscarriages or children who passed away.

fathers had finished primary school, with most completing secondary school (44%) or preparatory school (21%), and two completing a university-level training (5%). Clan membership influences marriage and land inheritance such that 93% of the fathers grew up in the village where the recordings took place, while only 53% of the mothers did.

Recordings. We used a novel combination of a lightweight stereo audio recorder (Olympus® WS-832) and wearable photo camera (Narrative Clip 1®) fitted with a fish-eye lens, to track children's movements and interactions over the course of a 9–11-hour period in which the experimenter was not present. Each recording was made during a single day at home in which the recorder and/or camera was attached to the child. Ambulatory children wore both devices on an elastic vest (see Figure 1). Non-ambulatory children wore the recorder in a onesie while their primary caregiver wore the camera on an elastic vest. The camera was set to take photos at 30-second intervals and was synchronized to the audio in post-processing to create a video file featuring the snapshot-linked audio from the child's recording.[4] We used these recordings to capture a wider range of the linguistic patterns children hear as they participate in different activities with different speakers over the course of their day (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2018a; Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017).

Data selection and annotation

We annotated video clips from 10 of the 55 children's recordings. We chose these 10 recordings to maximize variance in three demographic variables: child age (0–3;0), child sex, and maternal education. The sample is summarized in Table 1. We then selected one hour's worth of non-overlapping clips from each recording in the following order: nine randomly selected 5-minute clips, five 1-minute clips manually selected as the top "turn-taking" minutes of the recording, five 1-minute clips manually selected as the top "vocal activity"

[4] Documentation and scripts for post-processing are available at and https://github.com/marisacasillas/Weave.

Figure 1. The recording vest fit over children's chests with an audio recording device in the front horizontal pocket and a camera fitted with a fisheye lens attached to the a shoulder strap.

minutes of the recording, and one, manually selected 5-minute extension of the best 1-minute sample (see Figure 2). We created these different subsamples of each day to measure properties of (a) children's *average* language environments (random samples) and (b) their *most input-dense* language environments (turn-taking samples). The third sample (high-activity) gave us insight into children's productive speech abilities, but is discussed elsewhere (Casillas, in progress).

The turn-taking and high-activity clips were chosen by two trained annotators (the first author and a student assistant) who listened to each recording in its entirety at 1–2x speed while actively taking notes about potentially useful clips. Afterwards, the first author

Table 1

Demographic overview of the 10 children whose recordings we sampled.

HB ID	Age	Sex	Mot age	Mot edu	Ppl in house
CM50	01;25	M	26	none	8
CM07	03;18	M	22	preparatory	9
CM11	05;29	F	17	secondary	15
CM23	07;15	F	24	primary	9
CM38	10;21	M	24	secondary	5
CM04	14;10	M	21	none	9
CM17	22;03	F	31	preparatory	9
CM25	26;25	F	17	primary	5
CM47	32;05	F	28	secondary	5
CM55	36;02	M	28	primary	6

reviewed the list of candidate clips, listened again to each one (at 1x speed with multiple repetitions), and chose the best five 1-minute samples for each of the two types of activity. Good turn-taking activity was defined as closely timed sequences of contingent vocalization between the target child and at least one other person (i.e., frequent vocalization exchanges). The "best" turn-taking clips were chosen because they had the most and most clear turn-switching activity between the target child and the other speaker(s). Good vocal activity clips were defined as clips in which the target child produced the most and most diverse spontaneous (i.e., not imitative) vocalizations. The "best" vocal activity clips were chosen for representing the most linguistically mature and/or diverse vocalizations made by the child over the day. All else being equal, candidate clips were prioritized when they contained less background noise or featured speakers and speech that were not otherwise frequently represented (e.g., CDS from older males). The best turn-taking clips and vocal activity clips often overlapped; turn-taking clips were selected from the list of candidates first, and then vocal-activity clips were chosen from the remainder. Again, these manually selected clips were chosen from audio that did not overlap with the initial "random" clip selection. Therefore the true peak turn-taking and vocal-activity clips for the day could possibly occur during the random clips—the manually selected samples are the best examples from the remaining 8–10 hours of audio. The instructions for manually selecting clips and the notes from our Tseltal clip selection can be found at https://github.com/marisacasillas/Tseltal-CLE/blob/master/audio_scanning_instructions.md.

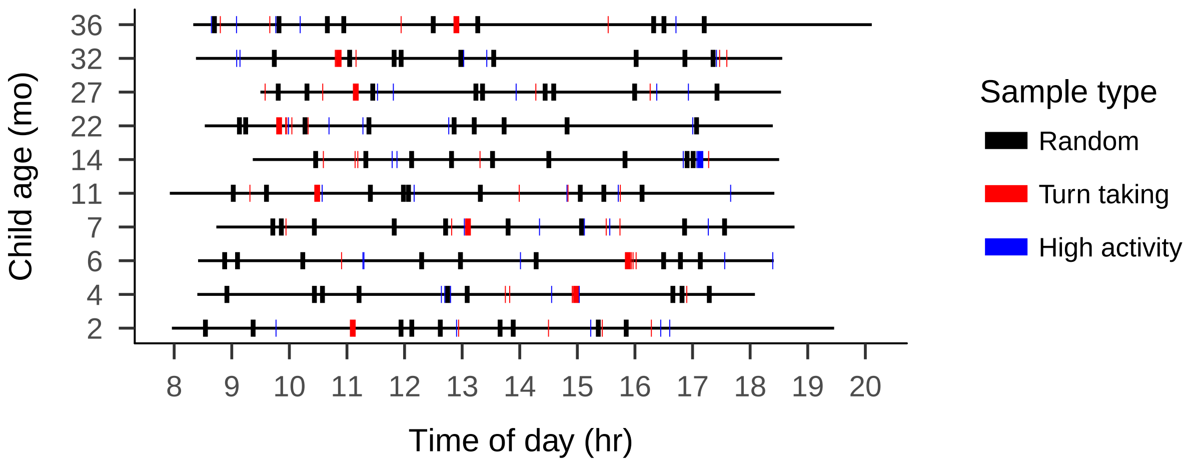

Figure 2. Recording duration (black line) and sampled clips (colored boxes) for each recording analyzed, sorted by child age.

The first author and a native speaker of Tseltal jointly transcribed and annotated each clip in ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) using the ACLEW Annotation Scheme (Casillas et al., 2017). The native Tseltal speaker lives in the community and knows most of the recorded families personally. The annotations include the transcription of (nearly) all hearable utterances in Tseltal, a loose translation of each utterance into Spanish, vocal maturity measures of each target child utterance (non-linguistic vocalizations/non-canonical babbling/non-word canonical babbling/single words/multiple words), and addressee annotations for all non-target-child utterances (target-child-directed/other-child-directed/adult-directed/adult-and-child-directed/animal-directed/other-speaker-type-directed). We annotated each utterance for intended addressee using contextual interactional information from the photos, audio, and preceding/following footage; we used an "unsure" category for utterances with no clear classification.[5] We exported each ELAN file as tab-separated values for all analysis.

Data analysis

Our aim in this paper is to describe the quantitative characteristics of Tseltal children's speech environments, as captured by the nine randomly selected five-minute clips from each child. We analyze five measures of children's speech environment: rate of target-child-directed speech (TCDS min/hr) and rate of other-directed speech (ODS min/hr), the rate of target-child-to-other turn transitions (TC–O transitions/min) and other-to-target-child turn transitions (O–TC transitions/min), and the duration of the target child's interactional sequences in seconds. For each of these five measures we investigate the effects of child age, time of day, household size, and number of speakers present. We then briefly report on a comparative analysis modelind the same effects on these five measures in the high turn-taking clips. We wrap up with two descriptive analyses: an initial estimate of the amount of time Tseltal children spend in high turn-taking interaction over the course of an entire day and an outline of the trajectory for early vocal maturity.

Results

Data analysis

Unless otherwise stated, all analyses were conducted with generalized linear mixed-effects regressions using the glmmTMB package and all plots are generate with

[5]Full documentation, including training materials, for the ACLEW Annotation Scheme can be found at https://osf.io/b2jep/wiki/home/.

ggplot2 in R (M. E. Brooks et al., 2017a; R Core Team, 2018; Wickham, 2009).[6] Notably, all five speech environment measures are restricted to non-negative values (min/hr, turn transitions/min, and duration in seconds), with a subset of them also displaying extra cases of zero in the randomly sampled clips (min/hr, turn transitions/min; e.g., when the child is napping). The consequence of these boundary restrictions is that the variance of the distributions becomes non-gaussian (i.e., a long right tail). We account for this issue by using negative binomial regression, whish is useful for overdispersed count data (M. E. Brooks et al., 2017b; Smithson & Merkle, 2013). When extra cases of zero are present due to, e.g., no speakers being present, we used a zero-inflation negative binomial regression, which creates two models: (a) a binary model to evaluate the likelihood of none vs. some presence of the variable (e.g., TCDS) and (b) a count model of the variable (e.g., “3” vs. “5” TCDS min/hr), using the negative binomial distribution as the linking function. Alternative analyses using gaussian models with logged dependent variables are available in the Supplementary Materials, but are qualitatively similar to the results we report here.

Our primary predictors were as follows: child age (months), household size (number of people), and number of non-target-child speakers present in that clip, all centered and standardized, plus squared time of day at the start of the clip (in decimal hours; centered on noon and standardized). We always used squared time of day to model the cycle of activity at home: the mornings and evenings should be more similar to each other than midday because people tend to disperse for chores after breakfast. To this we also added two-way interactions between child age and number of speakers present, household size, and time of day. Finally, we included a random effect of child, with random slopes of time of day, unless doing so resulted in model non-convergence. Finally, for the zero-inflation models, we included child age, number of speakers present, and time of day. We have noted below when models needed to deviate from this core design to achieve convergence. We only report

[6] The data and analysis code are freely available on the web ([retracted for review]), as is a summary of the results which will be updated as more transcriptions become available ([retracted for review]).

significant effects here; full model outputs are available in the Supplementary Materials.

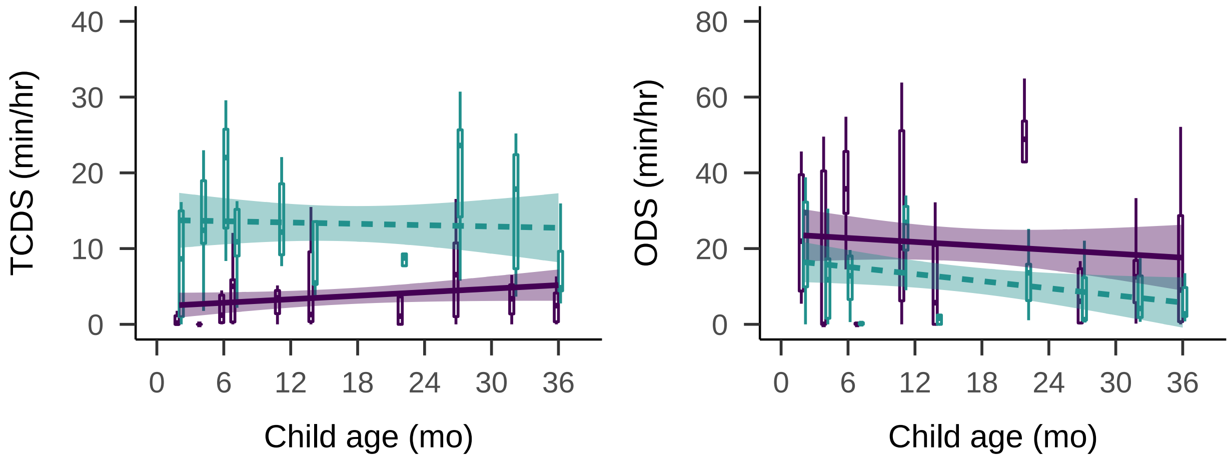

Figure 3. By-child estimates of minutes per hour of other-directed speech (left) and target-child-directed speech (right). Data are shown for the random (purple; solid) and turn taking (green; dashed) samples. Bands on the solid linear trends show 95% CIs.

Target-child-directed speech (TCDS)

The Tseltal children in our study were directly spoken to for an average of 3.63 minutes per hour in the random sample (median = 4.08; range = 0.83–6.55; Figure 3). These estimates are close to those reported for Yucatec Mayan data (Shneidman & Goldin-Meadow, 2012), which are plotted with our data, along with estimates from a few other populations in Figure 4 (US/Canada: Bergelson et al., 2018b; Tsimane: Scaff et al., in preparation, see Scaff and colleagues Scaff et al. (in preparation) for a more detailed comparison; US urban and Yukatek: Shneidman, 2010; Mozambique urban and rural, and Dutch: Vogt et al., 2015).[7]. We modeled TCDS min/hr in the random clips with a zero-inflated negative binomial regression, as described above.

The rate of TCDS in the randomly sampled clips was primarily affected by factors

[7] We convert the original estimates from Shneidman (2010) into min/hr by using the median utterance duration in our dataset for all non-target child speakers: (1029ms). Note that, though this conversion is far from perfect, Yukatek and Tseltal are related languages.

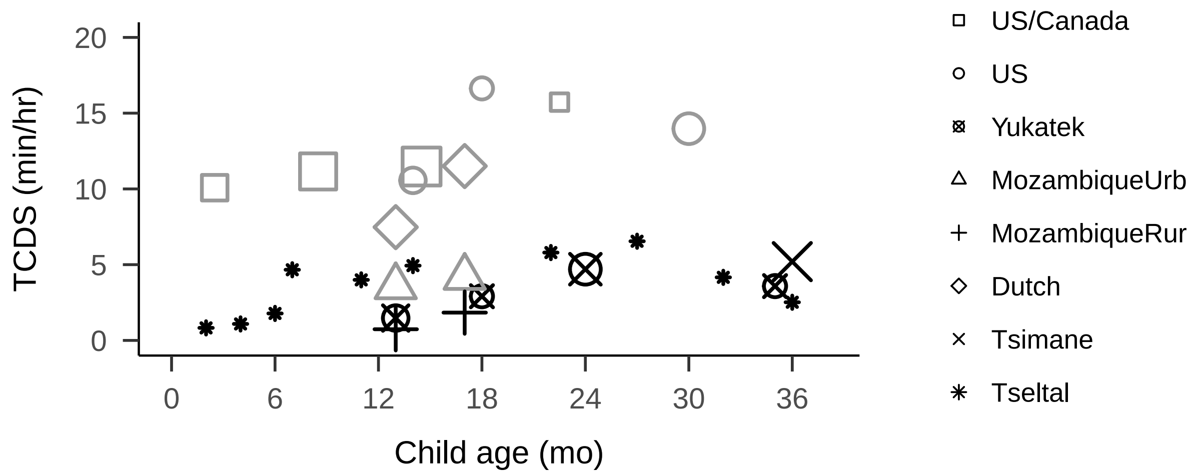

Figure 4. TCDS rate reported from daylong recordings made in different populations, including both urban (gray) and rural/indigenous (black) samples. Each point is the average TCDS rate reported for children at the indicated age, and size indicates number of children sampled (range: 1–26). See text for references to original studies.

relating to the time of day. The count model showed that, overall, children were more likely to hear TCDS in the mornings and evenings than around midday (B = 4.32, SD = 1.92, z = 2.25, p = 0.02). However, this pattern weakened for older children, some of whom even heard peak TCDS input around midday, as illustrated in Figure 5 (B = -5.22, SD = 1.97, z = -2.64, p = 0.01). There were no significant effects of child age, household size, or number of speakers present, no significant effects in the zero-inflation model.[8]

In contrast to findings from Shneidman and Goldin-Meadow (2012) on Yucatec Mayan, most TCDS in the current data came from adult speakers (mean = 80.61%, median = 87.22%, range = 45.90%–100), with no evidence for an increase in proportion TCDS from children with target child age (correlation between child age and proportion TCDS from children: Spearman's *rho* = -0.29; *p* = 0.42).

[8] This TCDS zero-inflation did not include the number of speakers present or time of day.

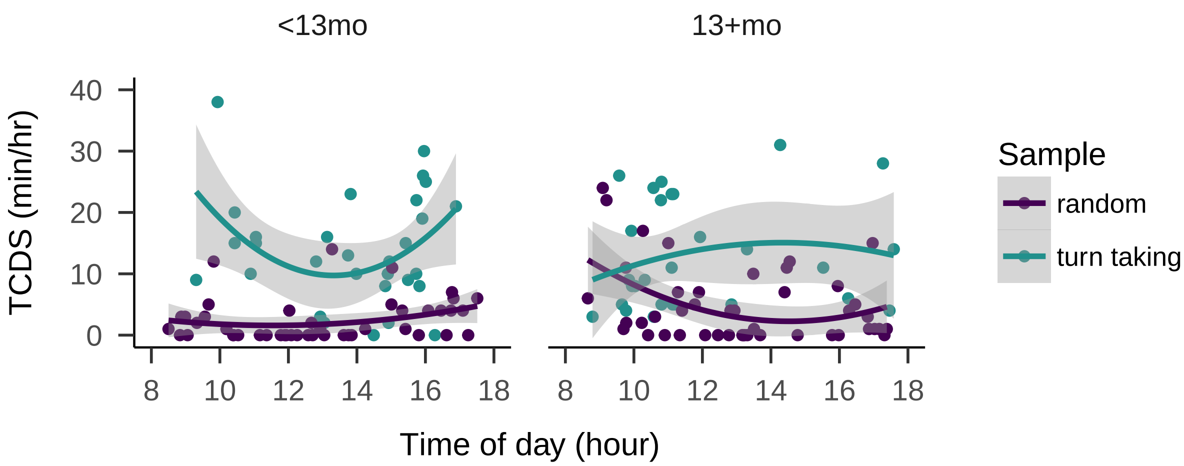

Figure 5. TCDS rate heard at different times of day by children 12 months and younger (left) and 13 months and older (right) in the randomly selected (purple) and turn-taking (green) clips.

Other-directed speech (ODS)

Children heard an average of 21.05 minutes per hour in the random sample (median = 17.80; range = 3.57–42.80): that is, 5–6 times as much speech as was directed to them. We modeled ODS min/hr in the random clips with a zero-inflated negative binomial regression, as described above.

The count model of ODS in the randomly selected clips revealed that the presence of more speakers was strongly associated with more ODS (B = 1.06, SD = 0.09, z = 11.54, p = 0). Additionally, more ODS occurred in the mornings and evenings (B = 2.70, SD = 1.14, z = 2.36, p = 0.02), and was also more frequent in large households for older children compared to younger children (B = 0.33, SD = 0.16, z = 2.01, p = 0.04). There were no other significant effects on ODS rate, and no significant effects in the zero-inflation models.[9]

Other-directed speech may have been so common because there were an average 3.44 speakers present other than the target child in the randomly selected clips (median = 3; range = 0–10), and (typically) more than half of the speakers were adults. However, these

[9]This ODS count model did not include by-child intercepts of time of day and its zero-inflation did not include the number of speakers present.

estimates may be comparable to North American infants (6–7 months) living in nuclear family homes (Bergelson et al., 2018a), so a high incidence of ODS may be common for infants in many sociocultural contexts.

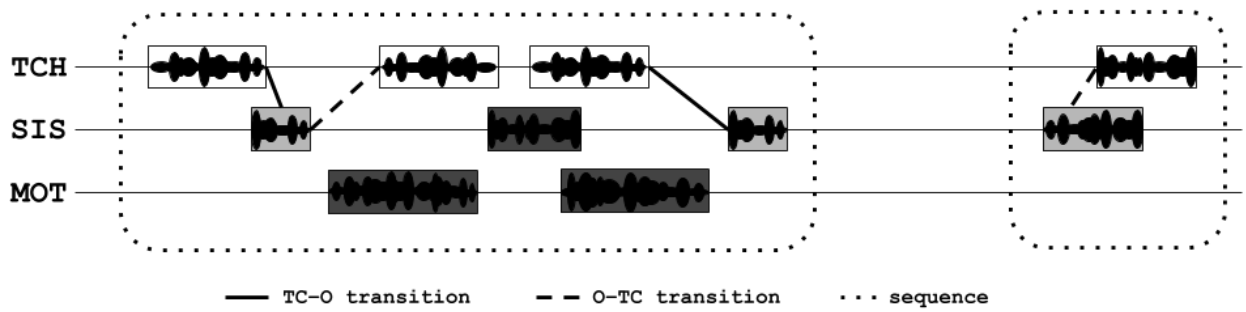

Figure 6. Illustration of a transcript clip between the target child (TCH), an older sister (SIS), and mother (MOT) in which transitions between the target child and other interlocutors are marked in solid and dashed lines and in which interactional sequences are marked with dotted lines. Light gray boxes indicate TCDS and dark gray boxes indicate ODS.

Target-child-to-other turn transitions (TC–O)

We detect contingent turn exchanges between the target child and other speakers based on turn timing Figure 6. If a child's vocalization is followed by a target-child-directed utterance within -1000–2000msec of the end of the child's vocalization (Casillas, Bobb, & Clark, 2016; Hilbrink, Gattis, & Levinson, 2015), it is counted as a contingent response (i.e., a TC–O transition). We use the same idea to find other-to-target-child transitions below (i.e., a target-child-directed utterance followed by a target child vocalization with the same overlap/gap restrictions). Each target child vocalization can only have one prompt and one response and each target-child-directed utterance can maximally count once as a prompt and once as a response (e.g., in a TC–O–TC sequence, the "O" is both a response and a prompt).

Gap and overlap restrictions are based on prior studies of infant and young children's turn taking (Casillas et al., 2016; Hilbrink et al., 2015), though the timing margins are increased slightly for the current dataset because the prior estimates come from relatively short, intense bouts of interaction in WEIRD parental contexts. Note, too, that much prior work has used maximum gaps of similar or greater length to detect verbal contingencies in caregiver-child interaction; and any work based on LENA® conversational blocks is thereby based on a 5-second silence maximum (Bergelson et al., 2018b; M. H. Bornstein, Putnick, Cote, Haynes, & Suwalsky, 2015; Broesch, Rochat, Olah, Broesch, & Henrich, 2016; Egeren, Barratt, & Roach, 2001; Y. Kuchirko, Tafuro, & Tamis-LeMonda, 2018; Romeo et al., 2018; Warlaumont, Richards, Gilkerson, & Oller, 2016); in comparison our timing restrictions are quite stringent.

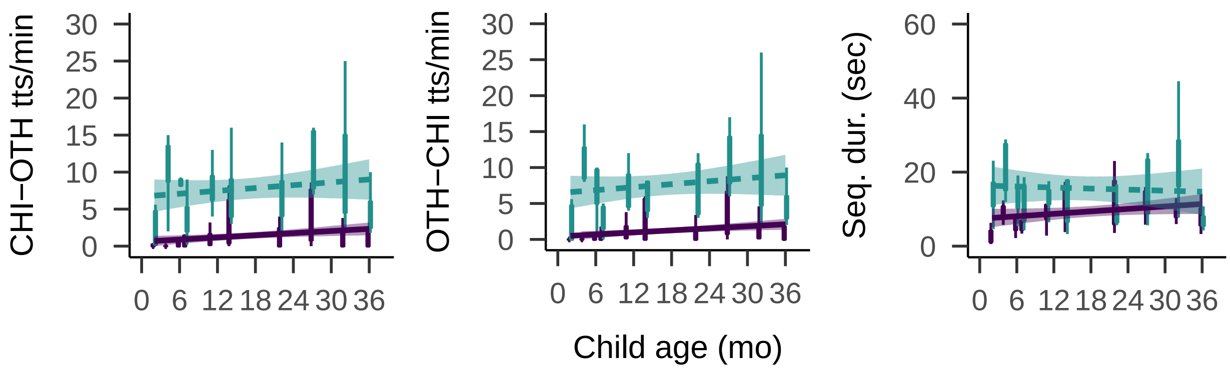

Figure 7. By-child estimates of contingent responses per minute to the target child's vocalizations (left), contingent responses per minute by the target child to others' target-child-directed speech (middle), and the average duration of contingent interactional sequences (right). Each datapoint represents the value for a single clip within the random (purple; solid) or turn taking (green; dashed) samples. Bands on the solid linear trends show 95% CIs.

Other speakers responded contingently to the target children's vocalizations at an average rate of 1.38 transitions per minute (median = 0.40; range = 0–8.60). We modeled TC–O transtions per minute in the random clips with a zero-inflated negative binomial regression, as described above.

The rate at which children hear contingent response from others was primarily influenced by factors relating to the child's age. Older children heard more contingent responses then younger children when there were more speakers present (B = 0.47, SD =

0.22, z = 2.11, p = 0.03). Also, as with the speech quantity measures, younger children heard more contingent responses in the mornings and evenings while this effect was less pronounced for older children (B = -6.46, SD = 2.56, z = -2.52, p = 0.01).There were no other significant effects on TC–O transition rate, and no significant effects in the zero-inflation model either.[10]

Other-to-target-child turn transitions (O–TC)

Tseltal children responded contingently to others' target-child vocalizations at an average rate of 1.17 transitions per minute (median = 0.20; range = 0–8.80). We modeled O–TC transtions per minute in the random clips with a zero-inflated negative binomial regression, as described above.

The rate at which children respond contingently to others (O–TC turn transitions per minute) was similarly influenced by child age and time of day: older children were less likely than young children to show peak response rates in the morning and evening (B = -7.30, SD = 2.61, z = -2.80, p = 0.01). There were no further significant effects in the count or zero-inflation models.[11]

Sequence duration

Sequences of interaction include periods of contingent turn taking with at least one target child vocalization and one target-child-directed prompt or response from another speaker. We use the same mechanism as before to detect contingent TC–O and O–TC transitions, but also allow for speakers to continue with multiple vocalizations in a row (e.g., TC–O–O–TC–OTH; Figure 7. Sequences are bounded by the earliest and latest vocalization for which there is no contingent prompt/response, respectively. Each target child vocalization can only appear in one sequence, and many sequences have more than one child vocalization. Because sequence durations were not zero-inflated, we modeled them in the random clips with negative binomial regression.

[10] This TC–O transition count model did not include by-child intercepts of time of day.

[11] This O–TC transition count model did not include by-child intercepts of time of day.

We detected 311 interactional sequences in the 90 randomly selected clips, with an average sequence duration of 10.13 seconds (median = 7; range = 0.56–85.47). The average number of child vocalizations within these sequences was 3.75 (range = 1–29; median = 3). None of the predictors significantly impacted sequence duration (all $p > 0.09$).[12]

Peak interaction

As expected, the turn-taking clips featured a much higher rate of contingent turn transitions: the average TC–O transition rate was 7.73 transitions per minute (median = 7.80; range = 0–25) and the average O–TC rate was 7.56 transitions per minute (median = 6.20; range = 0–26). The interactional sequences were also longer on average: 12.27 seconds (median = 8.10; range = 0.55–61.22).

Crucially, children also heard much more TCDS in the turn-taking clips—13.28 min/hr (median = 13.65; range = 7.32–20.19)—while also hearing less ODS—11.93 min/hr (median = 10.18; range = 1.37–24.42).

We modeled each of these five speech environment measures with parallel models to those used above (with no zero-inflation model for TCDS, TC–O, and O–TC rates, given the nature of the sample). The impact of child age, time of day, household size, and number of speakers was qualitatively similar (basic sample comparisons are visualized in Figure 3, Figure 4, and Figure 6) between the randomly selected clips and these peak periods of interaction with the following exceptions: older children heard significantly less ODS (B = -0.47, SD = 0.20, z = -2.39, p = 0.02), the presence of more speakers significantly decreased children's response rate to other's vocalizations (B = -0.26, SD = 0.12, z = -2.19, p = 0.03), and children's interactional sequences were shorter for older children (B = -0.24, SD = 0.10, z = -2.42, p = 0.02), shorter for children in large households (B = -0.21, SD = 0.10, z = -2.25, p = 0.02), and longer during peak periods in the mornings and afternoons (B = 2.76, SD = 1.11, z = 2.50, p = 0.01). Full model outputs can be compared in the Supplementary

[12] This sequence duration model did not include by-child intercepts of time of day.

Materials.

Peak minutes in the day. Now knowing the interactional characteristics of the "high" turn-taking clips, we looked for similarly interactive 1-minute sections in the random samples in order to estimate the number of high interactivity minutes in the whole day. To do this, we scanned each 60-second window (e.g., 0–60 sec, 1–61 sec, etc.[13]) of each random clip from each child and recorded the observed turn-transition rate. Only 6 of the 10 children showed at least one minute of their random sample that equalled or exceeded the grand average turn-transition rate (12.89 transitions per minute), and 7 of the 10 children showed at least one minute equalling or exceeding their own average turn transition rate from their turn-taking samples, as shown in Figure 8. Across children who did show turn-taking "peaks" in their random data (i.e., at or above rates from the sample-average from the turn-taking segments), periods of "peak" interaction were relatively long, ranging in duration from an average of 0 to 103 seconds across the 6 children with such peaks.

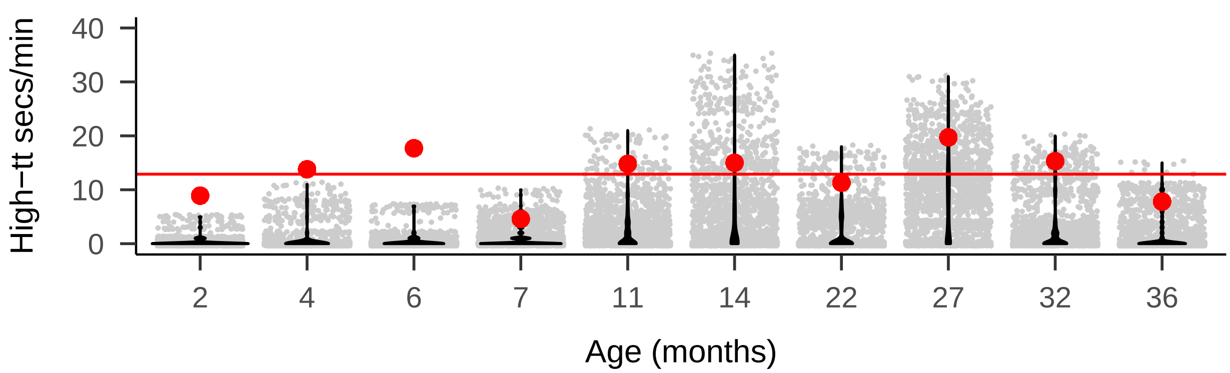

Figure 8. Turn-transitions rates, estimated over the last 60 seconds for each second of the random samples by child (nine 5-min clips each). The horizontal line indicates the group mean turn-transition rate in the turn-taking sample. The large points indicate the by-child mean turn-transition rate in the turn-taking sample.

Assuming approximately 12 waking hours, we therefore very roughly estimate that these Tseltal children spent an average of 100.16 minutes (1.67 hours) in high turn-taking,

[13]60 seconds is the smallest clip sample size in the turn-taking segments

dyadic interaction during their recording day. However, the range in the quantity of high turn-taking interaction varies enormously across children, starting at just a few minutes per day and topping out at more than 419.73 minutes (7 hours) in our sample.

Vocal maturity

Children's vocalizations appear to follow the normative benchmarks for productive speech development, as typically characterized by the *onset* of new production features. Decades of research in WEIRD populations has shown that, typically, children begin producing non-canonical babbles around 0;2, with canonical babbling appearing around 0;7, first words around 1;0 (P. K. Kuhl, 2004; Oller, 1980; Warlaumont & Finnegan, 2016; Warlaumont et al., 2016), and first multi-word utterances just after 1;6 (Braine & Bowerman, 1976; Fine & Lieven, 1993; Frank et al., in preparation; Slobin, 1970; Tomasello & Brooks, 1999). These benchmarks are mirrored in the Tseltal data (see Figure 9), which includes all annotated vocalizations from the random, turn-taking, and high vocal activity samples (N = 4725): there is a decline in the use of non-canonical babble and an accompanying increase in the use of canonical babble from 0;6 to 1;0. Recognizable words are also observed for every child from age 11;0 and older. Multi-word utterances already appear with the child at 1;2 and make up ~10–15% of children's utterances through the child at 2;3. The oldest two children use multi-word utterances in 33% and 45% of their vocalizations respectively.

These data are also consistent with usage statistics of speech-like vocalizations by WEIRD infants (Oller, 1980; Warlaumont & Finnegan, 2016; Warlaumont et al., 2016). In their Warlaumont et al. (2016) study, Warlaumont and colleagues found that the proportion of speech-like vocalizations (speech, non-word babble, and singing) was ~0.6 around age one in their SES-variable dataset of 106 children. We estimated the number of speech-like (canonical babbling and lexical speech) and non-speech-like (cries, laughter, and non-canonical babbles) vocalizations from Tseltal children 14 months and younger[14] across the random, turn-taking, and high vocal activity samples (N = 3020 from 6 children).

[14]We cannot compare speech-like vs. non-speech-like vocalizations after 1;6 due to a limitation of the

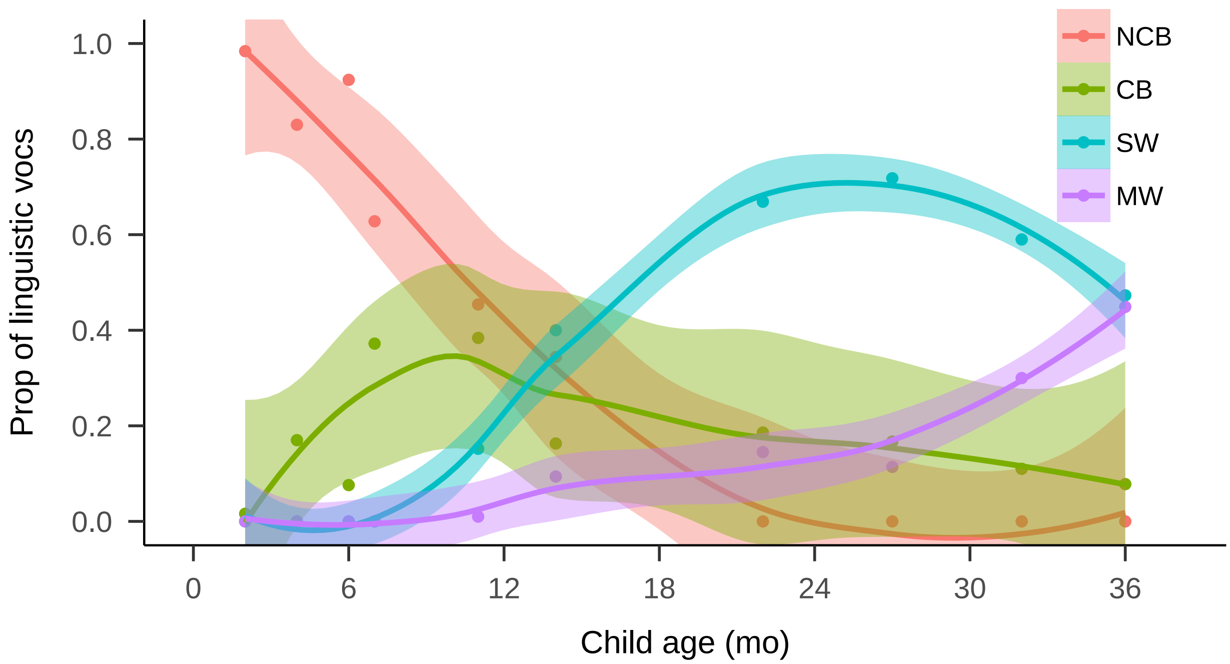

Figure 9. Proportion of vocalization types used by children across age (NCB = Non-canonical babble, CB = Canonical babble, SW = single word utterance, MW = multi-word utterance).

Between 2 and 14 months, Tseltal children demonstrated a large increase in the proportion of speech-like vocalizations (canonical babbling and lexical speech): from 9% before 0;6 to 58% between 0;10 and 1;2. Around age 1;0, their use of speech-like vocalizations (58%) is nearly identical to that estimated by Warlaumont et al. (2016) for American children (~60%).

Discussion

We analyzed 10 Tseltal Mayan children's speech environments in order to estimate how often they have the opportunity to attend and respond to speech directed to them. We also investigated how these speech environment characteristics changed (or stayed stable) across different ages, household sizes, time of day, and number of speakers present. To achieve representative estimates, we sampled audio segments randomly from each child's daylong recording, but we show that most of the same general patterns hold up during the "peak" turn-taking moments of the day as well. Finally, we roughly estimated the number of "high

ACLEW Annotation Scheme (Casillas et al., 2017).

interactivity" minutes Tseltal children encounter on a typical day and demonstrated that, despite the relatively small quantity of CDS, children's early vocal development was on-par with norms built from WEIRD children's data. These findings, which use a new methodology (i.e., daylong recordings with multiple sample types), partly replicate estimates of child language input and development in previous ethnographic and psycholinguistic work on Yucatec and Tseltal Mayan communities (P. Brown, 1997, 1998c, 2011, Tseltal: 2014; Shneidman, 2010; Yucatec: Shneidman et al., 2012). In what follows we briefly review each of the predictions made at the outset of the paper.

How much directed speech do Tseltal children hear?

The bulk of our analyses were aimed at understanding how much speech Tseltal children hear: we wanted to know how often they are directly spoken to and how often they might be able to listen to other-directed speech around them. As suggested by prior work, the children were only infrequently directly spoken to: an average of 3.63 minutes per hour in the random sample. Compared to other studies based on daylong recordings, the Tseltal average TCDS rate is approximately a third of that found for North American children (Bergelson et al., 2018b), but is comparable to that for Tsimane children (Scaff et al., in preparation). The CDS estimates also fall almost precisely in-line with those based on short-format recordings in a Yucatec Mayan village (Shneidman, 2010; Shneidman et al., 2012).

A novel contribution of this study is that we also included interactive measures to describe how often children were directly engaged with an interlocutor, either as a responder or as an addressee being responded to. We found that children's vocalizations were responded to at a rate of 1.38 speaker transitions per minute and that children responded to others' child-directed vocalizations at a rate of 1.17 transitions per minute. Prior work from a number of different domains suggests that contingent interaction (and the joint attention that likely accompanies it) is an ideal context for language learning since the child and interlocutor's coordinated attentional states decrease referential uncertainty, are a source of dynamic feedback, and can spur more interactions in the near future (M. H. Bornstein et al., 2015; Broesch et al., 2016; Egeren et al., 2001; Y. Kuchirko et al., 2018; Romeo et al., 2018; Warlaumont & Finnegan, 2016; Warlaumont et al., 2016). Because our measure is a novel one, we cannot directly compare Tseltal children's data with those of children growing up in other communities. That said, 1.38 and 1.17 transitions per minute suggests that contingent responses—more so than speech directly addressed to the child—are rare across children's day. Importantly, however, this may be due to the fact that children did not vocalize very often.

Preliminary analyses of children's vocal maturity showed that, on average, children only produced 472.50 vocalizations (many of which were crying, laughter, and non-canonical babble) during the entire 1-hour of clips sampled from their daylong recordings. This explanation resonates with the fact that, despite the low frequency of contingent turn-taking in the random sample, interactional sequences were fairly long: 10.13 seconds. During these long sequences, children tended to only vocalize 3.75 times, meaning that many of children's dyadic interactional sequences are marked by longer streams of directed input from another speaker, interspersed with only occasional responses from the child. Interactional peaks with contingent turn-taking *do* occur in the data, only rarely; our rough estimate is that Tseltal children participate in approximately 100.16 minutes of such interaction during a 12-hour waking day, most of which come in bursts of ~53 seconds long.

In sum, our results confirm prior claims that Tseltal children, like other Mayan children, are not often directly spoken to. When they are, much of their speech comes in interactional sequences in which children only play a minor part—directly contingent turn transitions between children and their interlocutors are relatively rare. However, we estimate that the average child under age 3;0 experiences more than one cumulative hour of high-intensity contingent interaction with CDS per day. If child-directed speech quantity linearly feeds language development (such that more input begets more output), then the estimates presented here would lead us to expect that Tseltal children are delayed in their language development, at least relative to North American children. However, our initial analyses of early vocal development suggest that Tseltal children, though they may not vocalize often, demonstrate vocal maturity comparable to children from societies in which CDS is known to be more frequent (Braine & Bowerman, 1976; Fine & Lieven, 1993; Frank et al., in preparation; P. K. Kuhl, 2004; Oller, 1980; Tomasello & Brooks, 1999; Warlaumont & Finnegan, 2016; Warlaumont et al., 2016). How do Tseltal children manage this feat?

Other-directed speech. One proposal is that Mayan children become experts at learning from observation during their daily interactions (León, 2011; Rogoff et al., 2003; Shneidman, 2010; Shneidman et al., 2012), thereby bridging the "gap" otherwise left by the lower rate of CDS. In the randomly selected clips, children were within hearing distance of other-directed speech for an average of 21.05 minutes per hour. That is substantially more than the ~7 minute per hour heard by North American children (Bergelson et al., 2018b), but comparable to the ~10 minutes per hour heard by Tsimane children (Scaff et al., in preparation). The large quantity of other-directed speech is likely due to the fact that Tseltal children (like Tsimane children) tend to live in households with more people compared to North American children (Bergelson et al., 2018a).

In our data, the presence of more speakers was associated with significantly more other-directed speech, both based on the number of individual voices present in the clip and on the number of people living in the household (for younger children). In comparison, children also heard more CDS when more speakers were present, but the effect was much weaker (0.2440 vs. 1.05622 more minutes per hour per speaker unit). This finding rings true with Brown's (2011, 2014) claim that Tseltal is a non-child-centric language community; the presence of more people somewhat increases talk to the child but really primarily increases talk amongst the other speakers. However, given that this increase in the number of speakers and amount of talk is also associated with an increase in the amount of overlapping speech (G. Cristia Alejandrina, 2018), we suggest that attention to other-directed speech is at least not the only learning mechanism needed to explain the robustness of early vocal development in Tseltal. Furthermore, just because speech is hearable does not mean children are attending to it. Follow-up work on the role of other-directed speech in children's speech development would need to clarify what constitutes viable "listened to" speech by the child.

Increased CDS with age. Another possibility is that CDS increases rapidly with child age (and vocalization competence). Combined with the idea that very early vocalizations follow a relatively species-specific, pre-programmed path that is then modulated by caregiver response and other factors (Oller, 1980; Oller, Griebel, & Warlaumont, 2016; Warlaumont & Finnegan, 2016; Warlaumont et al., 2016), a dramatic increase in directed speech with age might be expected. In her longitudinal studies of Yucatec Mayan children, Shneidman (2012) found that CDS increased significantly with age, from 55 utterances in an hour to 209 between 13 and 35 months. Her analyses show that most of the increase in CDS comes from other children speaking to the target child. Her findings are consistent with other reports that Mayan children are more often cared for by their older siblings from later infancy onward (2011, 2014). In our data, however, age effects were limited, and CDS from children was relatively rare (~10%) all the way up through age 3;0.

Child age alone had very little overall effect on the speech environment measures. In the random sample it was, at most, associated with marginal increases in TC–O transition rate, O–TC transition rate, and sequence duration. All other significant effects involving age related to the time of day measure, which is discussed below.

The non-increase in CDS with age may be due to the fact CDS from other children was infrequent in our data (cf. Shneidman & Goldin-Meadow, 2012). The relative lack of CDS may be due to the fact that: (a) the children were relatively young and so spent much of their time with their mothers, (b) these particular children did not have many older siblings, and (c) in the daylong recording context more adults were present to talk to each other than would be typical in a short-format recording (the basis for previous estimates). We conclude, from these findings, that an increase in CDS cannot explain the robust pattern of Tseltal vocal development either.

Learning from short periods of interaction. A third possibility is that these children learn effectively from short, routine language encounters. Bursty input appears to be the norm across a number of linguistic and interactive scales (Abney et al., 2018; Blasi et al., in preparation; Fusaroli et al., 2014), and experiment-based work suggests that children can benefit from massed presentation of new information (Schwab & Lew-Williams, 2016). We propose two mechanisms through which Tseltal children might capitalize on the distribution of speech input in their environment: they experience most language input during routine activities and they can consolidate experienced input during the downtime between interactive peaks. Neither of these mechanisms are proposed to be particular to Tseltal children, but might be employed to explain their efficient learning.

Tseltal children's linguistic input is not uniformly distributed over the day: all five measures of children's linguistic environment were more likely to occur in the mornings and afternoons than around midday, though younger children showed this pattern more robustly than older children. We had predicted a dip in linguistic input around midday because household members tend to disperse after breakfast to do their daily work before returning for another late afternoon meal. Young children, who are typically carried by their mothers for the majority of the day, followed this pattern more strongly than older children, who may have been more free to seek out interactions between mealtimes. A similar midday dip has been previously found for North American children's daylong recordings (Greenwood et al., 2011; Soderstrom & Wittebolle, 2013), suggesting that non-uniform distributions of linguistic input may be the norm for children in a variety of different cultural-economic contexts. Our paper is the first to show that those time of day effects change with age in the first few years on a number of speech environment features (TCDS, TC–O transitions, O–TC transitions, and (marginally) ODS).

Our impression from having transcribed these data is that the time of day effects likely arise from the activities that typically occur in the mornings and afternoons—meal preparation and dining times in particular—while napping could contribute to the midday dip (Soderstrom & Wittebolle, 2013). Indeed, time of day effects in daylong recordings at Canadian homes and daycares were substantially weakened when naptimes were excluded from the analysis (Soderstrom & Wittebolle, 2013). However, in the same Canadian data, the highest density speech input came during storytime and organized playtime (e.g., sing-alongs, painting), while mealtime was associated with less speech input. We expect that follow-up research which tracks activities in the Tseltal data will lead to very different conclusions: storytime and organized playtime are vanishingly rare in this non-child-centric community, and mealtime may represent a time of routine and rich linguistic experience. In both cases, however, the underlying association with activity (not hour) implies the possibility for action routines that may help children optimally extract information about what they will encounter and what they are expected to do in response, even over short periods (Bruner, 1983; Ferrier, 1978; tamis2018routine; Nelson, 1985; Shatz, 1978; Snow & Goldfield, 1983).

A more speculative possibility is that Tseltal children learn language on a natural input-consolidation cycle: the rarity of interactional peaks throughout the day may be complemented by an opportunity to consolidate new information. Sleep has been shown to benefit language learning tasks in both adults (Dumay & Gaskell, 2007; Frost & Monaghan, 2017; Mirković & Gaskell, 2016) and children (Gómez, Bootzin, & Nadel, 2006; Horváth, Liu, & Plunkett, 2016; Hupbach, Gomez, Bootzin, & Nadel, 2009; Williams & Horst, 2014), including word learning, phonotactic constraints, and syntactic structure. Our impression, both from the recordings and informal observations made during visits to the community, is that young Tseltal children take frequent naps, particularly at younger ages when they spend much of their day wrapped within the shawl on their mother's back. Mayan children tend to pick their own resting times (i.e., there are no formalized "sleep" times, even at bedtime Morelli, Rogoff, Oppenheim, & Goldsmith, 1992), and Mayan mothers take special care to keep infants in a calm and soothing environment in the first few months of life (Brazelton, 1972; e.g., León, 2000; Pye, 1992; E. Z. Vogt, 1976). There is little quantitative data on

Mayan children's daytime and nighttime sleeping patterns, but one study estimates that Yucatec Mayan children between 0;0 and 2;0sleep or rest nearly 15% of the time between morning and evening (Gaskins, 2000), again, at times that suited the child (@ Morelli et al., 1992). If Tseltal children's interactional peaks are bookended by short naps, it could contribute to efficient consolidation of new information encountered. How often Tseltal children sleep, how deeply, and how their sleeping patterns may relate to their linguistic development is an important topic for future research.

Limitations and Future Work

The current findings are based on a cross-sectional analysis of only 10 children. From each child, we have manually only analyzed a total of 1 of the 9–11 recording hours. The findings only take into account verbal input; the photo-linked audio we produce is not sufficient to analyze gaze and gestural behavior (P. Brown, 2014). In short, more data, and more kinds of data are needed to enrich this initial description of Tseltal children's early language environments. We have also used vocal maturity as an index of linguistic development in the current study, but further analysis of these children's receptive and productive lexical, morphological, and syntactic knowledge, including experiment and questionnaire based measures that build on past linguistic work (P. Brown, 1997, 1998b, 1998c, 1998a, 2011, 2014; P. Brown & Gaskins, 2014) is needed to establish trajectory of early language development in Tseltal.[15] To fully understand the extent to which language learning mechanisms are shared across ethnolinguistically diverse samples we cannot simply continue to compare developmental benchmarks. More promising long-term approaches include a focus on how within-community differences and/or cross-linguistic differences for related languages drive variation in learning (e.g., Pye, 2017; Weisleder & Fernald, 2013). The current analyses are based on a corpus that is under active development. As new data are added, up-to-date versions of the same analyses will be available on the same page where

[15] Other corpus- and experiment-based data and analyses from the Tseltal community are made available via the Casillas HomeBank corpus.

the current data and analysis scripts can be found: URL.

Conclusion

Based on the current data, we estimate that Tseltal children hear an average of 3.6 minutes of directed speech per hour. Contingent turn-taking is relatively rare throughout their day, and high-intensity interactive input comes in short bursts, typically in the mornings and early evenings for younger children. Despite this relatively small quantity of directed speech, Tseltal children's vocal maturity looks on-track with estimates based on WEIRD populations, in which children typically experience more child-directed speech. Our findings by and large replicate the descriptions put forth by linguistic anthropologists who have worked with Mayan communities for many decades. The real puzzle is then how Tseltal children efficiently extract information from their linguistic environments. We reviewed several proposals and outlined directions for future work. In our view, a promising avenue for continued research is to more closely investigate the activity/time-of-day effects and a possible input-consolidation cycle for language exposure in early infancy. By better understanding how Tseltal children learn language, we hope to uncover some of the ways in which human learning mechanisms are adaptive to the thousands of ethnolinguistic environments in which children develop.

Acknowledgements

Needs to be adjusted This work would not be possible without the efforts of Rebeca, Beti, Xun, their cousin, and Antun. We are enormously grateful to the participating families for sharing their lives with us. We also thank the community and community leaders in Tenejapa for supporting our work. Thanks too, to Maartje Weenink and Daphne Jansen, who have spent many hours identifying clips for transcription and errors in annotation. This work was supported by a NWO Veni Innovational Scheme grant (275-89-033) to MC and by an ERC Advanced Grant (269484 INTERACT) project to SCL. This manuscript was written using the papaja library in RStudio (Aust & Barth, 2018; RStudio Team, 2016).

References

Abney, D. H., Dale, R., Louwerse, M. M., & Kello, C. T. (2018). The bursts and lulls of multimodal interaction: Temporal distributions of behavior reveal differences between verbal and non-verbal communication. *Cognitive Science*, *XX*, XX–XX. doi:10.1111/cogs.12612

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.

Ambridge, B., Theakston, A. L., Lieven, E. V. M., & Tomasello, M. (2006). The distributed learning effect for children's acquisition of an abstract syntactic construction. *Cognitive Development*, *21*(2), 174–193.

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. *Journal of Memory and Language*, *51*(1), 55–70.

Aslin, R. N., Woodward, J. Z., LaMendola, N. P., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 117–134). New York, NY: Psychology Press.

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

Barrett, H., Broesch, T., Scott, R. M., He, Z., Baillargeon, R., Wu, D., ... Stephen Laurence. (2013). Early false-belief understanding in traditional non-Western societies. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1755), XX–XX. doi:10.1098/rspb.2012.2654

Bates, E., & Goodman, J. C. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive*

Processes, *12*(5–6), 507–584. doi:doi.org/10.1080/016909697386628

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2018a). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *XX*, XX–XX.

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2018b). What do north american babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *XX*, XX–XX.

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

Bornstein, M. H., Putnick, D. L., Cote, L. R., Haynes, O. M., & Suwalsky, J. T. D. (2015). Mother-infant contingent vocalizations in 11 countries. *Psychological Science*, *26*(8), 1272–1284.

Braine, M. D. S., & Bowerman, M. (1976). Children's first word combinations. *Monographs of the Society for Research in Child Development*, 1–104.

Brazelton, T. B. (1972). Implications of infant development among the mayan indians of mexico. *Human Development*, *15*(2), 90–111.

Brinchmann, E. I., Braeken, J., & Lyster, S.-A. H. (2019). Is there a direct relation between the development of vocabulary and grammar? *Developmental Science*, *22*(1), e12709.

Broesch, T., Rochat, P., Olah, K., Broesch, J., & Henrich, J. (2016). Similarities and differences in maternal responsiveness in three societies: Evidence from Fiji, Kenya, and the United States. *Child Development*, *87*(3), 700–711.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017a). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. Retrieved from https://journal.r-project.org/archive/2017/RJ-2017-066/index.html

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., … Bolker, B. M. (2017b). Modeling zero-inflated count data with glmmTMB.

bioRxiv. doi:10.1101/132753

Brown, P. (1997). Isolating the cvc root in tzeltal mayan: A study of children's first verbs. In E. V. Clark (Ed.), *Proceedings of the 28th annual child language research forum* (pp. 41–52). Stanford, CA: CSLI Publications/University of Chicago Press.

Brown, P. (1998a). Children's first verbs in tzeltal: Evidence for an early verb category. *Linguistics*, *36*(4), 713–753.

Brown, P. (1998b). Conversational structure and language acquisition: The role of repetition in tzeltal adult and child speech. *Journal of Linguistic Anthropology*, *2*, 197–221. doi:10.1525/jlin.1998.8.2.197

Brown, P. (1998c). Early tzeltal verbs: Argument structure and argument representation. In E. V. Clark (Ed.), *Proceedings of the 29th annual stanford child language research forum* (pp. 129–140). Stanford, CA: CSLI Publications.

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.

Brown, P. (2014). The interactional context of language learning in tzeltal. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), *Language in interaction: Studies in honor of eve v. clark* (pp. 51–82). Amsterdam, NL: John Benjamins.

Brown, P., & Casillas, M. (in press). Childrearing through social interaction on rossel island, png. In A. J. Fentiman & M. Goody (Eds.), *Esther goody revisited: Exploring the legacy of an original inter-disciplinarian* (pp. XX–XX). New York, NY: Berghahn.

Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. In N. J. Enfield, P. Kockelman, & J. Sidnell (Eds.), *Handbook of linguistic anthropology* (pp. 183–222). Cambridge, UK: Cambridge University Press.

Brown, R. (1977). Introduction. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and interaction* (pp. 1–30). Cambridge, UK: Cambridge University

Press.

Bruner, J. (1983). *Child's talk*. Oxford: Oxford University Press.

Butterworth, G. (2003). Pointing is the royal road to language for babies. In S. Kita (Ed.), *Pointing* (pp. 17–42). Psychology Press.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*(28), 11278–11283.

Casillas, M. (in progress). The trajectory of early verbal development for children acquiring tseltal. *XX*, *XX*, XX–XX.

Casillas, M., Bobb, S. C., & Clark, E. V. (2016). Turn taking, timing, and planning in early language acquisition. *Journal of Child Language*, *43*, 1310–1337.

Casillas, M., Brown, P., & Levinson, S. C. (2017). Casillas HomeBank corpus.

Casillas, M., Bunce, J., Soderstrom, M., Rosemberg, C., Migdalek, M., Alam, F., ... Garrison, H. (2017). Introduction: The aclew das template [training materials]. Retrieved from https://osf.io/aknjv/

Childers, J. B., & Tomasello, M. (2002). Two-year-olds learn novel nouns, verbs, and conventional actions from massed or distributed exposures. *Developmental Psychology*, *38*(6), 967–978. doi:10.1037//0012-1649.38.6.967

Chouinard, M. M., & Clark, E. V. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language*, *30*(3), 637–669.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *20*(4), 477–488. doi:10.1016/S0163-6383(97)90037-0

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, *7*(3), 157–170. doi:10.1111/lnc3.12015

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child*

Development, XX–XX. doi:10.1111/cdev.12974

Cristia, G., Alejandrina. (2018). Talker diarization in the wild: The case of child-centered daylong audio-recordings. *Proceedings of Interspeech 2018*. doi:doi:10.21437/Interspeech.2018-2078

Demuth, K., Moloi, F., & Machobane, M. (2010). 3-year-olds' comprehension, production, and generalization of Sesotho passives. *Cognition*, *115*(2), 238–251.

Dumay, N., & Gaskell, M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*(1), 35–39.

Egeren, L. A. van, Barratt, M. S., & Roach, M. A. (2001). Mother-infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology*, *37*(5), 684–697.

Estigarribia, B., & Clark, E. V. (2007). Getting and maintaining attention in talk to young children. *Journal of Child Language*, *34*(4), 799–814.

Ferrier, L. J. (1978). Some observations of error in context. In N. Waterson & C. Snow (Eds.), *The development of communication* (pp. 301–309). New York: Wiley Press.

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

Fortier, M. E., Kellier, D., Fernández Flecha, M., & Frank, M. C. (under review). Ad-hoc pragmatic implicatures among Shipibo-Konibo children in the Peruvian Amazon. doi:10.31234/osf.io/x7ad9

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

Frost, R. L. A., & Monaghan, P. (2017). Sleep-driven computations in speech processing. *PloS One*, *12*(1), e0169538. doi:doi:10.1371/journal.pone.0169538

Fusaroli, R., Razczaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy.

New Ideas in Psychology, *32*, 147–157. doi:10.1016/j.newideapsych.2013.03.005

Ganek, H., Smyth, R., Nixon, S., & Eriks-Brophy, A. (2018). Using the Language ENvironment analysis (LENA) system to investigate cultural differences in conversational turn count. *Journal of Speech, Language, and Hearing Research*, *61*, 2246–2258. doi:10.1044/2018_JSLHR-L-17-0370

Garcia, R., Roeser, J., & Höhle, B. (2018). Thematic role assignment in the L1 acquisition of Tagalog: Use of word order and morphosyntactic markers. *Language Acquisition*, *XX*, XX–XX. doi:10.1080/10489223.2018.1525613

Gaskins, S. (1996). How Mayan parental theories come into play. *Parents' Cultural Belief Systems: Their Origins, Expressions, and Consequences*, 345–363.

Gaskins, S. (1999). Children's daily lives in a Mayan village: A case study of culturally constructed roles and activities. In A. Göncü (Ed.), *Children's engagement in the world: Sociocultural perspectives* (pp. 25–60). Oxford: Berg.

Gaskins, S. (2000). Children's daily activities in a Mayan village: A culturally grounded description. *Cross-Cultural Research*, *34*(4), 375–389.

Gaskins, S. (2006). Cultural perspectives on infant–caregiver interaction. In N. J. Enfield & S. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 279–298). Oxford: Berg.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224.

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.

Gómez, R. L., Bootzin, R. R., & Nadel, L. (2006). Naps promote abstraction in language-learning infants. *Psychological Science*, *17*(8), 670–674.

Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, *32*(2), 83–92. doi:10.1177/1525740110367826

Hart, B., & Risley, T. R. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Paul H. Brookes Publishing.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Beyond WEIRD: Towards a broad-based behavioral science. *Behavioral and Brain Sciences*, *33*(2–3), 111–135.

Hernik, M., & Broesch, T. (2018). Infant gaze following depends on communicative signals: An eye-tracking study of 5- to 7-month-olds in Vanuatu. *Developmental Science*, *XX*, XX–XX. doi:10.1111/desc.12779

Hilbrink, E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother–infant interaction. *Frontiers in Psychology*, *6:1492*, 1–12.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., … Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071–1083. doi:10.1177/0956797615581493

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378.

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88.

Horváth, K., Liu, S., & Plunkett, K. (2016). A daytime nap facilitates generalization of word meanings in young toddlers. *Sleep*, *39*(1), 203–207.

Hupbach, A., Gomez, R. L., Bootzin, R. R., & Nadel, L. (2009). Nap-dependent learning in infants. *Developmental Science*, *12*(6), 1007–1012.

Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in spanish-learning children. *Developmental Science*, *11*(6), F31–F39.

doi:10.1111/j.1467-7687.2008.00768.x]

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*, 343–365.

Kuchirko, Y., Tafuro, L., & Tamis-LeMonda, C. S. (2018). Becoming a communicative partner: Infant contingent responsiveness to maternal language and gestures. *Infancy*, *23*(4), 558–576.

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

León, L. de. (1998). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, *8*(2), 131–161.

León, L. de. (2000). The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants. *Journal of Linguistic Anthropology*, *8*(2), 131–161.

León, L. de. (2011). Language socialization and multiparty participation frameworks. In A. Duranti, E. Ochs, & and B.B. Schieffelin (Eds.), *Handbook of language socialization* (pp. 81–111). Malden, MA: Wiley-Blackwell. doi:10.1002/9781444342901.ch4

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). New York, NY, US: Cambridge University Press. doi:10.1017/CBO9780511620690.005

Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language*, *24*(1), 187–219.

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., & De Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cognitive Science*, *36*(4), 698–713.

ManyBabies Collaborative. (2017). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*, *XX*, XX–XX. doi:10.31234/osf.io/s98ab

Marchman, V. A., Martínez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, *7*(2), 212–224.

Masataka, N. (2003). *The onset of language*. Cambridge University Press.

Mastin, J. D., & Vogt, P. (2016). Infant engagement and early vocabulary development: A naturalistic observation study of Mozambican infants from 1;1 to 2;1. *Journal of Child Language*, *43*(2), 235–264. doi:0.1017/S0305000915000148

Mirković, J., & Gaskell, M. G. (2016). Does sleep improve your grammar? Preferential consolidation of arbitrary components of new linguistic knowledge. *PloS One*, *11*(4), e0152489.

Morelli, G. A., Rogoff, B., Oppenheim, D., & Goldsmith, D. (1992). Cultural variation in infants' sleeping arrangements: Questions of independence. *Developmental Psychology*, *28*(4), 604.

Nelson, K. (1985). *Making sense: The acquisition of shared meaning*. New York: Academic Press.

Newport, E. L., Gleitman, H., & Gleitman, L. R. (1977). Mother, i'd rather do it myself: Some effects and non-effects of maternal speech style. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and interaction* (pp. 109–150). Cambridge, UK: Cambridge University Press.

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in developmental psychology: A call to action. *Journal of Experimental Child Psychology*, *162*, 31–38.

Oller, D. K. (1980). The emergence of the sounds of speech in infancy. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology, volume 1: Production* (pp. 93–112). New York, NY: Academic Press.

Oller, D. K., Griebel, U., & Warlaumont, A. S. (2016). Vocal development as a guide to modeling the evolution of language. *Topics in Cognitive Science*, *8*, 382–392.

doi:DOI:10.1111/tops.12198

Peter, M., Durrant, S., Bidgood, A., Pine, J., & Rowland, C. (in preparation). Individual differences in speed of language processing and its relationship with language development. *XX*, (XX), XX–XX.

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, *13*(1), 85–100.

Pye, C. (1992). The acquisition of ki'che'. In D. I. Slobin (Ed.), *The crosslinguistic study of language acquisition, volume 3* (pp. 221–308). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pye, C. (2017). *The comparative method of language acquisition research*. University of Chicago Press.

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/

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*, 880–891.

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*, 1008.

Rogoff, B., Mistry, J., Göncü, A., Mosier, C., Chavajay, P., & Heath, S. B. (1993). Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development*.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003). Firsthand learning through intent participation. *Annual Review of Psychology*, *54*(1), 175–203. doi:10.1146/annurev.psych.54.101601.145118

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science*, *29*(5), 700–710.

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205.

RStudio Team. (2016). *RStudio: Integrated development environment for r*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Salomo, D., & Liszkowski, U. (2013). Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child Development*, *84*(4), 1296–1307.

Scaff, C., Stieglitz, J., Casillas, M., & Cristia, A. (in preparation). Language input in a hunter-forager population: Estimations from daylong recordings.

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

Segal, J., & Newman, R. S. (2015). Infant preferences for structural and prosodic properties of infant-directed speech in the second year of life. *Infancy*, *20*(3), 339–351. doi:10.1111/infa.12077

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

Shneidman, L. A. (2010). *Language input and acquisition in a Mayan village* (PhD thesis). The University of Chicago.

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.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2012). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672–686.

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*. CRC Press: Boca Raton.

Snow, C. E., & Goldfield, B. A. (1983). Turn the page please: Situation-specific language acquisition. *Journal of Child Language*, *10*(3), 551–569.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review*, *27*(4), 501–532.

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS One*, *8*, e80646.

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, *20*(6), XX–XX. doi:doi.org/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). London, UK: Psychology Press.

VanDam, M., Warlaumont, A. S., Bergelson, E., Cristia, A., De Palma, P., & MacWhinney, B. (2016). HomeBank: An online repository of daylong child-centered audio recordings. *Seminars in Speech and Language*.

Vogt, E. Z. (1976). *Tortillas for the gods*. Cambridge, MA: Harvard University Press.

Vogt, P., Mastin, J. D., & Schots, D. M. A. (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.

Warlaumont, A. S., & Finnegan, M. K. (2016). Learning to produce syllabic speech sounds via reward-modulated neural plasticity. *PloS One*, *11*(1), e0145096.

Warlaumont, A. S., Richards, J. A., Gilkerson, J., & Oller, D. K. (2016). A social feedback loop for speech development and its reduction in Autism. *PloS One*, *11*(1), e0145096.

Wasow, T. (1997). Remarks on grammatical weight. *Language Variation and Change*, *9*(1), 81–105.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152.

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

Williams, S. E., & Horst, J. S. (2014). Goodnight book: Sleep consolidation improves word learning via storybooks. *Frontiers in Psychology*, *5*, 184.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: A professional framework for multimodality research. In *Proceedings of the fifth international conference on language resources and evaluation*.

Yurovsky, D. (2018). A communicative approach to early word learning. *New Ideas in Psychology*, *50*, 73–79.