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 for younger children, and high interactional peaks tended to occur in ~1-minute bursts of turn taking. These findings only partly 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 American English-learning children. Given the present 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 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 (e.g., Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015; Hoff, 2006). In several languages, child-directed speech (CDS[1]) has been demonstrated to be distinct from adult-directed speech (ADS) in that it is linguistically adapted for young listeners (Cristia, 2013; Soderstrom, 2007), interactionally rich (Bruner, 1983; Butterworth, 2003), preferred by infants (Cooper & Aslin, 1990; ManyBabies Collaborative, 2017; Segal & Newman, 2015), and appears to facilitate early word learning (Cartmill et al., 2013; Hoff, 2003; Hurtado, Marchman, & Fernald, 2008; Rowe, 2008; Weisleder & Fernald, 2013). However, ethnographic reports from a number of traditional, non-Western communities suggest that children easily acquire their language(s) even when they are only infrequently directly addressed (Brown, 2011). If so, frequent 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 vocal development of 10 Tseltal Mayan children growing up in a community where caregivers have previously been reported to infrequently directly address speech to young children (Brown, 1998b, 2011, 2014).

Child-directed speech

Prior work in Western contexts has shown that the amount of CDS children hear influences their language development; more CDS is associated with faster-growing receptive and productive vocabularies (e.g., Hart & Risley, 1995; Hoff, 2003; Ramírez-Esparza,

[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.

García-Sierra, & Kuhl, 2014; Shneidman & Goldin-Meadow, 2012), faster lexical retrieval (Hurtado et al., 2008; Weisleder & Fernald, 2013), and faster syntactic development (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010). Given that CDS is designed for a child hearer, it is more likely than ADS or other-directed speech to align with the child's attention, and may thereby comparatively facilitate early language development. There are, however, a few significant caveats to this body of work relating CDS quantity and language development.

First, while there is overwhelming evidence linking CDS quantity to vocabulary size, links to grammatical development are more scant (but see 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 it facilitates syntactic learning (see also Yurovsky, 2018). On the other hand, there is a wealth of evidence that syntactic knowledge is lexically specified (e.g., 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)—what is good for the lexicon may also be good for syntax. For now, a direct link between CDS and grammatical development still needs further exploration .

Second, most work on CDS quantity uses summary measures that average over the ebb and flow of the day (e.g., average proportion CDS). In reality, verbal behaviors are highly structured during interaction: while some occur at regular intervals , others occur in shorter, more intense bursts separated by long periods of inactivity . Infants' and adults' vocal behavior is clustered across multiple time scales of daylong recordings (Abney, Smith, & Yu, 2017) and noun and verb use is bursty across languages (Blasi, Schikowski, Moran, Pfeiler, & Stoll, in preparation). Even in experimental settings, two-year-olds have been shown to learn novel words better from a massed presentation of object labels versus a distributed one (Schwab & Lew-Williams, 2016; but see Ambridge, Theakston, Lieven, & Tomasello, 2006).

The existence of multi-scale temporal structure in children's language experience implies new roles for attention and memory in development; more work is needed to know 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 effects of CDS quantity (Brown & Gaskins, 2014; Henrich, Heine, & Norenzayan, 2010; Lieven, 1994; M. Nielsen, Haun, Kärtner, & Legare, 2017). While we gain valuable insight by looking at within-population variation (e.g., different socioeconomic groups), 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 (Brown & Gaskins, 2014; Gaskins, 2006; Lieven, 1994; Ochs & Schieffelin, 1984). 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 (Brown, 2011, 2014; 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 language development, partly due to a lack of directly comparable methods (Brown, 2014; Brown & Gaskins, 2014). If, however, children in these communities do acquire language without delay, despite infrequent CDS, developmental language science would need to re-work current ideas about the precise role of CDS quantity in early language development.

Developmental language research using modern psycholinguistic methods has supported the idea that children in some indigenous, non-Western communities hear very little CDS. Scaff, Cristia, and colleagues (2017; in preparation) estimate based on daylong recordings that Tsimane children, growing up in a forager-horticulturalist population in the Bolivian lowlands, hear maximally ~4.8 minutes of CDS per hour between 0;6 and 3;0 (Cristia et al., 2017; Scaff et al., in preparation; see also Mastin & Vogt, 2016; Vogt, Mastin,

& Schots, 2015).

Shneidman and colleagues (2010; 2012) analyzed speech from one-hour at-home video recordings of children between 1;0 and 3;0 in a Yucatec Mayan and a North American community. Their analyses yielded four main findings: compared to the American children, (a) 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 CDS proportion by 3;0, and (d) most of the added CDS came from other children (e.g., older siblings/cousins). The lexical diversity of the CDS Yucatec Mayan children heard at 24 months—particularly from adult speakers—predicted their vocabulary knowledge at 35 months, suggesting that CDS characteristics still played a role in that non-Western indigenous context.

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 use CDS until the children themselves begin to actively initiate verbal interactions (Brown, 2011, 2014). Nonetheless, Tseltal children develop language with no apparent delays (Brown, 2011, 2014; Brown & Gaskins, 2014; Liszkowski et al., 2012; see also Pye, 2017). We provide more details on the community and dataset in the Methods section. We analyze five 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 potentially overhear, (c) the rate of contingent responses to their vocalizations, (d) the rate of their contingent responses to others' vocalizations, and (e) the duration of their interactional dyadic sequences. We then also roughly estimate the number of minutes per day children spent in "high turn-taking" interaction and outline a basic trajectory for their early vocal development.

Based on prior work, we predicted that Tseltal Mayan children would be infrequently directly addressed, that the amount of CDS and contingent responses they heard would increase with age, that most CDS would come from other children, and that, despite this, their early vocal development would be 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

Corpus

The children in this dataset come from a small-scale, subsistence farming community in the highlands of Chiapas (Southern Mexico). The vast majority of children grow up speaking Tseltal monolingually at home. Nuclear families are typically organized into patrlinieal clusters of large, multi-generation households. 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 (Brown, 1998b, 2011, 2014). During their waking hours, infants are typically tied to their mother's back while she goes about her work for the day. When not on their mother's back, young children are often cared for by other family members, especially older siblings. Typically, CDS is limited until children themselves begin to initiate interactions, usually around age 1;0. Interactional exchanges, when they do occur, are often brief or non-verbal (e.g., object exchange routines) and take place within a multi-participant context (Brown, 2014). Interactions tend to focus on appropriate actions and responses (not on words and their meanings), and young children are socialized to attend the events taking place around them (see also de León, 2000, 2011; Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003). By age five, most children are competent speakers who engage in daily chores and the caregiving of their younger siblings. The Tseltal approach to caregiving is similar to that described for other Mayan communities (de León, 2011; Gaskins, 1996, 1999; e.g., León, 1998; Pye, 1986; Rogoff et al., 1993, 2003; Shneidman & Goldin-Meadow, 2012).

The current data come from the Casillas HomeBank Corpus (Casillas, Brown, & Levinson, 2017a; 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-Western communities: the Tseltal Mayan community described here and a Papua New Guinean community described elsewhere (Brown, 2011, 2014; Brown & Casillas, in press). This Tseltal corpus, primarily collected in 2015, includes recordings from 55 children born to 43 mothers. The participating families typically only had 2–3 children (median = 2; range = 1–9), due to the fact that they 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). Based on data from living children, we estimate that, 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). 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. Note that all ages should be taken with a grain of salt because documentation of birthdates in the village is not rigorous. Household size, defined in our dataset as 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% 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 school (12%) or university (2%; 1 mother); the remainder (23%) had no schooling or did not complete primary school. All fathers had finished primary school, with most completing secondary school (44%) or preparatory school (21%), and two completing a university-level training (5%). While 93% of the fathers grew up in the village where the recordings took place, only 53% of the mothers did because of the way clan membership influences marriage and land inheritance.

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 interactions over the course of a 9–11-hour period at home in which the experimenter was not present. Ambulatory children wore both devices at once (see Figure 1) while other 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 generate snapshot-linked audio.[2] We used these recordings to capture a wide range of the linguistic patterns children encounter as they participate in different activities over the course of their day (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2018; Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011; Catherine S. Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017).

Figure 1. The recording vest included an audio recorder in the front horizontal pocket and a camera with a fisheye lens on the shoulder strap.

[2]Post-processing scripts are available at https://github.com/marisacasillas/Weave.

Table 1

Demographic overview of the 10 children whose recordings we sampled.

Age	Sex	Mot age	Mot edu	People in house
0;01.25	M	26	none	8
0;03.18	M	22	preparatory	9
0;05.29	F	17	secondary	15
0;07.15	F	24	primary	9
0;10.21	M	24	secondary	5
1;02.10	M	21	none	9
1;10.03	F	31	preparatory	9
2;02.25	F	17	primary	5
2;08.05	F	28	secondary	5
3;00.02	M	28	primary	6

Data selection and annotation

We chose 10 children's recordings based on maximimal spread in child age (0;0–3;0), child sex, and maternal education (see Table 1; all had native Tseltal-speaking parents). We selected one hour's worth of non-overlapping clips from each recording in the following order: nine randomly selected 5-minute clips, five manually selected 1-minute top "turn-taking" clips, five manually selected 1-minute top "vocal activity" clips, and one, manually selected 5-minute extension of the best 1-minute clip (see Figure 2). We created these different subsamples to measure properties of (a) children's *average* language environments ("Random"), (b) their most *input-dense* language environments ("Turn-taking"), and (c) their most *mature vocal behavior* ("Vocal activity").

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. The first author then reviewed the list of candidate clips and chose the best five 1-minute samples for each of the two activity types. Note that, because the manually selected clips did not overlap with the initial "random" clip selection, the "true" peak turn-taking and vocal-activity clips for the day could have possibly occurred during the random clips. High-quality 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). High-quality vocal activity clips were defined as clips in which the target child produced the most and most diverse spontaneous (i.e., not imitative) vocalizations (see full instructions at https://git.io/fhdUm).

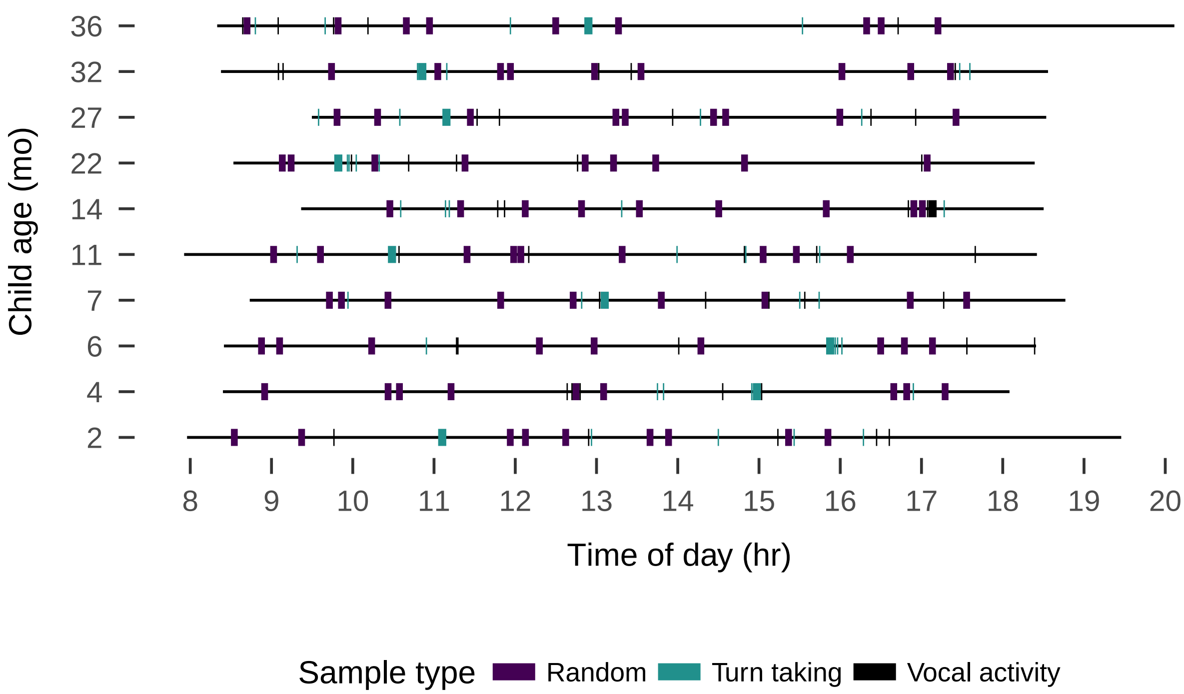

Figure 2. Recording duration (black line) and sampled clips (colored boxes) for each of the 10 recordings analyzed, sorted by child age.

The first author and a native speaker of Tseltal who knows all the recorded families personally jointly transcribed and annotated each clip in ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) using the ACLEW Annotation Scheme (Casillas et al., 2017b). Utterance-level annotations include: an orthographic transcription (Tseltal), a loose translation (Spanish), a vocal maturity rating for each target child utterance (non-linguistic/non-canonical babbling/canonical babbling/single words/multiple words), and the intended addressee type for all non-target-child utterances (target-child/other-child/adult/adult-and-child/animal/other-speaker-type). Intended addressee was detemined by using contextual and interactional information from the photos, audio, and preceding/following footage; utterances with no clear intended addressee were marked as "unsure". We annotated lexical utterances as single- or multi-word based on the word boundaries provided by the single native speaker who reviewed all transcription; Tseltal is a mildly polysynthetic language so, on average, there is more than one morpheme per word.[3]

Data analysis

In what follows we first describe Tseltal children's speech environments based on the nine randomly selected 5-minute clips from each child, including: the 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. We investigate the effects of child age, time of day, household size, and number of speakers present on each of these five measures. We next repeat these analyses, only this time looking at the high turn-taking clips. We then 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 a basic trajectory for early Tseltal vocal development.

[3] Documentation, including training materials can be found at https://osf.io/b2jep/wiki/home/.

Results

Statistical models

All analyses were conducted in R with generalized linear mixed-effects regressions using the glmmTMB package, and all plots were generated with ggplot2 (M. E. Brooks et al., 2017a; R Core Team, 2018; Wickham, 2009).[4] Notably, all five dependent measures are restricted to non-negative (0–infinity) values. This implicit boundary restriction at zero causes the distributional variance of our measures to become non-gaussian (i.e., with a long right tail). We handle this issue by using a negative binomial linking function in the regression, which estimates a dispersion parameter (in addition to the mean and variance) that allows the model to more closely fit our non-negative, overdispersed data (M. E. Brooks et al., 2017b; Smithson & Merkle, 2013). When, in addition to this, extra cases of zero were evident in the distribution (e.g., TCDS min/hr was zero because children were alone), we also added a zero-inflation model to the regression. A zero-inflation negative binomial regression creates two models: (a) a binary model to evaluate the likelihood of none vs. some presence of the variable (e.g., no vs. some 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 mixed-effects regressions 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 time of day at the start of the clip (as a factor; morning: up until 11:00; midday: 11:00–13:00; and afternoon: 13:00 onwards). We also added two-way interactions between child age and: (a) number of speakers present, (b) household size, and (c) time of day. We also included a random effect of child. For the zero-inflation models, we included number of speakers present. We only report significant effects in the main text; full model

[4]Data and analysis code can be found at https://github.com/marisacasillas/Tseltal-CLE.

outputs are available in the Supplementary Materials.

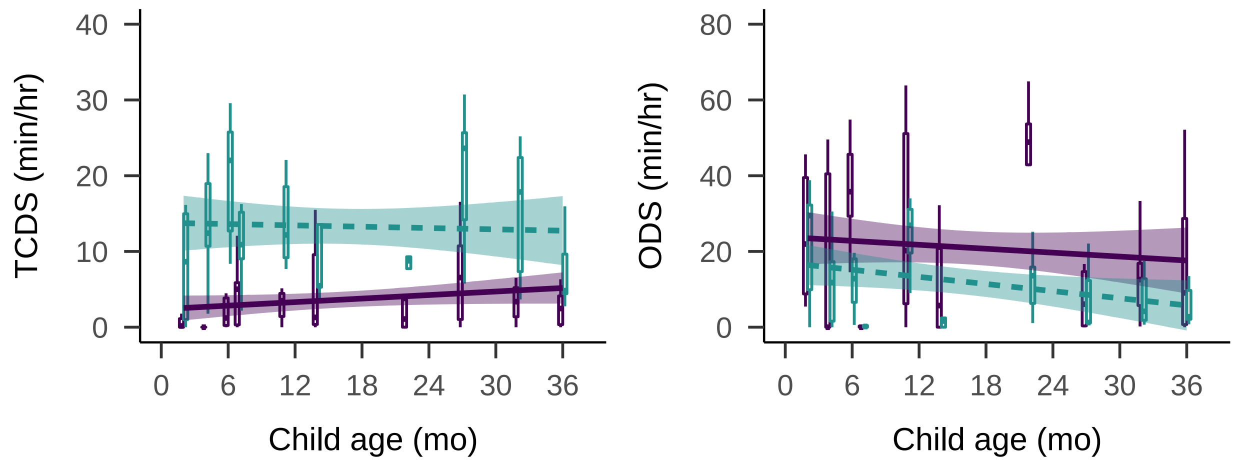

Figure 3. By-child estimates of minutes per hour of target-child-directed speech (left) and other-directed speech (right). Data are shown for the random (purple; solid) and turn taking (green; dashed) samples. Bands on the linear trends show 95% CIs.

Target-child-directed speech (TCDS)

The children in our sample 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 children (Shneidman & Goldin-Meadow, 2012), as illustrated in Figure 4 (see Scaff et al. (in preparation) for more detailed cross-language comparisons).[5] We modeled TCDS min/hr in the random clips with a zero-inflated negative binomial regression. The rate of TCDS in the randomly sampled clips was primarily affected by factors relating to the time of day (see Figure 5). The count model showed that the children were more likely to hear TCDS in the mornings than around midday (B = 0.82, SD = 0.40, z = 2.06, p = 0.04), with no difference between morning and afternoon (p = 0.29) or midday and afternoon (p = 0.19) TCDS rates. Time of day effects varied by age: older children were significantly more likely to hear TCDS at midday (B = 0.73, SD = 0.36, z =

[5] We convert Shneidman (2010)'s utterance/hr estimates to min/hr with the median Tseltal utterance duration for non-target child speakers: (1029ms) because Yucatec and Tseltal are related languages.

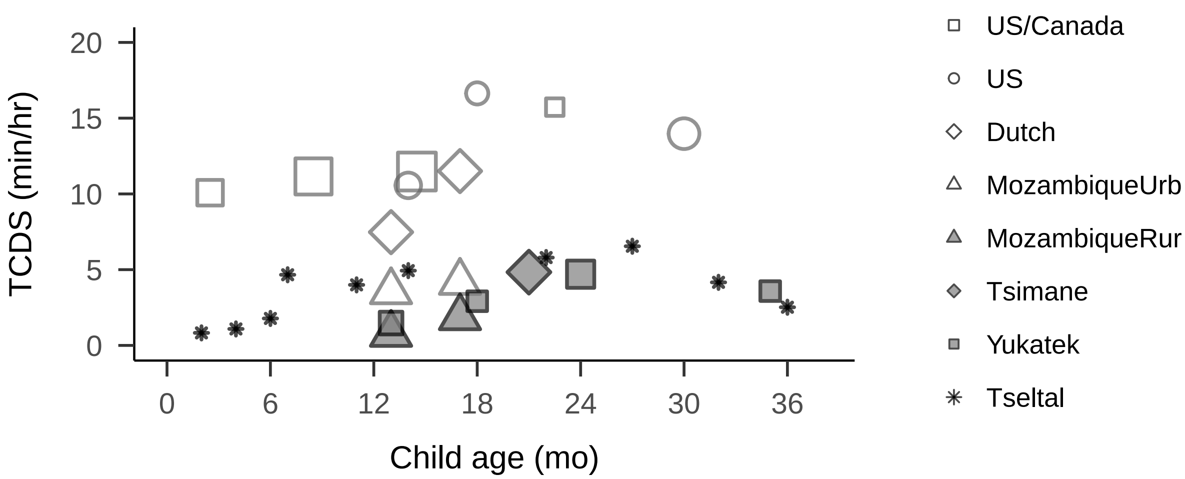

Figure 4. TCDS rates reported from at-home recordings in different populations, including both urban (empty shape) and rural/indigenous (filled shape) samples. Each point shows the average TCDS rate at the indicated age, while size indicates the number of children sampled (range: 1–26). Data sources: Bergelson et al. (2019) US/Canada; Shneidman (2010) US and Yucatec; Vogt, Mastin, and Schots (2015) Dutch, Mozambique urban and rural; Scaff et al. (in preparation) Tsimane.

2.04, p = 0.04) and marginally more likely to hear it in the morning (B = 0.46, SD = 0.28, z = 1.65, p = 0.10) compared to the afternoons. Older target children were also significantly more likely to hear TCDS when more speakers were present, compared to younger children (B = 0.61, SD = 0.20, z = 3.06, p = 0.00). There were no significant effects of target child age or household size, and no significant effects in the zero-inflation model.

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 of TCDS from children with target child age (Spearman's *rho* = -0.29; *p* = 0.42).

Other-directed speech (ODS)

Children heard an average of 21.05 minutes of ODS per hour in the random sample (median = 17.80; range = 3.57–42.80): that is, nearly six times as much speech as was

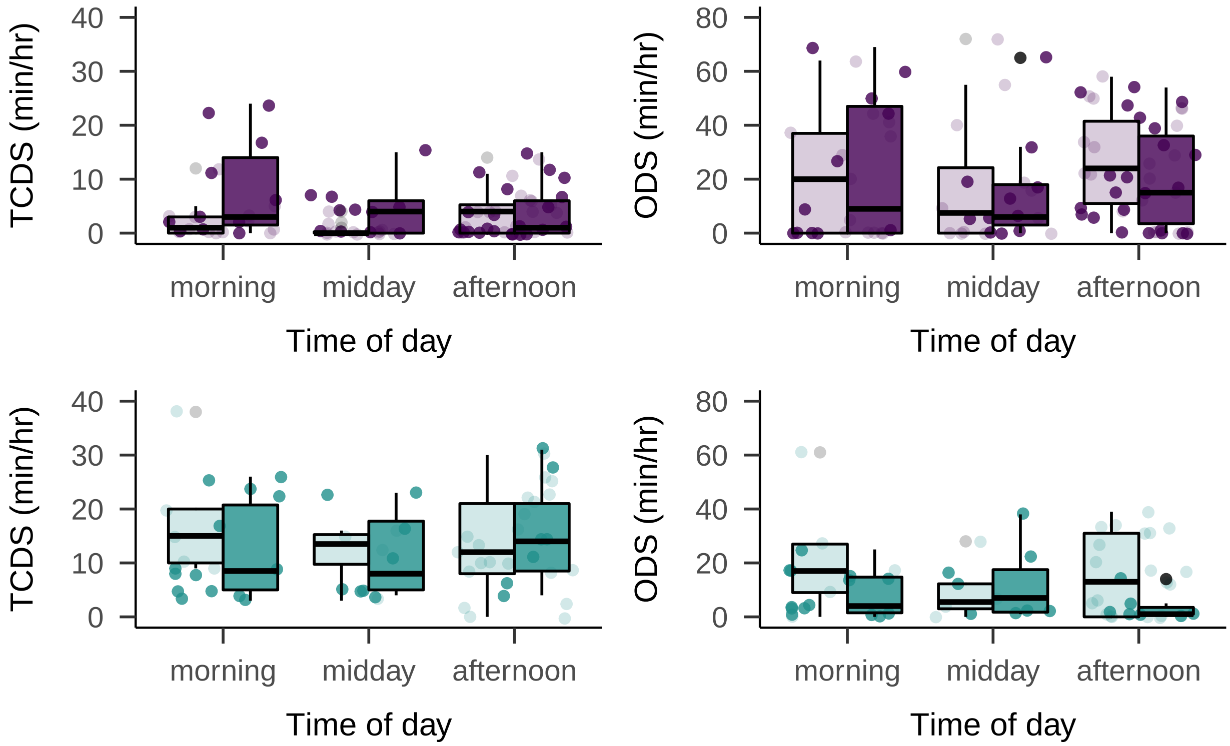

Figure 5. TCDS (left) and ODS (right) rates heard at different times of day in the random (top) and turn-taking (bottom) clip samples by children age 1;0 and younger (light) and 1;0 and older (dark).

directed to them, on average. We modeled ODS min/hr in the random clips with a zero-inflated negative binomial regression. The count model of ODS in the randomly selected clips revealed that the presence of more speakers was strongly associated with more ODS (B = 0.65, SD = 0.09, z = 7.32, $p < 0.0001$). There were an average of 3.44 speakers present other than the target child in the randomly selected clips (median = 3; range = 0–10), more than half of whom were typically adults. Older target children were also significantly less likely to hear ODS in large households, compared to younger children (B = 0.32, SD = 0.13, z = 2.41, p = 0.02).

Like TCDS, ODS was also strongly affected by time of day (see Figure 5). Compared to midday, target children were significantly more likely to hear ODS in the mornings (B = 0.36, SD = 0.17, z = 2.09, p = 0.04) and marginally more likely to hear it in the afternoons (B =

0.29, SD = 0.16, z = 1.89, p = 0.06), with no significant difference between ODS rates in the mornings and afternoons (p = 0.63). As before, ODS rate varied across the day by target child age: older children were significantly more likely to hear ODS in the afternoon than at midday (B = 0.38, SD = 0.17, z = 2.21, p = 0.03), with no significant differences between afternoon and morning (p = 0.10) or midday and morning (p = 0.63). There were no other significant effects on ODS rate, and no significant effects in the zero-inflation models.

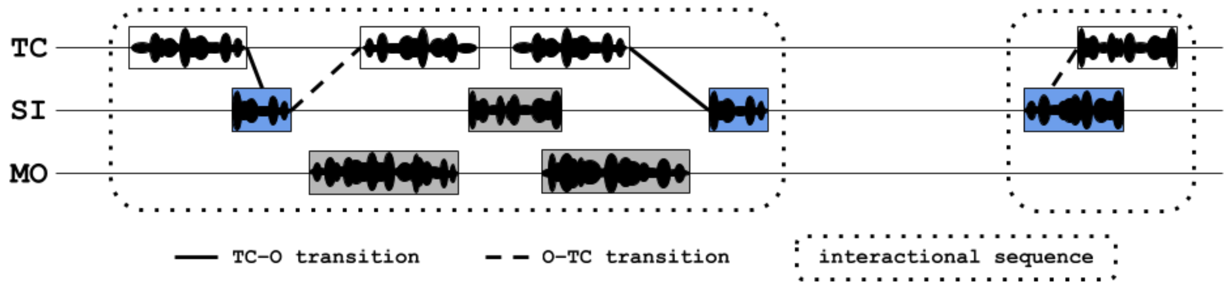

Figure 6. Illustration of an annotated audio clip including the target child (TC), an older sister (SI), and mother (MO). Transitions between the target child and others are marked with solid and dashed lines. Interactional sequences are boxed in with dotted lines. Box color indicates TCDS (blue) and ODS (light gray).

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

Contingent responses by or to the target child are likely to occur at moments in which the child and another speaker are attentionally aligned; the rate at which these responses is an index of the frequency of these joint moments of high-quality linguistic evidence. We measured two types of contingent responses: target-child-to-other and other-to-target-child. We detect these contingent turn transitions based on utterance onset/offset times and the annotations of intended addressee for each non-target-child utterance (Figure 6). If a child's vocalization is followed by a target-child-directed utterance within -1000msec to 2000msec after its end (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 (i.e., a target-child-directed utterance followed by a target child vocalization with the same timing restrictions). In our analysis, each target child vocalization can have maximally 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). These timing restrictions are broadly based on prior studies of infant and young children's spontaneous turn taking (e.g., M. H. Bornstein, Putnick, Cote, Haynes, & Suwalsky, 2015; T. Broesch, Rochat, Olah, Broesch, & Henrich, 2016; Casillas et al., 2016; Hilbrink et al., 2015).

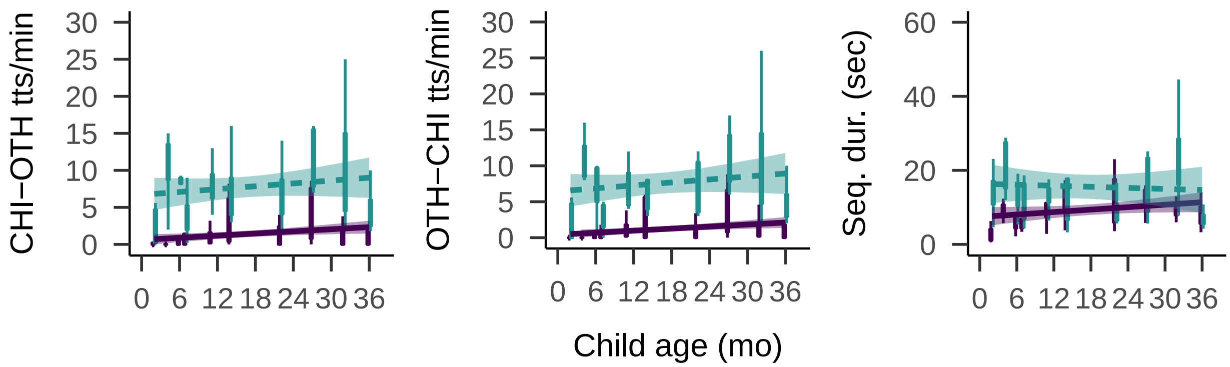

Figure 7. By-child estimates of target-child-to-other contingent responses (left), other-to-target-child contingent responses (middle), and the average duration of interactional sequences (right). Each boxplot represents the variance across clips within the random (dark purple; solid) or turn taking (light green; dashed) samples for each child. Bands on the 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; Figure 7). We modeled TC–O transtions per minute in the random clips with a zero-inflated negative binomial regression. The rate at which target children heard contingent responses from others was primarily influenced by factors relating to target child age. The rate of contingent responses to target child vocalizations varied across the day by target child age: older children heard significantly more contingent responses around midday (B = 1.08, SD = 0.44, z = 2.44, p = 0.01) and in the morning (B = 0.94, SD = 0.37, z = 2.51, p = 0.01), compared to the afternoon, with no significant difference between morning and midday (p = 0.77). Older target children also heard significantly more contingent responses then younger ones when there were more speakers present (B = 0.56, SD = 0.23, z = 2.48, p = 0.01).There were no further significant effects in the count or zero-inflation models.

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

The children in our sample responded contingently to others' target-child vocalizations at an average rate of 1.17 transitions per minute (median = 0.20; range = 0–8.80; Figure 7). We modeled O–TC transtions per minute in the random clips with a zero-inflated negative binomial regression, excluding by-child intercepts of time of day to achieve convergence. The rate at which target children responded contingently to others (O–TC turn transitions per minute) was similarly influenced by child age and time of day: older children responded contingently to others' utterances significantly more often around midday (B = 1.46, SD = 0.46, z = 3.13, p = 0.00) and in the morning (B = 1.33, SD = 0.42, z = 3.19, p = 0.00), compared to the afternoon, with no significant difference between morning and midday (p = 0.81). Overall, older children responded to others' utterances at a marginally higher rate (B = 1.14, SD = 0.66, z = 1.74, p = 0.08). Older target children also gave significantly more contingent responses then younger ones when there were more speakers present (B = 0.52, SD = 0.22, z = 2.30, p = 0.02).There were no further significant effects in the count or zero-inflation models.

Sequence duration

We defined sequences of interaction as periods of contingent turn taking with at least one target child vocalization and one target-child-directed prompt or response from another speaker. To detect sequences of interaction, we used the same mechanism as before to detect contingent TC–O and O–TC transitions, but also allowed for speakers to continue with multiple vocalizations in a row (e.g., TC–O–O–TC–OTH; Figure 6). We bounded sequences by the earliest and latest vocalization for which there is no contingent prompt or response, respectively. In our analysis, each target child vocalization can only appear in one sequence. We modeled these sequence durations in the random clips with negative binomial regression alone (i.e., with no zero-inflation model). 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; Figure 7). 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 &ge 0.21).

Speech environment characteristics at peak interaction

As expected, the high-quality turn taking clips featured a much higher rate of contingent turn transitions: the average TC–O transition rate was 7.73 transitions per minute (~5.5x the random sample rate; median = 7.80; range = 0–25) and the average O–TC rate was 7.56 transitions per minute (~6.5x the random sample rate; median = 6.20; range = 0–26). The interactional sequences were also slightly longer on average: 12.27 seconds (~1.2x the random sample rate; median = 8.10; range = 0.55–61.22). Crucially, children also heard much more TCDS in the turn-taking clips—13.28 min/hr (nearly 4x the random sample rate; median = 13.65; range = 7.32–20.19)—while also hearing less ODS—11.93 min/hr (nearly half the random sample rate; median = 10.18; range = 1.37–24.42).

We analyzed each of these speech environment measures with parallel models to those used for the random sample, though this time we did not include a zero-inflation model for TCDS, TC–O, and O–TC rates—given the criteria for selecting a turn-taking clip, the child is never alone, and so there are no extra-zero cases. As a whole, children's speech environments appeared very different when viewed through the lens of interactional peaks rather than randomly sampled clips (see Figures 3, 5, and 7), particularly with respect to time of day effects and the number of speakers present, which we focus on here. Full model outputs are available in the Supplementary Materials.

Time-of-day effects were consistently weaker in the turn-taking sample. TCDS rates showed no time-of-day effects and no interaction between time-of-day and age, and ODS rates *did* show a dip, but later in the day than what we saw in the random sample (i.e., afternoon, not midday; afternoon-vs.-midday: B = 0.70, SD = 0.29, z = 2.39, p = 0.02, afternoon-vs.morning: B = 0.72, SD = 0.25, z = 2.91, p = 0.00). Older children were also significantly more likely to hear ODS around midday compared to the morning (B = -0.56, SD = 0.28, z = -1.99, p = 0.05), but heard significantly less ODS overall than younger children (B = -0.45, SD = 0.21, z = -2.19, p = 0.03). There were no time-of-day effects at all on contingent response rates (TC–O and O–TC) in the turn-taking sample. However, running counter to this overall pattern, sequence duration in in the turn-taking sample did show significant time-of-day effects not found in the random sample: sequences were significantly longer in the afternoon compared to morning and midday (afternoon-vs.-morning: B = -0.32, SD = 0.15, z = -2.12, p = 0.03; afternoon-vs.-midday: B = 0.38, SD = 0.15, z = 2.61, p = 0.01).

Effects relating to the number of speakers present were also somewhat weaker in the turn-taking sample, though inconsistently. In the random sample, older children heard more TCDS, and participated in more contingent responses (both TC–O and O–TC) when more speakers were present, but this effect did not hold up in the turn-taking sample on any of the three measures. On the other hand, the number of speakers present was associated with significantly more ODS in both the random and turn-taking samples (random sample: B = 0.71, SD = 0.11, z = 6.63, $p < 0.001$), suggesting that the number of speakers is a robust predictor of ODS quantity across different contexts.

Peak minutes in the day. Having now established the interactional timing characteristics of the "high" turn-taking clips, we looked for similarly temporally-contingent 1-minute sections of interaction 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.) of each random clip and recorded the observed turn-transition rate. We then compared the resulting 1-minute transition rates to those typical for the high turn taking sample.

Only 6 of the 10 children showed at least one minute of their random sample that equalled or exceeded the average contingent transition rate (12.89 transitions/min), and 7 of the 10 children showed at least one minute equalling or exceeding their *own* average turn transition rate from their turn-taking sample. Across the 6 children who did show turn-taking "peaks" in their random data, peak periods were relatively long, at an average of 88.95 seconds (median = 90.67 seconds; range = 71–103 seconds). Assuming approximately 14 waking hours (Hart & Risley, 1995), we therefore very roughly estimate that the Tseltal child spends an average of 116.85 minutes (1.95 hours) in high turn-taking, dyadic interaction during their day. Importantly, 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 489.69 minutes (8.16 hours) in our sample. Much more data, particularly from other Tseltal children in this age range, is required to get a stable estimate of peak minutes in the day.

Vocal maturity

Tseltal 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 post-industrial, typically Western populations has shown that, typically, children begin producing non-canonical babbles around 0;2, with canonical babbling appearing sometime around 0;7, first words around 1;0, with first multi-word utterances appearing just after 1;6 (Fine & Lieven, 1993; Frank et al., in preparation; P. K. Kuhl, 2004; Slobin, 1970; Tomasello & Brooks, 1999; Warlaumont & Finnegan, 2016; Warlaumont, Richards, Gilkerson, & Oller, 2016). These benchmarks are mirrored in the Tseltal children's vocalizations, which are summarized in Figure 9 based on all annotated vocalizations from the random, turn-taking, and high vocal activity samples (N = 4725 vocalizations). 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 observed for every child from age 11;0 and older. Multi-word utterances appear in all recordings at 1;2 and later, making up 45% of the oldest child's (3;0) vocalizations.

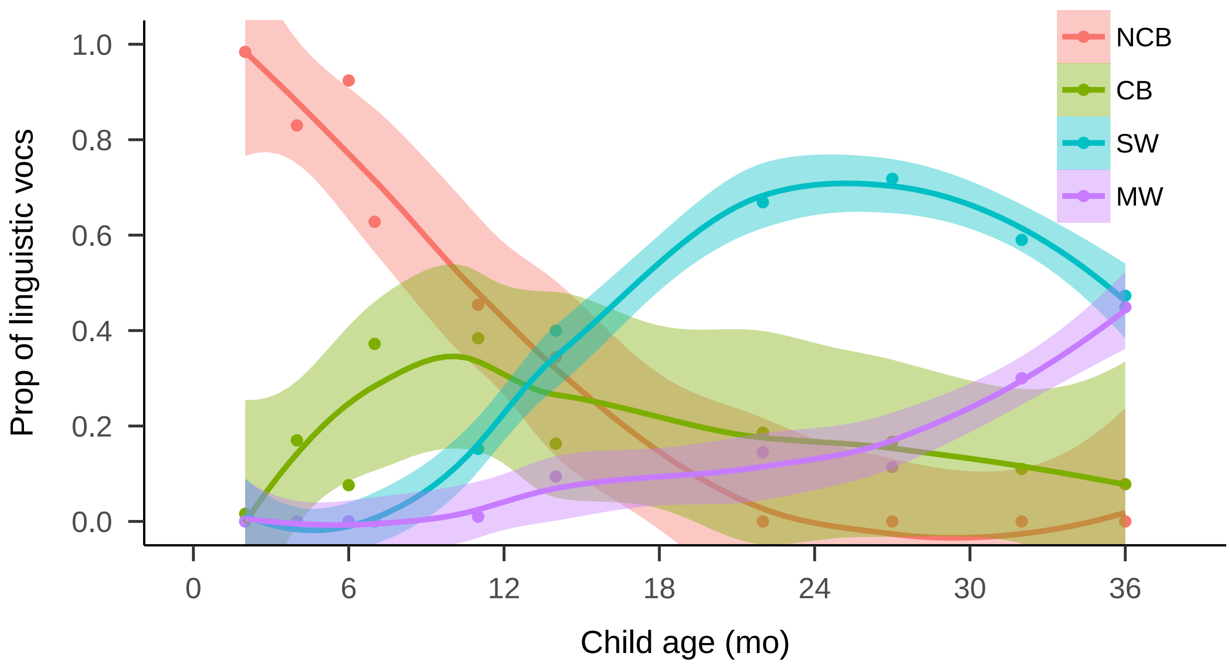

Figure 8. Proportion of vocalization types used by children across age (NCB = Non-canonical babble, CB = Canonical babble, SW = single word utterance, MW = multi-word utterance).

These data are also consistent with usage statistics of speech-like vocalizations by English-acquiring infants (Oller, 1980; Warlaumont & Finnegan, 2016; Warlaumont et al., 2016). Between 2 and 14 months, these 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 around age 1;0 in a variable SES sample (~60%).[6]

[6] Speech-like vs. non-speech-like comparisons are limited to age 1;6 in the ACLEW Annotation Scheme.

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. We found that, over the course of a whole recording day, children were directly spoken to infrequently, but were often surrounded by a large quantity of other-directed speech. The quantity of speech and contingent responding in their environment appeared stable across the first three years of life but was, over the course of individual recordings, influenced by the time of day and number of speakers present. When we focused on the speech quantity characteristics of interactional peaks, we saw that many children received most of their direct speech for the day in short, high-intensity bursts of turn-taking. Speech quantities during these interactional peaks were less affected by time of day and number of speakers present, suggesting that there is at least one current of stable, reliable, high-engagement linguistic interaction available to Tseltal children in the first few years of life. The children's spontaneous vocal productions demonstrated that, despite the relative infrequency of directly addressed speech, Tseltal children's early vocal development was on-par with norms based on post-industrial, typically Western populations. These findings only partly replicate estimates of child language input and development in previous work on Yucatec Mayan and Tseltal Mayan communities (Yucatec: Shneidman et al., 2012; Shneidman, 2010; Tseltal: P. Brown, 1997, 1998, 2014, 2011), and bring new questions to light regarding the distribution and sources of child-directed speech in Mayan children's speech environments.

Robust learning with less child-directed speech

The bulk of our analyses were aimed at understanding how much speech Tseltal children hear: we wanted to know how often they were directly spoken to and how often they might have been able to listen to speech directed to others. Consistent with 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., 2019), but is comparable to that for Tsimane children (Scaff et al., in preparation) and Yucatec Mayan children (Shneidman, 2010; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2012) in a similar age range. Meanwhile, we found that the children had an enormous quantity of other-directed speech in their environment, averaging 21.05 minutes per hour in the random sample, which is more than has been previously reported for other cultural settings (e.g., Bergelson et al., 2019; Scaff et al., in preparation).

We also created two novel interactive measures to describe how often children were directly engaged with an interlocutor, either as a responder or as an addressee being responded to. Children's vocalizations were responded to at a rate of 1.38 transitions per minute and children responded to others' child-directed vocalizations at a rate of 1.17 transitions per minute. Contingent interaction (and the joint attention that likely accompanies it) is a fertile context for language learning because the participants' coordinated attentional states decrease referential uncertainty, increase the chances of dynamic feedback, and can spur further interactions (M. H. Bornstein et al., 2015; T. Broesch et al., 2016; 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, contingent responses are rare across the day—more rare than CDS in general. The rarity of contingent responses may be due to the fact that the children did not vocalize very often: preliminary analyses showed that they only produced an average of 472.50 (median = 453; range = 245–753) vocalizations during their full one hour of annotated audio (including the high vocal activity minutes), and much of which was crying and laughter. Interestingly, although interactional sequences were fairly long when they occurred (mean = 10.13 seconds), children tended to only vocalize 3.75 times per sequence. In other words, many of children's dyadic interactions—sometimes containing the bulk of their directed speech for the day—are marked by longer streams of speech from an interlocutor, interspersed with only occasional responses from the child.

In sum, our daylong recording 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 monotonically feeds language development (such that more input begets more (advanced) output), then the estimates presented here would lead us to expect Tseltal to be delayed in their language development. However, our analyses suggest that Tseltal children, though they do not vocalize often, demonstrate vocal maturity comparable to children from societies in which CDS is known to be more frequent (Fine & Lieven, 1993; Frank et al., in preparation; P. K. Kuhl, 2004; Slobin, 1970; Tomasello & Brooks, 1999; Warlaumont & Finnegan, 2016; Warlaumont et al., 2016). How might Tseltal children manage this feat?

Other-directed speech. One proposal is that Mayan children become experts at learning from observation during their daily interactions (de León, 2011; Rogoff et al., 2003; Shneidman, 2010; Shneidman et al., 2012). In the randomly selected clips, children were within hearing distance of other-directed speech for an average of 21.05 minutes per hour. This large quantity of ODS is likely due to the fact that Tseltal children tend to live in households with more people compared to North American children (Bergelson et al., 2018). 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). The presence of more speakers had no overall impact on the quantity of TCDS children experienced, but older children were more likely than younger children to hear TCDS when more speakers were present. These findings ring true with Brown's (2011, 2014) claim that Tseltal is a non-child-centric language community; the presence of more people primarily increases talk amongst the other speakers but, as children become more sophisticated language users, they are more likely to participate in talk. However, given that an increase in the number of speakers is also likely associated with an increase in the amount of overlapping speech (Cristia, Ganesh, Casillas, & Ganapathy, 2018), we suggest that attention to ODS is unlikely to be the primary mechanism underlying the robustness of early vocal development in Tseltal. Furthermore, just because speech is hearable does not mean the children are attending to it. Follow-up work on the role of ODS in language development would need better define what constitutes likely "listened to" speech by the child.

Increased CDS with age. Another possibility is that speakers more frequently address children who are more communicatively competent (i.e., increased TCDS with age, e.g., Warlaumont et al., 2016). In her longitudinal studies of Yucatec Mayan children, Shneidman (2012) found that TCDS increased significantly with age, from 55 utterances in an hour to 209 between 13 and 35 months. However, most of the increase came from other children speaking to the target child, a finding 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, there was no evidence for an overall increase in TCDS with age, neither from adult speakers nor from child speakers. This non-increase in TCDS with age may be due to the fact TCDS from other children was overall infrequent in our data (cf. Shneidman & Goldin-Meadow, 2012), possibly because: (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 (as used in Shneidman & Goldin-Meadow, 2012). That aside, we conclude from these findings, that an increase in TCDS with age is also unlikely to explain the robust pattern of Tseltal vocal development.

Learning during interactional bursts. A third possibility is that children learn effectively from short, routine language encounters. Bursty input appears to be the norm across a number of linguistic and interactive scales (e.g., Abney et al., 2017; Blasi et al., in preparation), 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: (a) they experience most language input during routine activities and (b) they consolidate their language experiences 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: children were most likely to encounter speech, particularly directed, contingent speech in the mornings and late afternoons, compared to midday. Older children, who are less often carried and were therefore more free to seek out interactions, showed these time of day effects most strongly, receiving more TCDS when more speakers were present before and after the household disperses for farming work. 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). These time of day effects likely arise from the activities that typically occur in the mornings and late afternoons—meal preparation and dining in particular—while napping could contribute to the midday dip (Soderstrom & Wittebolle, 2013). That said, in data from North American children, 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 tracking 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 a role for action routines that help children optimally extract information about what they will encounter and what they are expected to do in response, even over short periods (see, e.g., Bruner, 1983; Snow & Goldfield, 1983; Catherine S Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2018).

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 (Frost & Monaghan, 2017; Mirković & Gaskell, 2016) and children (Gómez, Bootzin, & Nadel, 2006; Horváth, Liu, & Plunkett, 2016; Hupbach, Gomez, Bootzin, & Nadel, 2009), 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 (e.g., de León, 2000; Pye, 1992). 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;0 sleep 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 (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 (Brown, 1997, 1998b, 1998a, 2011, 2014; Brown & Gaskins, 2014) is needed to establish trajectory of early language development in Tseltal (Casillas et al., 2017a). 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 the current data and analysis scripts can be found: ADD-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., 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

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. (2018). 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. (2019). What do north american babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724. doi:doi:10.1111/desc.12724

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.

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. (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.

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., 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. (2017a). Casillas HomeBank corpus.

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

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, A., Ganesh, S., Casillas, M., & Ganapathy, S. (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

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

de León, L. (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

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.

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

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.

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.

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.

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.

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.

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.

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.

Ochs, E., & Schieffelin, B. (1984). Language acquisition and socialization: Three developmental stories and their implications. In R. A. Schweder & R. A. LeVine (Eds.), *Culture theory: Essays on mind, self, and emotion* (pp. 276–322). Cambridge University Press.

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.

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.

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

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

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., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2018). Routine language: Speech directed to infants during home activities. *Child Development*.

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, 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.

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

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.