A cross-linguistic examination of young children’s everyday language experiences

John Bunce1,2, Melanie Soderstrom2, Elika Bergelson3, Celia Rosemberg4, Alejandra Stein4, Florencia Alam4, Maia Julieta Migdalek4, & Marisa Casillas5,6

1 California State University, East Bay

2 University of Manitoba

3 Harvard University

4 Centro Interdisciplinario de Investigaciones en Psicología Matemática y Experimental - CONICET

5 University of Chicago

6 Max Planck Institute for Psycholinguistics

Author note

This research was supported by the Social Sciences and Humanities Research Council of Canada (435-2015-0628, 869-2016-0003) and by the Natural Sciences and Engineering Research Council of Canada (501769-2016-RGPDD) to Melanie Soderstrom; by the National Endowment for the Humanities (HJ-253479-17), National Institutes of Health Grant DP5-OD019812, and National Science Foundation BCS-1844710 to Elika Bergelson; a CONICET grant, PIP 80/2015, and a MINCyT grant, PICT 3327/2014 to Celia Rosemberg; and an NWO Veni Innovational Scheme Grant (275-89-033) to Marisa Casillas. We thank Anne Warlaumont and Caroline Rowland for contributing their datasets to this project and for helpful feedback on this manuscript. Finally, we thank the families who participated in the recordings that made this research possible.

Correspondence concerning this article should be addressed to John Bunce, California State University East Bay, Department of Human Development and Women’s Studies, 25800 Carlos Bee Blvd, Hayward, CA 94542. E-mail: [john.bunce@csueastbay.edu](mailto:john.bunce@csueastbay.edu)

Abstract

We present an exploratory cross-linguistic analysis of the quantity of target-child-directed speech and adult-directed speech in North American English (US & Canadian), United Kingdom English, Argentinian Spanish, Tseltal (Tenejapa, Mayan), and Yélî Dnye (Rossel Island, Papuan), using annotations from 69 children aged 2–36 months. Using a novel methodological approach, our cross-linguistic and cross-cultural findings support prior work suggesting that target-child-directed speech quantities are stable across early development, while adult-directed speech decreases. A preponderance of speech from women was found to a similar degree across groups, with less target-child-directed speech from men and children in the North American samples than elsewhere. Consistently across groups, children also heard more adult-directed than target-child-directed speech. Finally, the numbers of talkers present in any given clip strongly impacted children’s moment-to-moment input quantities. These findings illustrate how the structure of home life impacts patterns of early language exposure across diverse developmental contexts.

*Keywords:* child-directed speech; cross-linguistic; cross-cultural; language development; linguistic input; addressee

*Word count:*

A cross-linguistic examination of young children’s everyday language experiences

Across cultures, children’s early language experiences vary substantially with respect to who talks to them, what is talked about, and what the children themselves are expected to contribute (e.g., Brown, 2011; Brown & Gaskins, 2014; Casillas, Brown, & Levinson, 2020; de León, 2011; Demuth & Mputhi, 1979; Gaskins, 2006; Ochs & Schieffelin, 1984; Pye, 1986; Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003; Shneidman & Goldin-Meadow, 2012; Vogt, Mastin, & Schots, 2015). For example, home pedagogical techniques, such as caregiver use of rhetorical questions and directly addressed instructions, are more common in some cultural contexts than others (e.g., US versus Mayan groups, see e.g., Gaskins (1996); Rogoff et al. (2003); Shneidman, Gaskins, and Woodward (2016)).

Research today, primarily revolving around urban, Western contexts, situates child-directed speech (CDS)—and more specifically interactive speech produced by adult caregivers—as a keystone for early language development (e.g., Cartmill et al., 2013; Hoff, 2003; Ramírez-Esparza, García-Sierra, & Kuhl, 2017a, 2017b, 2014). Recent findings converge on the idea that so-called “high quality” (interactive, one-on-one) CDS is a consistent and robust predictor of children’s growing vocabulary (e.g., Ramírez-Esparza et al., 2014; Rowe, 2008). However, the focus of most research so far using CDS to predict vocabulary outcomes reflects the political and economic priorities of growing, urban societies, especially their need for a unified and literate workforce. These priorities may not generalize across other cultural contexts, where other language phenomena (e.g., grammatical systems or rhetorical practices) may prove more relevant (Ochs & Kremer-Sadlik, 2020; Sperry, Sperry, & Miller, 2019).

Recent cross-linguistic and cross-cultural work on typically developing children supports the idea that there is, in fact, significant natural variation in children’s exposure to CDS. For example, Shneidman (2010; Shneidman & Goldin-Meadow, 2012) found an almost ten-fold difference in the proportion of child-directed speech in the linguistic environments of Chicago (US) and Yucatec Mayan (Mexico) children before age three. Scaff, Cristia, and colleagues find that Tsimane’-acquiring children (Bolivia) are directly spoken to infrequently, with recent estimates as low as approximately one half minute per hour (Cristia, Ganesh, Casillas, & Ganapathy, 2018; Scaff, Casillas, Stieglitz, & Cristia, under review). Casillas and colleagues (Casillas et al., 2020, 2021) found surprisingly similar, but relatively infrequent rates of directed input in two rural populations with substantially different approaches to child language socialization (Tseltal Mayan (Mexico) and Rossel Island Papuan (Papua New Guinea)). Cristia (2023) pulls these findings and more together in a systematic review, highlighting a consistent difference in higher versus lower input rates between urban and rural societies, respectively. A recurrent theme across much of this work examining CDS in rural and developing populations has been the role of input from other children (e.g., siblings, cousins, and other peers; see also Loukatou, Scaff, Demuth, Cristia, and Havron (2022) and Cristia, Gautheron, and Colleran (2023) for perspectives from Sesotho and multilingual children in Vanuatu).

It is not yet understood how differences in CDS exposure play a role in how children process or learn language in their first few years. The emerging evidence on this topic in a cross-cultural context is complex. For example, Ramirez-Esparza and colleagues (2017b) found that CDS heard in a group context (as opposed to one-on-one interactions) was related to vocabulary development in US Spanish-English bilinguals but not monolinguals from the same population. Consistent with this view, studies of populations where caregiver CDS is relatively rare have found that young children meet language development milestones at roughly the same rate as children growing up in contexts where adult CDS is very common (Casillas et al., 2020, 2021), though lexical development may be more sensitive (Ramírez-Esparza et al., 2017b; Shneidman & Goldin-Meadow, 2012).

In short, there is a great deal yet to learn about how language learning is supported by CDS *and* other sources of input. These other sources may include adult conversations that the infant observes (passively or actively), CDS produced by other children, and multiparty interactions (Casillas et al., 2020, 2021; Cristia et al., 2018, 2023; Loukatou et al., 2022; Scaff et al., under review).

The present study takes a first step toward describing multiple sources of input—not just CDS—across a linguistically and culturally diverse sample of young children. Specifically, we examine how child age and cultural group influence the quantities of directly addressed and overhearable adult speech that children encounter in five culturally distinct settings. Before we dive in to the methods and findings, we will set up the present study with three brief overviews of relevant work on measuring children’s linguistic input: first we define the constructs of child-directed and adult-directed speech as we use them here, then we review the major factors known to influence the quantities of each in children’s language environments, and we then describe prior approaches to estimating each from daylong audio recordings.

## What counts as “child-directed” input?

A great deal of prior work has contrasted child- and adult-directed speech, but what gets counted as “child-directed” varies from study to study. There are two basic approaches. First, these two terms (“CDS” and “ADS”) can be used to denote the intended addressee, such that child-directed speech is speech that is *addressed* to a child and adult-directed speech is speech that is *addressed* to an adult. Second, the terms can be used to denote the speech *register* or other characteristics of the speech. In this second approach, speech that contains the prosodic, lexical, grammatical, and affective characteristics typically associated with speech to children is classified as child-directed speech, regardless of the actual intended addressee, with the same for characteristics of speech to adults resulting in classification as adult-directed speech. In the present study, we will measure linguistic input quantities based on the utterance’s intended *addressee* (e.g., separating speech exclusively directed to the target child versus to another child versus to an adult, etc.).

While qualitative properties of different input types are also important to consider for language development (e.g., response contingency; see Masek, Ramirez, McMillan, Hirsh-Pasek, and Golinkoff (2021) for a recent review), input *quantities* are ideal for roughly comparing the linguistic material children encounter in their daily lives. Input qualities, such as syntactic simplification, high and variable pitch, and vocal timbre modifications have been recognized across culturally and linguistically diverse datasets of child-directed speech and sign (e.g., Bornstein et al., 1992; Brentari, González, Seidl, & Wilbur, 2011; Broesch & Bryant, 2015; Farran, Lee, Yoo, & Oller, 2016; Fernald et al., 1989; Hilton et al., 2022; Masataka, 1998), but these characteristics still vary across cultural groups and, sometimes may not be present at all (e.g., Broesch & Bryant, 2015; Brown, 2014; Hilton et al., 2022; Ochs & Schieffelin, 1984; Pye, 1986). Since these variations in qualitative features across cultural groups are not yet well understood, we restrict ourselves to defining CDS and ADS based on addressee rather than qualitative features.

Quantity-centered estimates of directed speech can also capture some aspects of input “quality”. Addressees have an advantage in comprehension compared to non-addressed others precisely because the utterance in question is suited to the addressee’s immediate comprehension, including timing, word choice, and omitted information (Bell, 1984; Foushee, Srinivasan, & Xu, 2021; Schober & Clark, 1989). Such addressee effects are theorized to exist universally across cultures and communities, given that they arise from general mechanisms of human coordination (Clark, 1996).

Relating back to the idea of input “quality”, consider that adult-directed speech is generally more challenging for young children to comprehend and engage with, while speech that is directly addressed to the target child is highly suited for their comprehension and interactive engagement. By classifying speech by addressee, we can capture clear potential for difference in ease of comprehension (one “quality” factor) of different input sources. In the present paper, we compare adult-directed speech (ADS) quantities against target-child addressed (TCDS) speech quantities—that is, speech addressed *specifically* to the child under study, rather than to another nearby child. These measures represent two distinct sources of linguistic input, so our present study could thus be described as measuring the quantity of two quality types.

## Factors shaping input quantity

A broad spectrum of factors has been suggested to influence the quantity of CDS children encounter in their daily lives. Much less work has investigated factors influencing the quantity of ADS children encounter. We briefly summarize the primary factors examined in prior work, from the macro scale to the micro scale. These factors inform the present study’s analyses.

On the macro end of the spectrum, CDS quantities are thought to be influenced through group membership, for example via socioeconomic group membership or via culturally held beliefs and practices around childrearing (e.g., see Rowe and Weisleder (2020) for a recent review; see Gaskins (2006) and Ochs and Schieffelin (1984) for reviews regarding language socialization and culture). For example, meta-analyses of nearby adult talk in daylong audio data (Piot, Havron, & Cristia, 2022) and CDS in naturalistic, unstructured interaction data (Dailey & Bergelson, 2022) suggest a small but significant positive correlation of linguistic input quantity with socioeconomic status (but see also Bergelson et al., in revision).

Also on the macro scale, there is immense cross-cultural variability in how often children are talked to, by whom, and what is talked about (e.g., de León, 2011; Demuth & Mputhi, 1979; Gaskins, 2006; Ochs & Schieffelin, 1984; Pye, 1986; Rogoff et al., 2003; Rosemberg et al., 2020; Vogt et al., 2015). With respect to CDS quantity, ethnographic and linguistic evidence from two rural communities—Tseltal (Mayan) and Yélî Dnye (Papuan)—suggests that adults in neither community engage in Western-style pedagogical language, but that Tseltal caregivers encourage children to attend to the interactions around them whereas Yélî caregivers more often place infants at the center of social attention (Brown, 2011, 2014; Brown & Casillas, in press). That said, prior work has not found differences in baseline TCDS rate between the two communities, suggesting that the translation from cultural patterns of childrearing to measures of input quantity is complex (Casillas et al., 2020, 2021). As noted above, Shneidman and Goldin-Meadow (2012) *did* observe clear differences in comparing US and Yucatec Mayan children’s input, with the US children under age three hearing significantly more directed input, and mostly from adults. This evidence concords with Cristia’s (2023) characterization of the primary split in input quantities being in rural versus urban populations, more so than culture-to-culture differences.

In the meso part of the spectrum, children’s available interactants also shape what kinds of CDS they hear, with prior results pointing to a greater availability of CDS from adults compared to children, and among adults, women compared to men (e.g., Bergelson, Casillas, et al., 2019; Shneidman & Goldin-Meadow, 2012). As noted above, however, a recurrent theme in work on rural populations is the presence of other children and hence much more CDS from siblings, cousins, and other children (Casillas et al., 2020, 2021; Cristia et al., 2023; Loukatou et al., 2022; Rosemberg et al., 2020; Scaff et al., under review).

A separate meso-level factor is child age: prior work does not consistently demonstrate evidence of change in CDS quantity with age but it does demonstrate age-related change for ADS quantity (Bergelson, Casillas, et al., 2019; Casillas et al., 2020, 2021). For example, Bergelson, Casillas, and colleagues (2019) find that ADS rates decreased significantly over the first few years while CDS quantity rates held constant. Converging evidence for this pattern then came from Casillas and colleagues’ studies of Tseltal (2020) and Yélî Dnye (2021). Ramírez-Esparza and colleagues (2017b) also find increases in non-canonical CDS with age—non-CDS-styled CDS and multiparty CDS—in their sample of US English and English-Spanish bilinguals. Shneidman and Goldin-Meadow (2012) did observe change with age in CDS rate, but only due to increased CDS from other children.

Lastly, on the micro end of the spectrum, CDS and ADS rates fluctuate moment to moment given factors such as the ongoing activity (e.g., playing or eating), the number of potential interactants present, the physical condition of the target child and their surrounding family (e.g., sleeping or awake, stationary or in motion), and more. Soderstrom and Wittebolle (2013) found that linguistic input rates systematically varied depending on the activity context in Canadian daylong recordings (see also Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011). Similar patterns emerged in the daylong recording data of Tseltal and Yélî Dnye-acquiring children (Casillas et al., 2020, 2021). Though we will not, in the present study, have information about activity context, we will at least be able to account for the number of individual talkers present, because when there are more talkers there is more talk. That is, the presence of each additional talker increases competition for the conversational floor, and when four or more talkers are present, group conversations often split into simultaneous sub-conversations, multiplying the amount of observable talk (conversational “schism” see, e.g., Sacks, Schegloff, and Jefferson (1974) and Holler et al. (2021)).

Minimally, the number of talkers present can be considered a “nuisance” variable to help explain fluctuations in CDS and ADS rate over the day. More informatively, however, the number of talkers may serve as a proxy for activities that involve dense family interaction time versus activities where smaller groups of individuals are on their own.[[1]](#footnote-22)

## Extracting CDS and ADS from daylong recordings

Ecologically valid estimates of ADS and CDS input rates are now possible via long-format (e.g., daylong) recordings of children’s home language environments (e.g. LENA, Greenwood et al. (2011); Bergelson, Amatuni, Dailey, Koorathota, and Tor (2019); see Pisani, Gautheron, and Cristia (in prep) for a review). However, to date these recording systems cannot reliably and automatically differentiate between CDS and ADS. Studies that have leveraged daylong recordings have therefore relied on manual annotation to supplement any automated output, taking several different approaches. For example, Weisleder and Fernald (2013) manually classified 5-min blocks of time as primarily child-directed or adult-directed, while Ramírez-Esparza and colleagues (2017a, 2017b, 2014) manually annotated speech-dense clips of audio as having: (1) speech addressed to the child, (2) speech containing the parentese register features of CDS versus ADS (independent of addressee) and (3) who was present as a conversational partner. Moving from the clip level to the utterance level, Bergelson, Casillas, and colleagues (2019) extracted individual utterances using LENA’s automated utterance annotations and then annotated them as child- or adult-directed, based on recognizable CDS and ADS register characteristics.

While these studies examine CDS and ADS in large and highly naturalistic datasets, they either take a very coarse perspective (e.g., examining 5-minute intervals), or tell us about input patterns during the day’s interactional peaks rather than illustrating patterns in children’s average language experiences over the course of a day. In order to extract a representative measure of linguistic input, i.e., how much language children encounter from different types of people in different types of activities across their day (including typical “down” time), we must take random or periodic samples of the language environment (Casillas & Cristia, 2019) rather than only analyzing interactional peaks or estimating across time periods. To gather accurate and representative estimates of natural, at-home CDS and ADS in the present study, we therefore randomly sampled clips from daylong audio recordings and fully transcribed all hearable speech, annotating intended addressee for each utterance in each clip.[[2]](#footnote-24)

## The current work

We examine baseline rates of target-child-directed speech (TCDS) and adult-directed speech (ADS) in the daylong recordings of children growing up in five culturally and linguistically distinct groups: North American English (“NA English”; US & Canadian), United Kingdom English (“UK English”; England), Argentinian Spanish (“Arg. Spanish”; Argentina), Tseltal (Tenejapa, Mayan, Mexico), and Yélî Dnye (Rossel Island, Papuan, Papua New Guinea). As detailed below, some of these corpora include samples from multiple, distinct sub-populations (e.g., NA English includes both US and Canadian English), so we refer to each of these samples as “language groups”. This unique metacorpus draws on seven pre-existing collections of daylong recordings (“corpora”) that were gathered by different research teams, with a variety of different recording devices (i.e., not all LENA), and for a range of different research purposes. Our primary objective was to quantitatively measure the exposure of young children in these groups to two different sources of linguistic input—TCDS and ADS—and to examine several factors associated with variation in this exposure—age, language group, talker type, and number of talkers present.

To accomplish this goal, we defined a second, critical objective: to generate an audio sampling and annotation approach that could be fruitfully employed across recordings made in culturally and linguistically diverse populations (Soderstrom et al., 2021). As motivated above, our analyses focus on two distinct types of linguistic input: TCDS and ADS. Our annotation scheme additionally allows us to examine other types of input, e.g., OCDS (other-child-directed speech, i.e., speech directed to children other than the target child; see Figure 1). For the sake of simplicity, we report data for OCDS in the Supplementary Materials, where we combine it with TCDS to generate parallel analyses of all-CDS (TCDS + OCDS), with similar results to what is reported below.

### Hypotheses.

The specific aims of our analysis are to examine how TCDS and ADS vary across age, language group, talker type, and number of talkers in a highly naturalistic and culturally diverse set of daylong recordings. Before analysis, we established a specific set of hypotheses—with corresponding regression formulae—regarding TCDS and ADS. These hypotheses are based on prior work (see Tables 2 and 3 for detailed overviews). The models for TCDS and ADS are similar but slightly different, as the specific hypotheses about these two input sources slightly differ:

Based on the work cited above, we expected for TCDS rate to vary across language groups (e.g., to be higher-rate in more urban groups) and to come most often from women, but with a greater presence of other-child-produced TCDS in some groups (Tseltal, Yélî Dnye, Argentina). We did not expect any effects of child age on TCDS rate. We expected that TCDS rate would be higher when more talkers were present, given the idea that more talkers results in more talk.

We expected for ADS rate to vary across language groups (e.g., to be lower-rate in more urban contexts), to decrease significantly with child age (especially in groups with high ADS rates early on, e.g., Yélî Dnye), to come most often from women, with greater contributions from children in some groups (Tseltal, Yélî Dnye, Argentina). We also expected that ADS rate would be higher when more talkers were present.

Please note that we limited our hypotheses to simple effects and two-way interactions. We might anticipate other, more complex effects (e.g., the three-way interaction of age-language group-talker type on TCDS rate), but given the limited size of our metacorpus (*N* = 10 recordings per corpus maximum) we leave these effects to be tested in future, larger datasets.

The present paper is the first to bring together all these different factors known to influence TCDS and ADS and to examine their joint effects across multiple language groups (though see Bergelson et al. (in revision) for a mega LENA-based analysis). We examine these factors in order to identify axes of consistency and variation across the multi-corpus sample though, as a reminder, we are not attempting to examine all possible interactions and instead take a hypothesis-driven approach to analysis.[[3]](#footnote-26) Importantly, while we identify key points of theoretically relevant cultural variation across the samples in this study, we do not argue that these language groups represent the full spectrum of diversity in linguistic input experiences, even within the specific populations we have sampled.

# Methods

## Metacorpus Construction

We use the Analyzing Child Language Experiences around the World (ACLEW) metacorpus (Soderstrom et al., 2021) of long-form audio recordings of children’s everyday language environments, comprising seven corpora from five culturally and linguistically distinct groups, labeled here as: North American English (NA English), United Kingdom English (UK English), Argentinian Spanish (Arg. Spanish), Tseltal, and Yélî Dnye.[[4]](#footnote-29) Each group is represented by a single corpus except North American English, for which we had access to three corpora. Recordings for each corpus were originally collected for the unique research purposes of the individual lab contributing the corpus, and therefore there is variation across corpora in the recruitment practices, recording equipment (i.e., not all LENA), recording length, target child ages (see Supplementary Materials Section 3), and other demographic characteristics (see Table 1 for an overview).

### Sampling technique.

Our sampling and annotation scheme needed to be suitable for daylong recordings of different durations made with different recording devices, and for variable annotation situations (e.g., in a lab or in the field).

We selected a single day’s recording for 10 children from each corpus, except the McDivitt-Winnipeg corpus from which we selected 9 recordings due to a sampling error (total recordings *N* = 69); this sample size per corpus reflects what was possible with the smallest corpora in our sample. We used a script to select recordings that were as balanced within and across corpora in reported child gender (male/female), maternal education (below high school–advanced degree), and child age (0;2–3;0; see OSF\_URL\_MASKED\_FOR\_REVIEW for details). The range of available ages was more limited in North American English compared to the other corpora but our statistical approach accounts for this (also see the Supplementary Materials Section 3). Five of the included recordings overlap with those used in Bergelson, Casillas, et al. (2019) and the same Tseltal and Yélî Dnye annotations have been analyzed somewhat differently in separate work (Casillas et al., 2020, 2021).

Each dataset and contributing lab came with a specific set of constraints on what was possible for manual annotation work (e.g., teams of undergraduate students versus individual collaborations with native speakers in remote field sites), so we settled on two basic techniques for sub-sampling and transcribing data from these long-format recordings. These methods for sampling and preparing clips for annotation are illustrated in Figure 1.

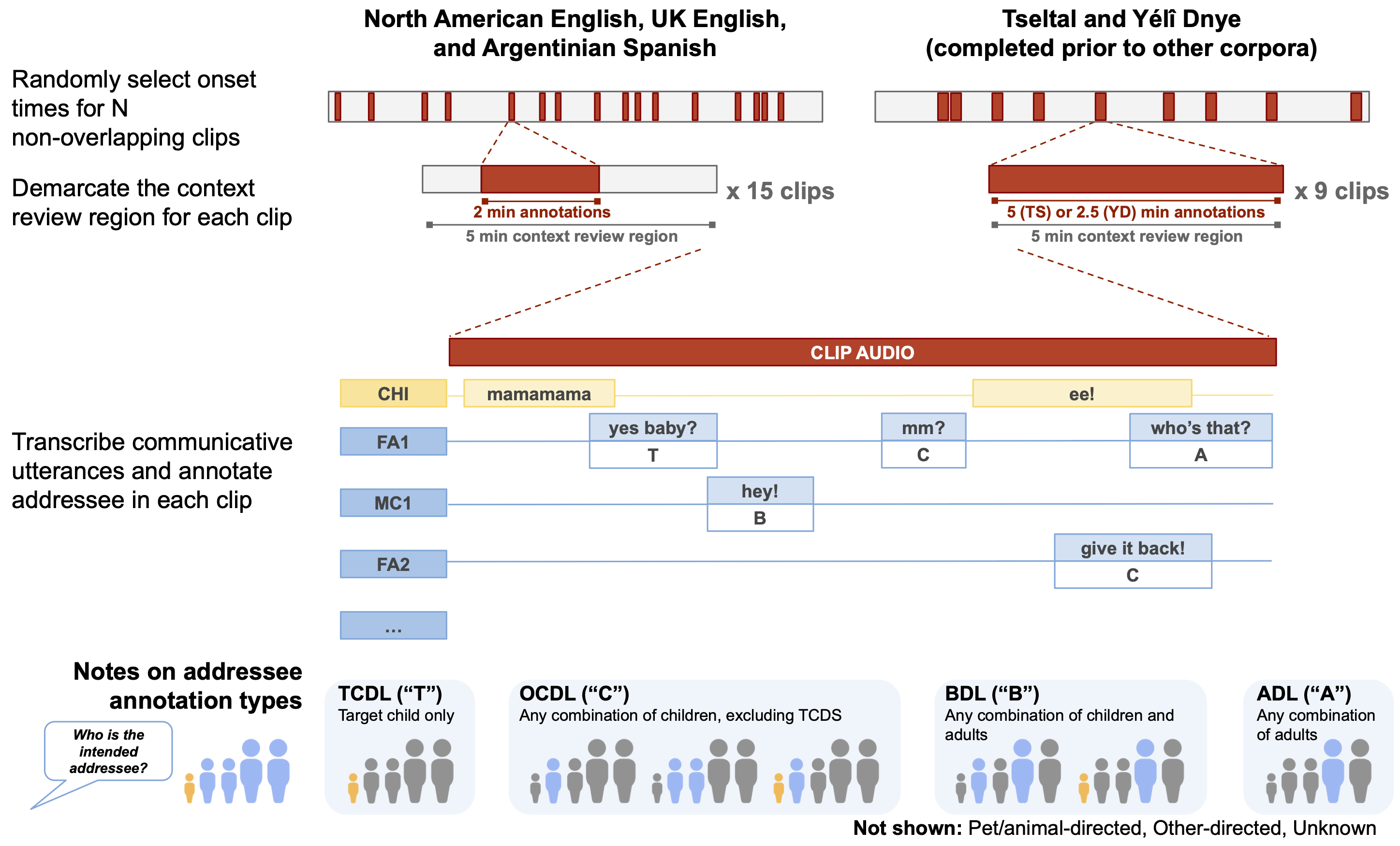


Figure 1: Summary of clip selection and annotation method across corpora.

For North American English, UK English, and Argentinian Spanish (49 of the 69 recordings), we wrote a Python script to randomly pick start times for 15 two-minute clips from throughout the day of each recording, excluding any possibility of clip overlap. The script selected the start and stop times of each clip, as well as the start and stop times of an associated four-minute context period for each clip (see Figure 1, upper left). Thus each of the 15 clips per recording contained one minute of prior context, followed by two minutes of audio to be transcribed and annotated, followed by two more minutes of additional context. The start and stop times of the context and to-be-transcribed clips were then added automatically to a single ELAN (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) audio annotation file that spanned the entire recording. This process resulted in 30 total minutes of annotation per recording.

The Tseltal and Yélî Dnye corpora (20 of the 69 recordings) used a similar method, except only 9 clips were randomly selected. However, the clips were longer than in the other corpora. Tseltal clips were 5 minutes long and Yélî Dnye clips were 2.5 minutes long, resulting in a total of 45 minutes and 22.5 minutes of annotation per recording for the Tseltal and Yélî Dnye corpora, respectively. The five-minute clips in Tseltal had no additional context; this length of clip already provides significant context. The 2.5-minute clips for Yélî Dnye were followed by an additional 2.5 minutes of recording context. Thus, the total context review period for annotation clips across all corpora was five minutes (Figure 1).

Minor deviations in the sampling process between corpora are not expected to have meaningful effects on the analyses: all clips are short and randomly selected from throughout the child’s waking day. These deviations arose because the Tseltal and Yélî Dnye datasets required significant contributions from native local speakers in each remote community sampled, and so the annotation workflow was adapted to suit the associated researcher’s fieldwork schedule. We originally planned for 5-minute clips for the Yélî Dnye data, parallel to the Tseltal data, but this proved infeasible given the time-intensive nature of annotating this community’s high-density language environments, so clip duration was reduced from 5 to 2.5 minutes.

The final clip collection therefore consists of 35.8 hours of transcribed and annotated recording time, of which 16.3 hours consists of communicative vocalizations. Given the constraints across corpora on transcription work hours, this was near the ceiling of manual annotation data we could generate. It was unknown in advance how many recording minutes would be needed to produce meaningful results. That said, Casillas and colleagues (2020, 2021) found that the present amount and distribution of recording minutes were sufficient to detect many of the effects predicted here. Their findings are especially promising for the current set of analyses, which includes a similar statistical approach and re-uses those two datasets (now with additional corpora for comparison). Recent studies (Cychosz, Villanueva, & Weisleder, 2020; Marasli & Montag, 2023; Micheletti et al., 2020) have started building up a more general approach to sampling naturalistic behavior from daylong recordings, but our lack of prior knowledge about the distribution of different input densities from different types of talkers across these groups prevented us from being able to confidently peg our sampling technique to anticipated underlying effects. To counteract what we anticipated would be limited statistical power, we planned to only analyze effects for which we had strong *a priori* predictions (see an overview in Tables 2 and 3).

## Annotation technique

Each of the randomly selected segments were annotated using the ACLEW Annotation Scheme (<https://osf.io/b2jep/>, Casillas et al., 2017; Soderstrom et al., 2021), an ELAN-based approach (Wittenburg et al., 2006). Each annotator undergoes a rigorous and independent training and testing process to ensure intra- and inter-lab consistency in coding. Annotators segmented and transcribed all hearable human communicative vocalizations in the samples, with a separate tier for each individual talker to allow for overlapping talk. Each tier was identified by the talker’s perceived age and gender category (adult/child/unknown and female/male/unknown; e.g., FA1 = female adult 1 in Figure 1). All utterances (except the target child’s) were also annotated for the intended addressee (exclusively target-child-directed/other type of child-only-directed/adult-only directed/mixed-age-directed/animal-directed/other-directed/unknown) on the basis of any available contextual and interactional information within the audio recordings.

Annotator reliability was checked by the complete re-annotation of one-minute from each recording by a new annotator. We then compared the original minute’s annotations to the re-coded minutes’ annotations. A full reliability report is available at URL\_MASKED\_FOR\_REVIEW, but to briefly summarize, error estimates for talker type annotations (e.g., disagreements about whether the talker is the target child or a different child) are far better than prior work has found between human and LENA (i.e., automated) annotations. Further, comprehensive kappa scores reflect moderate-to-substantial agreement (cross-corpus *k* range = 0.55–0.68) for talker types and slight-to-substantial agreement (cross-corpus *k* range = 0.32–0.64) for addressee, with wide variability in agreement between corpora. Despite the fact that CDS can have some cross-linguistically recognizable features (e.g., Bornstein et al., 1992; Fernald et al., 1989; Hilton et al., 2022), we had expected somewhat lower reliability scores for addressee annotations because the reliability annotators did not always understand the language of the file they were annotating; their annotation decisions were thus less informed by lexicosyntactic content than the (native-speaking) original annotators’. Most cases of disagreement arose when one annotator indicated silence or overlapping talk where the other annotator indicated talk from a single person—confusion between actual addressee categories was relatively low (see Supplementary Materials Section 8 for more details).

Table   
 *Details for the corpora in the dataset (Bergelson et al., 2016; Casillas, Brown & Levinson, 2017; McDivitt & Soderstrom, 2016; Rosemberg et al., 2015; Rowland et al., 2018; Warlaumont et al., 2016). Parentheses following the mean indicate the range across participants.*

| Corpus | Language | Ages | Household Size | Region | Participants | Recording Lengths |
| --- | --- | --- | --- | --- | --- | --- |
| Bergelson | North American English | 11.2 mo (7–17mo) | 4.2 (3–7) | Northeast (US) | 10 (4 males) | 802.8 min (665–960 min) |
| McDivitt-Winnipeg | North American English | 12.4 mo (2–32 mo) | 3.3 (2–5) | Winnipeg Manitoba (Canada) | 9 (4 males) | 532.6 min (118–837 min) |
| Warlaumont | North American English | 6.3 mo (3–9 mo) | Not available | Central Valley California (US) | 10 (5 males) | 815.8 min (612–960 min) |
| LuCiD Lang0-5 | UK English | 20 mo (11–31 mo) | 3.6 (3–5) | North West (England) | 10 (5 males) | 931.5 min (845–960 min) |
| Rosemberg | Argentinian Spanish | 17.1 mo (9–27 mo) | 5.5 (3–12) | Buenos Aires (Argentina) | 10 (6 males) | 242 min (125–303 min) |
| Casillas-Tseltal | Tseltal (Mayan) | 16.1 mo (2–36 mo) | 7 (7.5; 4–14) | Chiapas (Mexico) | 10 (5 males) | 554.5 min (492–578 min) |
| Casillas-Yélî Dnye | Yélî Dnye (Papuan) | 14.2 mo (2–36 mo) | 6.5 (3–11) | Milne Bay Province (Papua New Guinea) | 10 (5 males) | 488.2 min (430–555 min) |
| Total |  | 13.9 mo (2–36 mo) | 5.1 (2–14) |  | 69 (34 males) | 625.2 min (118–960min) |

## Data analysis

All statistical analyses were conducted in R with the glmmTMB package (Brooks et al., 2017; R Core Team, 2019) and all figures were generated with ggplot2 (Wickham, 2016). Analysis scripts and raw anonymized data are available at URL\_MASKED\_FOR\_REVIEW. Our two dependent measures were the rates of TCDS and ADS (both expressed in minutes per hour). We calculated TCDS and ADS input rate for each clip for each of three talker types: female adults (here “women”), male adults (here “men”), and children (here “children”, including both male and female children). All other utterances (e.g., language addressed to animals and language produced by electronic devices) were excluded. As motivated above, we designate TCDS versus ADS utterances based on who they were perceivably addressed to: ‘TCDS’ includes communicative utterances that were addressed exclusively to the target child (from an adult or another child). ‘ADS’ includes communicative utterances addressed to one or more adults (from an adult or from another child).

TCDS and ADS input rate cannot be negative. In practice, they are modally zero or close to zero. Given our random sampling technique, which can include periods of silence, many clips include no TCDS or ADS. These “down” times for input are part of the representative pattern of children’s language experience [[5]](#footnote-39) but also present an analytical challenge: observed cases of 0 TCDS/ADS in many clips combined with a skewed non-negative distribution of > 0 TCDS/ADS in other clips. This distribution of TCDS/ADS across sampled clips cannot be modeled with the assumption of normality. We therefore used zero-inflated negative binomial mixed-effects regressions for our analyses. This regression type uses a two-model approach to overcome non-negative, overdispersed data with extra cases of zero—the case for the present data (Brooks et al., 2017; Smithson & Merkle, 2013).

The two model components constructed for the analyses of TCDS and ADS are: (1) a **zero-inflation model**, which uses a logistic regression to model the likelihood of the *presence* of ‘zero’ cases in the data (e.g., answering questions like ‘are zero-TCDS clips less likely for older target children?’) and (2) a **count model**, which uses linear regression to model how the *rate* of TCDS/ADS is influenced by the predictors of interest (e.g., answering questions like ‘is TCDS rate higher for older target children?’). The *a priori* predictions we laid out above can be applied to both model components, as shown in Tables 2 and 3.

The simple effects included in the models were target child age (centered and standardized from age in months), number of talkers present in that clip (centered and standardized from the unique number of talkers across all clips), talker type (woman versus man/child), and language sample (North American English versus UK English/Argentinian Spanish/Tseltal/Yélî Dnye). We only included interactions for which we had a strong *a priori* hypothesis and thus the models for TCDS and ADS differ slightly in their structure (see the Results for the regression formulae).

We modeled language group and talker type as factorial variables, which limits our ability to make comparisons among language groups; e.g., if Tseltal were the reference level, the model outcomes for language group would give pairwise comparisons between Tseltal and all the other language groups, but not pairwise comparisons between other language groups, for example, between Argentinian Spanish and UK English. We selected ‘North American English’ and ‘women’ as our default reference levels for reporting model estimates below given that North American English and linguistic input from female adults are the most well represented in (a) the current dataset and (b) prior work done on our sampled populations. In addition, we are interested in establishing under-studied patterns that may be present in our dataset—effects that diverge from groups that are currently over-represented in the literature. Setting these levels as a reference gives us a first glimpse into the variation that has gone under-examined in past work. This analysis should not be understood as positioning North American English as a global “norm”.

That said, there is interest among our author team and others in the field in *pairwise* comparisons between language groups. To allow curious readers to inspect more closely how the reported effects below are impacted by the selected reference level of language group, we include versions of our models with each language group as the reference level in the Supplementary Materials (i.e., five versions of the TCDS and ADS model each; Section 6). Here in the main text our results focus on models of TCDS and ADS with North American English as the reference level for language group and women as the reference level for talker type.

Table   
 *Predictions for TCDS analysis. Asterisk indicates previously observed with daylong child language data (Casillas et al., 2020, 2021; Scaff et al., under review). The ‘Supported’ column reflects the extent to which each finding from the current study aligns with its predicted outcome.*

| Predictor | Hypothesized outcome | Supported |
| --- | --- | --- |
| Count model |  |  |
| Target child age | No change in TCDS rate with age\* | Yes |
| Talker type | Women > men, children\* | Yes |
| # Talkers present | More talkers leads to more TCDS\* | Yes |
| Corpus | NA English shows more TCDS than Argentinian Spanish, Tseltal, and Yélî Dnye | Yélî Dnye only |
| Talker type x Corpus | More TCDS from children in Argentinian Spanish, Tseltal, and Yélî Dnye | more TCDS from children in all four corpora versus NA English, more TCDS from men in Argentinian Spanish and Yélî Dnye versus NA English |
| Target child age x Talker type | More TCDS from children for older target children | Yes |
| Zero-inflation model |  |  |
| Target child age | More zero-TCDS clips for young children (e.g., during naps) | No |
| Corpus | More zero-TCDS clips in Argentinian Spanish, Tseltal, and Yélî Dnye than NA English | No |

Table   
 *Predictions for ADS analysis. Asterisk indicates predictions on the basis of related or similar data. The ‘Supported’ column reflects the extent to which each finding from the current study aligns with its predicted outcome*

| Predictor | Hypothesized outcome | Supported |
| --- | --- | --- |
| Count model |  |  |
| Target child age | Decrease in ADS rate with age\* | Yes |
| Talker type | Women > men, children\* | Yes |
| # Talkers present | More talkers leads to more ADS\* | Yes |
| Corpus | Argentinian Spanish, Tseltal, and Yélî Dnye show more ADS than NA English | No |
| Talker type x Corpus | More ADS from children in Argentinian Spanish, Tseltal, and Yélî Dnye than NA English, no specific predictions regarding men | No—UK English only |
| Target child age x Corpus | Bigger ADS decrease across age in Argentinian Spanish, Tseltal, and Yélî Dnye | No |
| Zero-inflation model |  |  |
| Target child age | More zero-ADS clips for older children | (untested due to convergence issues) |
| Corpus | More zero-ADS clips in NA English than Argentinian Spanish, Tseltal, and Yélî Dnye | No |

# Results

Descriptive statistics for observed TCDS and ADS rates by language group and talker type are shown in Table 4 and in Figure 2. A visual summary of statistical model outcomes from the count models of TCDS and ADS rate is shown in Figure 3. Further, marginal means plots of model-predicted TCDS and ADS rates across age, language group, and talker type are available in Supplementary Materials Section 7. In Tables 2 and 3 we provide a high-level summary of which hypothesized outcomes were statistically supported in the regressions described below.

Table   
 *Average input rates per clip across participants for each language group. Note that these descriptive statistics are raw rates and therefore reflect overall differences between corpora without controlling for, e.g., number of talkers present, which are accounted for in the statistical analyses. Parentheses following the mean indicate the median and range across participants in each group.*

| Language | TCDS rate | ADS rate | Mean proportion TCDS |
| --- | --- | --- | --- |
| NA English | 3.49 (3.24; 0–10.12) | 8.06 (6.86; 0–19.32) | 0.30 |
| UK English | 3.69 (3.72; 1.22–7.15) | 4.38 (4.42; 0.6–9.59) | 0.46 |
| Arg. Spanish | 4.77 (3.19; 1.4–9.38) | 10.83 (10.24; 1.59–23.93) | 0.31 |
| Tseltal | 3.54 (3.94; 0.83–6.55) | 11.08 (8.35; 2.78–33.08) | 0.24 |
| Yélî Dnye | 3.13 (2.95; 1.58–6.26) | 19.87 (17.1; 7.25–38.54) | 0.14 |

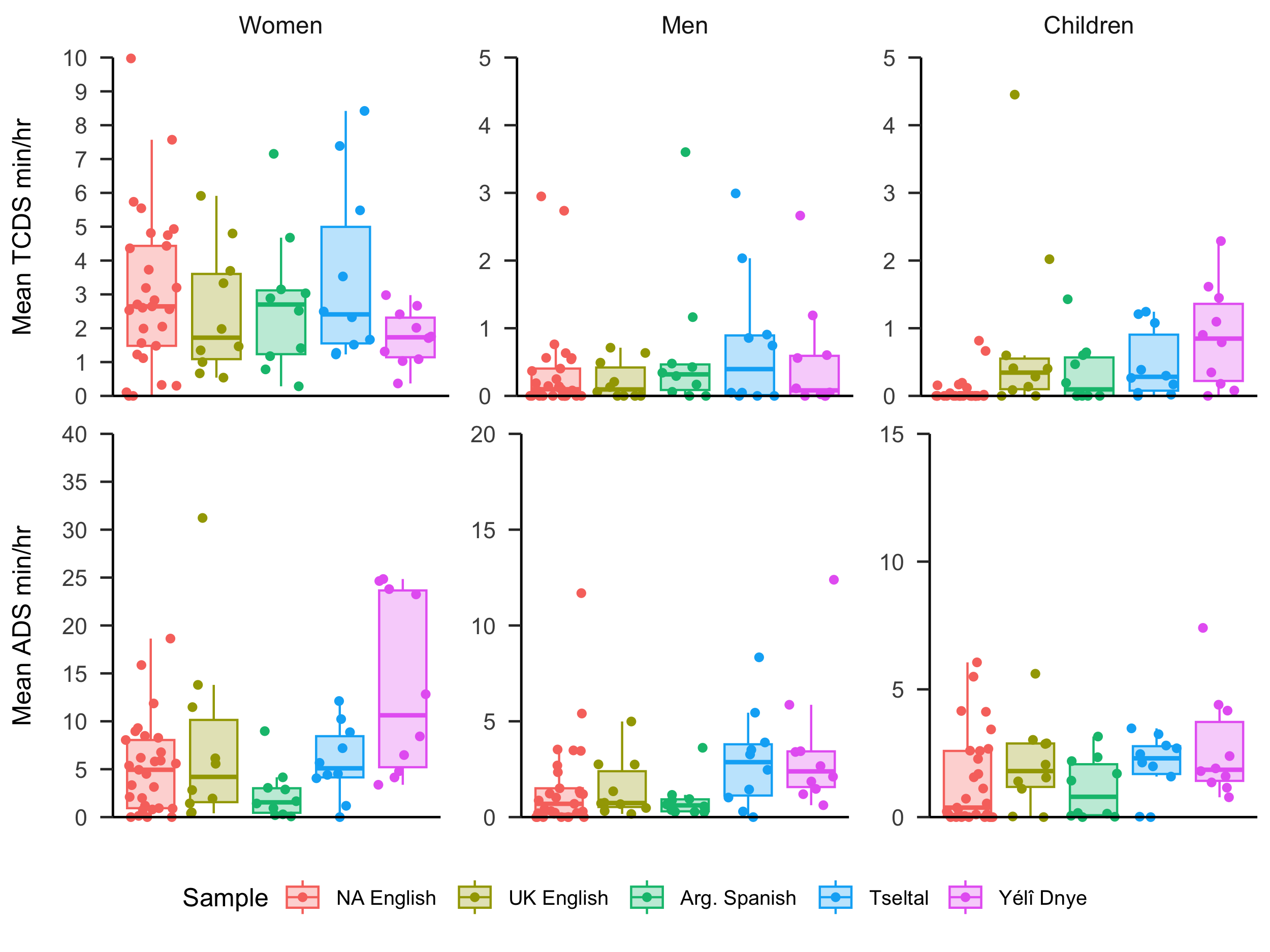


Figure 2: Mean by-recording rates of TCDS (above) and ADS (below) min/hr rates across language groups and talker types. For example, the upper-leftmost datapoint shows a recording with an average of 10 minutes per hour of TCDS from women talkers in North American English. Left-to-right order of language group within each panel matches the order shown in the legend.

## Target-child-directed Speech

On average, across all recordings, children were exposed to 3.66 minutes of TCDS per hour (median = 3.24), with substantial individual variation between children (range = 0–10.12). Our model of TCDS rate included target child age (numeric; standardized), talker type (factorial; woman/man/child), the number of talkers present in the clip (numeric; standardized), and language group (factorial; NA English/UK English/Argentinian Spanish/Tseltal/Yélî Dnye), with two additional two-way interactions (talker type by language group and child age by talker type) and random intercepts by child. The zero-inflation model component included child age and language group as predictors (*N* = 2745 clips, log-likelihood = -2,703.72, overdispersion estimate = 8.94; formula = TCDS.min.p.hr ~ child.age + talker.type + num.tlkrs.in.clip + lang.grp + talker.type:lang.grp + child.age:talker.type + (1 | child.id), ziformula = ~ child.age + lang.grp).

As a reminder, we report results from the model of TCDS with North American English as the reference level for language group and women as the reference group for talker type. Identical models with the full range of alternate reference levels for language group are available in the Supplementary Materials (Section 6). Unless otherwise noted, the significant effects reported below are qualitatively similar (i.e., significant in the same direction) in all alternate models.

### Effects of child age, talker type, and number of talkers present.

As predicted, we found no evidence that TCDS changed with age (*B* = -0.03, *SE* = 0.09, *z* = -0.31, *p* = 0.76). TCDS rate was significantly lower for men compared to women (*B* = -2.03, *SE* = 0.19, *z* = -10.69, *p* < 0.001) and for children compared to women (*B* = -3.54, *SE* = 0.37, *z* = -9.64, *p* < 0.001). TCDS rate was also significantly higher when there were more talkers present (*B* = 0.33, *SE* = 0.04, *z* = 7.62, *p* < 0.001).

### Effects relating to language group.

The baseline rate of TCDS input in Yélî Dnye was estimated to be significantly lower than North American English (*B* = -0.95, *SE* = 0.32, *z* = -2.97, *p* < 0.01), with no evidence for difference in baseline TCDS rate between North American English and any other language group (all *p*’s ≥ 0.58). TCDS input rate from men varied between language groups: compared to North American English, the TCDS rate from men was significantly higher in both Argentinian Spanish (*B* = 0.70, *SE* = 0.31, *z* = 2.29, *p* = 0.02) and Yélî Dnye (*B* = 0.75, *SE* = 0.37, *z* = 2.02, *p* = 0.04). Similarly, TCDS from children varied between language groups: compared to North American English, TCDS from children was significantly higher rate in all other language groups (UK English: *B* = 1.10, *SE* = 0.51, *z* = 2.16, *p* = 0.03; Argentinian Spanish: *B* = 1.58, *SE* = 0.45, *z* = 3.48, *p* < 0.001; Tseltal: *B* = 1.91, *SE* = 0.49, *z* = 3.91, *p* < 0.001; Yélî Dnye: *B* = 2.81, *SE* = 0.46, *z* = 6.11, *p* < 0.001).

### Interaction between child age and talker type.

We found no evidence that TCDS from men changed with age relative to TCDS from women (*B* = -0.13, *SE* = 0.13, *z* = -1.01, *p* = 0.31). In contrast, TCDS from children increased with age relative to TCDS from women (*B* = 0.29, *SE* = 0.12, *z* = 2.35, *p* = 0.02).

The zero-inflation regression component did not suggest any additional evidence for effects of child age or language group (North American English versus other groups) on the likelihood of a clip containing zero TCDS (all *p*’s ≥ 0.27).

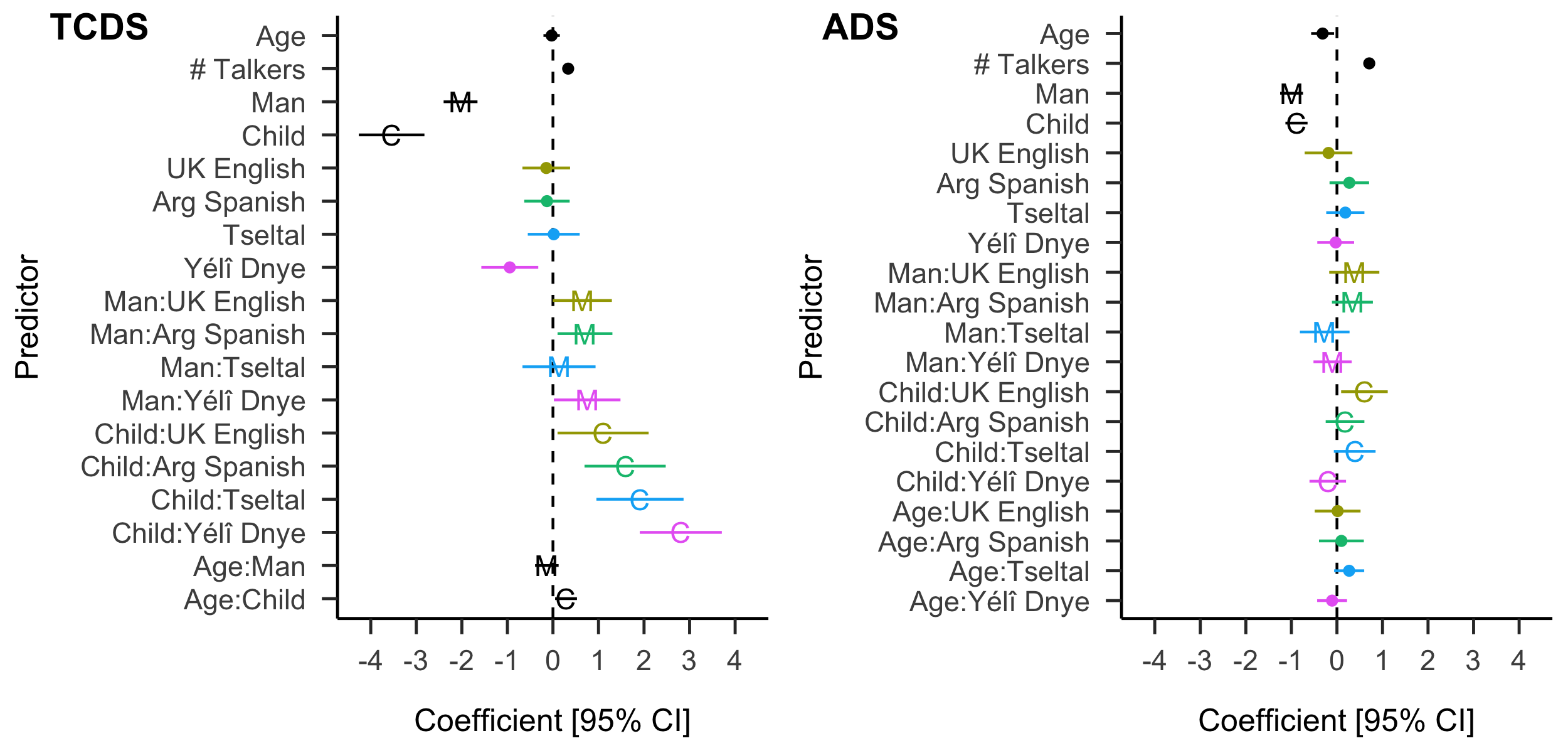


Figure 3: Coefficients with 95% confidence intervals from the count models of TCDS (left) and ADS (right) for all included fixed effects in the model with NA English and women set as the reference levels for language group and talker type, respectively. Intervals not overlapping with zero indicate significance. Color indicates population, ‘C’ and ‘M’ indicate effects related to child- and man-produced utterances, respectively. For example, both the left and the right panels show that both child- and man-produced input are significantly less frequent compared to the reference levels of woman-produced input. Note that the fixed effects included in each model are determined by the predictions laid out above separately for TCDS and ADS.

## Adult-directed speech

On average, across all recordings, children were exposed to 10.08 minutes of ADS per hour (median = 8.34), again with considerable variation between children (range = 0–38.54). Our model of ADS rate included target child age (numeric; standardized), talker type (factorial; woman/man/child), number of talkers in the clip (numeric;standardized), and language group (factorial; NA English/UK English/Argentinian Spanish/Tseltal/Yélî Dnye), with two additional two-way interactions (talker type by language group and child age by language group) and random intercepts by child. The zero-inflation model component only included language group; we had planned to also include child age in the zero-inflation component, but its inclusion led to issues of model non-convergence. Child age remained a predictor in the count model (*N* = 2745 clips, log-likelihood = -4,190.69, overdispersion estimate = 15.69; formula = ADS.min.p.hr ~ child.age + talker.type + num.tlkrs.in.clip + lang.grp + talker.type:lang.grp + child.age:lang.grp + (1 | child.id), ziformula = ~ lang.grp).

Again, we report results from the model of ADS with North American English as the reference level for language group and women as the reference group for talker type, and unless otherwise noted, the significant effects reported are qualitatively similar in all alternate models found in the Supplementary Materials (Section 6).

### Effects of child age, talker type, and number of talkers present.

ADS rate decreased significantly with age (*B* = -0.31, *SE* = 0.13, *z* = -2.45, *p* = 0.01), but this effect was non-significant in some alternate models with other reference levels (see Supplementary Materials Section 6) and so should be considered preliminary. We note that, across all alternate models, the estimate for an effect of age remained numerically negative. ADS rate was also significantly lower for men compared to women (*B* = -1, *SE* = 0.13, *z* = -7.77, *p* < 0.001) and for children compared to women (*B* = -0.89, *SE* = 0.12, *z* = -7.16, *p* < 0.001). This result, suggesting a difference between children and women, depends on which language group is chosen for the reference level. The predictor is non-significant, though still numerically negative, when UK English is set as the reference level (see Supplementary Materials Section 6). As with TCDS, ADS was significantly higher when there were more talkers present (*B* = 0.71, *SE* = 0.03, *z* = 21.54, *p* < 0.001).

### Effects relating to language group.

There was no evidence for differences between baseline ADS input rate in North American English and any other language group (all *p*’s ≥ 0.22). There was also no evidence that the difference in women’s and men’s ADS input rates varied between North American English and any other language group (all *p*’s ≥ 0.14). In contrast, the difference in women’s and children’s ADS input rates was significantly smaller in UK English compared to North American English (*B* = 0.60, *SE* = 0.26, *z* = 2.31, *p* = 0.02). There was also no evidence that age-related change in ADS input rates varied between North American English and any other language group (all *p*’s ≥ 0.11).

The zero-inflation regression component did not suggest any additional evidence for effects of child age or language group (North American English versus other groups) on the likelihood of a clip containing zero ADS (all *p*’s ≥ 0.99).

Readers interested in exploring further pairwise comparisons of TCDS and ADS effects between language groups (e.g., Tseltal versus UK English) are encouraged to view alternate versions of the models of TCDS and ADS in the Supplementary Materials (Section 6).

# Discussion

We examined how two input sources, TCDS and ADS, vary in children’s early language environments, depending on child age, talker type, language group, and number of talkers present. Our data come from a metacorpus of 69 daylong recordings from children under three in five culturally and linguistically distinct groups. The present paper is the first to examine the joint effects of these factors across multiple language groups, shedding light on typical patterns in children’s early language experiences across these diverse contexts. This project also presented a successful model for sampling and annotating child language data in a unified manner across diverse language groups. In this discussion we highlight four major findings: (1) Minimal effects of age, (2) women’s input predominates, men’s is rare, and children’s varies between cultural groups, (3) more talkers leads to more talk, and (4) minimal evidence for *baseline* differences in TCDS and ADS input rates between language groups. While many of the predictions we made were supported, some were not (Tables 2 and 3). In what follows, we briefly discuss each of the four major findings highlighted, raising the most relevant implications of each.

## Minimal effects of age

TCDS rate showed no significant change across this developmental period (0;0–3;0) while ADS rate significantly decreased with target child age. The result replicates prior findings on daylong TCDS and ADS in these groups (Bergelson, 2020; Bergelson, Casillas, et al., 2019; Casillas et al., 2020, 2021). However, this significant effect of target child age on ADS rate should be taken as preliminary given that alternate reference-level models do not always show this effect. The lack of evidence for an increase in TCDS rate with age is consistent with our predictions may appear inconsistent with the findings reported in Ramírez-Esparza and colleagues’ (2017a) study. The substantial differences in their and our constructs (i.e., “parentese” register versus TCDS) and measurement approach (i.e., clip-by-clip classification versus utterance-based transcript analysis) unfortunately prevent direct comparison between these two studies, but future work may examine both approaches within the same corpus to get a more comprehensive view of how child age impacts the quantities of different input sources.

It is not yet clear what would lead to this decrease in ADS with target child age. One existing proposal is that children become independently able to wander away from adult conversation as they gain mobility and independence (Bergelson, Casillas, et al., 2019). This proposal is consistent with our main result, but confirming it would require information beyond the current recordings, and we would moreover need to explain why the decrease in ADS is sensitive to which language group is selected as the reference.

The lack of evidence for an increase in TCDS during this early period, when we *know* that children experience immense growth in their linguistic knowledge and processing capacity, aligns with recent work reasoning that growth in early linguistic skills reflects children’s changing efficiency and sophistication in extracting relevant information from their ambient linguistic environments, *not* direct changes to their linguistic input (see Bergelson, 2020 for a review). Rather than attributing developmental change to changes in the input, this theoretical approach looks instead to changes in children’s ability to engage in real-time language prediction and use of already acquired world and symbolic language knowledge (Bergelson, 2020; Meylan & Bergelson, 2022; Snedeker, Geren, & Shafto, 2007). To this account, the current findings add an important cross-linguistic perspective: this basic idea may apply across diverse linguistic contexts.

## Women’s input predominates, men’s is rare, and children’s varies between cultural groups

Regarding the talkers producing children’s input, we found that women predominate in children’s language environments. The prevalence of woman-produced language over man- and child-produced language was evident for both TCDS and ADS. However, the extent to which women’s input predominates—especially for TCDS—varied. TCDS produced by men was significantly higher rate in Argentinian Spanish and Yélî Dnye compared to North American English. TCDS produced by children was significantly higher rate in all language groups compared to North American English, and TCDS produced by children increased more with age relative to TCDS produced by women. In contrast, we found very little evidence for change in the rates of ADS from different talkers across age or language group: only UK English showed a significant difference from North American English, with a significantly smaller gap between children and women’s ADS rates.

One implication of these findings is that, across these diverse language groups, women’s input plays an outsize role in their children’s input, both in terms of directed and observable language. While there was very clear cross-cultural variation in the contribution of different talker types, this central role for woman-produced linguistic input was clear across our dataset. We are far from the first researchers to make this observation for child language input (see, e.g., Bateson, 1979; Bergelson, Casillas, et al., 2019; Bruner, 1983; Cooper & Aslin, 1989; Mannle, Barton, & Tomasello, 1992), and talker-specific effects on early linguistic representations have been demonstrated previously in experimental tests of implicit language knowledge (e.g., Bergelson & Swingley, 2018; Hillairet de Boisferon et al., 2015; Houston & Jusczyk, 2000; Martin et al., 2015). However, our findings underscore how cross-culturally pervasive these effects may be, urging further work on the talker-specific properties of infants’ early linguistic representations and the mechanisms by which these early representations become more robust to different talker types over time.

## More talkers leads to more talk

As predicted, and consistent with prior work (Casillas et al., 2020, 2021; Sacks et al., 1974), we found that more talkers leads to more input, both for TCDS and ADS. As discussed in more detail below, systematic cross-cultural differences in multi-party interaction suggest that this effect can be partly understood as cultural rather than situation-specific. Here we briefly discuss the component of this effect that is not cultural—the simple fact that, all else being equal, more talkers leads to more talk.

Our interpretation of this basic effect, which is not mutually exclusive with the cultural interpretation, is that the presence of more talkers increases competition for the floor (Holler et al., 2021; Sacks et al., 1974). When there are four or more individuals present, as is the average case in all but the English-speaking groups (see Supplementary Materials Section 2), there is an opportunity for conversational schism. In cases of schism, subgroups break off into smaller conversations (e.g., 2 two-person conversations), potentially doubling the observable talk in the input (Sacks et al., 1974). Future work might selectively examine subsets of daylong data to more precisely characterize how interactionally driven factors such as the number of talkers present in a clip accounts for fluctuations in a child’s linguistic input rate, both within and across language groups.

It is notable that multi-party talk and, in general, observable talk between others is abundant across these diverse groups, with raw ADS estimates typically far outpacing exclusive, directed input (i.e., TCDS). While a variety of language-learning processes may indeed benefit from early exposure to observable talk (e.g., syntactically complex prosodic structures in adult-adult conversation), ADS learning effects in infancy and early toddlerhood have gone largely unexamined (though see, e.g., Akhtar, Jipson, and Callanan (2001) and Oshima-Takane, Goodz, and Derevensky (1996)). Given the current results, it may be worthwhile to dig deeper into how observable talk and multi-party talk influences the very early stages of language learning, especially the early development of non-referential linguistic knowledge (e.g., regarding phonology, see Cristia (2020), regarding conversational turn taking, see Dunn and Shatz (1989)).

## Minimal evidence for *baseline* differences in language group

Regarding effects of cultural group on baseline TCDS and ADS input rates, we had predicted that children acquiring Argentinian Spanish, Tseltal, and Yélî Dnye would encounter less TCDS and more ADS compared to North American English-acquiring children. By and large, our results do not support these hypotheses. When it came to TCDS, we only observed one case where input rates differed: Yélî Dnye’s baseline TCDS rate was significantly lower than that of North American English. When it came to ADS, we found no evidence for differences in baseline ADS rate between North American English and other language groups. This set of results may come as a surprise, considering that the raw rates of TCDS and ADS clearly vary between language groups, in ways that often align with our original predictions (e.g., the raw ADS rate of Yélî Dnye is nearly two and a half times larger than that of North American English; Table 4). A reasonable question is then why our model estimates don’t reflect these differences.

Many of these apparent inconsistencies with mean overall TCDS and ADS rate come from systematic between-group differences in the number and composition of talkers. For example, Yélî Dnye children had an average of 6.06 talkers present in addition to the target child, while North American English children only had 1.81. So, even if the baseline rate of TCDS is significantly lower per talker in Yélî Dnye (as our model above suggests), there are *many more* talkers present in the Yélî children’s acoustic environment compared to the North American children. Consequently, the overall experienced TCDS by children in these two groups appears overall comparable (Table 4).

We tested this idea in a *post-hoc* analysis where we removed number and type of talker from the regression models, only leaving child age and language group as predictors (Supplementary Materials Section 4, i.e., formula = min.p.hr ~ child.age + lang.grp + + (1 | child.id), ziformula = ~ child.age + lang.grp). To the extent that number of talkers and talker type are correlated with language group, their associated variance will be incorporated with language group effects in these simpler models (Wurm & Fisicaro, 2014), giving us an interpretive view closer to that implied by the by-language-group averages in Table 4. The simpler models suggested no evidence for difference in TCDS rate between North American English and the other groups, and suggest significantly higher ADS in Yélî Dnye (and significantly lower likelihood of a zero-ADS clip) compared to North American English. In sum, the results for Yélî Dnye look very different depending on whether variance in the number of talkers (and thereby variance in the quantity of TCDS/ADS) can be attributed to the language group (i.e., in the “simple” models) or whether it’s pulled out as a separate predictor (i.e., in the primary models). This point is important for two reasons.

First, the same data can lead to very different conclusions depending on what variance is treated as culture-specific vs. culture-general. From an ethnographic perspective, it may be completely valid to consider features like number and composition of talkers a part of children’s specific cultural and linguistic milieu. The number of talkers present, after all, likely relates to cultural practices around childcare (e.g., alloparenting), household organization (e.g., multigenerational housing), and daily activities (e.g., food preparation routines; see Gaskins (1999) and Casillas (2023) for more discussion of these issues). Put differently, variation in the number of talkers present is *cultural* and not just situation specific. However, our perspective is that, in order to understand how these factors might generally and cross-culturally influence children’s linguistic input, we need to analyze them as (partly) separate from culture. Doing so gives us a glimpse into how basic processes of conversational coordination and caregiving may shape children’s input in broadly similar ways across diverse human groups and thus give us insight into how children learn language so robustly across widely varied home environments.

Second, and complementarily, a lack of cultural group effects (beyond differences in the prevalence of multiparty talk) does *not* imply that early language environments are cross-culturally and cross-linguistically similar. Our measures represent highly simplified quantifications of two sources of linguistic input—one designed specifically for the target child and one designed for adults—but capture nothing about the content of naturalistic input or its integration into children’s interactive or multi-modal experiences (e.g., Bergelson, Amatuni, et al., 2019; Broesch, Rochat, Olah, Broesch, & Henrich, 2016; de León, 1998; Kuchirko, Tafuro, & Tamis-LeMonda, 2018; Montag, 2020; Rowe, 2012). The measure we use here, while a crucial starting point, is too coarse to make detailed conclusions regarding qualitative similarities or differences in children’s early language experiences. To do so, we would need much more than the present data can offer, at least: (1) detailed generative models of how much input children encounter, from whom, and under what conditions—to which the present study contributes, (2) an understanding of the content of that input and how it fluctuates under different conditions, and (3) documentation of the local cultural, social, economic, and material realities that may radically change the experienced linguistic input.

Finally, we note that our findings do not cleanly divide between so-called “WEIRD” and “non-WEIRD” (Henrich, Heine, & Norenzayan, 2010) groups. For example, Yélî Dnye and Tseltal—the two rural subsistence communities represented here—do not pattern together in our data, and neither do the two historically related urban populations—North American and UK English (see Cristia (2023) for further discussion). This highlights the importance of considering each population in its own right when making claims about cultural similarities and differences. While ultimately we hope to pinpoint areas of similarity and systematic variation in language development across a wide variety of developmental contexts, it is far too early to make universalist claims about patterns in children’s real-world language experiences. The WEIRD or non-WEIRD distinction, while helpful to illustrate cultural biases in behavioral research, can also unfortunately reinforce those same biases by grouping together very distinct cultural groups in opposition to a Western, primarily North American, groups (see Singh, Cristia, Karasik, Rajendra, and Oakes (2023) for further discussion relating specifically to infant research).

## Limitations

There were minor methodological variations in sampling due to the logistical constraints in doing annotation across corpora. We have no specific reason to believe that these discrepancies impacted our findings in any meaningful way. Of greater concern, however, is whether our collection of annotated clips constitute enough data to reveal true underlying effects. We sampled randomly over the course of the daylong recording to capture a representative sample of young children’s input, which often includes “down” time moments. Also, given the diversity of populations in our metacorpus, random sampling was the most straightforward way to ensure that our sampling method itself did not introduce confounds across corpora (e.g., if we had picked high-vocal-activity segments only or otherwise activity-centered moments like “play”). However, the highly zero-inflated nature of children’s daily experiences (Mendoza & Fausey, 2021) provides challenges both for statistical modeling and for interpreting the findings.

Best estimates to date suggest that our sample size (22.5–45 minutes per recording) *is* reasonable for obtaining preliminary stable estimates (Cychosz et al., 2020; Micheletti et al., 2020). For example, Marasli and Montag (2023) examine estimated versus true word counts from daylong recordings using a variety of random sampling schemes, finding that a total of 30 minutes of randomly sampled 1–5-minute clips generates accurate average word count estimates, with varying but symmetrical rates of error depending on clip size (shorter is better). Word count and utterance duration are highly correlated, and utterance duration directly corresponds to our measure of quantity (see DeAnda, Bosch, Poulin-Dubois, Zesiger, and Friend (2016) and also Räsänen, Seshadri, Lavechin, Cristia, and Casillas (2021) for evidence from the corpora used here). Therefore we consider the currently sampled data as sufficient for an accurate approximation of input rates from daylong recordings. However, we leave it to future work to refine this assumption based on a greater diversity of daylong recording types.

Importantly, we see the present study as an initial assessment of differences between these populations in children’s home linguistic experiences, and do not believe that any single study should be considered the final word in comparisons of this nature. Indeed, another weakness of the current study is a lack of deep incorporation of existing ethnographic and language socialization claims about these populations (Brown & Gaskins, 2014; Gaskins, 2006; Ochs & Schieffelin, 1984). What our findings *do* highlight is that specific facets of behavioral patterns (e.g., housing arrangements, child caregivers, etc.) are actually visible in quantitative measures of children’s language environment in ways that allow us to identify axes of cross-cultural variation relevant for developing generalizable theories of language learning. By looking deeper at the local context for each dataset, we would better understand variation within each. For example, we have not yet been able to meaningfully analyze variation in socioeconomic status among the included language groups.[[6]](#footnote-65) We strongly urge readers to take caution in generalizing our results (or those of other researchers) beyond the current data, to new populations.

Finally, the present set of analyses examines input without taking into account patterns of target child vocalization, turn-taking with the target child, or examining how overlapping vocalizations would change the estimates presented here (e.g., Broesch et al., 2016; de León, 1998; Donnelly & Kidd, 2021; Elmlinger, Goldstein, & Casillas, 2023; Kuchirko et al., 2018; Scaff et al., under review). Examinations of the target children’s vocalizations and their active interaction with other talkers is outside of the scope of the present paper but is an active area of work by the present author team. Determining what overlapping vocalizations may be seriously degraded in children’s perception of their input (Erickson & Newman, 2017; Hall III, Grose, Buss, & Dev, 2002) is also beyond the scope of the present paper, complicated by the varying types and levels of background noise, the time spent outdoors, and the activity contexts in which overlap is embedded (e.g., two simultaneous adult conversations versus simultaneous chanting of a phrase by three children during a game). Surely excluding all overlapping talk would reduce the estimates presented here, but we are unconvinced that doing so would contribute much more to our understanding than the current data do. For research directly considering this issue of overlapping speech and its impact on input estimates, we point readers to work by Scaff and colleagues (Scaff et al., under review).

## Conclusion

Our findings revealed that, across a diverse set of cultural and linguistic contexts, the quantity of input directed to children during the first three years is both relatively low and stable across age. Overhearable adult-directed input is much more available, but preliminary evidence suggests that it decreases across age. Language group also impacts who input comes from, especially when it comes to *directed* input from other children, which is more common in some groups than others. However, women’s input predominates overall. Finally, the number of talkers who are present matters a great deal for the amount of language encountered, both target-child directed and adult-directed. These results add to a growing body of work quantifying the outsize role women’s input plays in children’s early language exposure across diverse cultural groups. It also highlights the fact that children’s relative exposure to input from other talker types—especially language from other children—is an important and understudied aspect of their early linguistic input. Finally, it underscores the importance of understanding how other aspects of everyday life drive patterns in language exposure (e.g., the number of others present), opening up pathways for future work to more precisely pinpoint the nature of these differences and their relationship to early language development.

# References

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430. https://doi.org/<https://doi.org/10.1111/1467-8624.00287>

Bateson, M. C. (1979). The epigenesis of conversational interaction: A personal account of research development. In M. Bullowa (Ed.), *Before speech* (pp. 63–79). New York, NY: Cambridge University Press.

Bell, A. (1984). Language style as audience design. *Language in Society*, *13*(2), 145–204. <https://doi.org/doi:10.1017/S004740450001037X>

Bergelson, E. (2016). *Bergelson SEEDLingS HomeBank corpus*. https://doi.org/<https://doi.org/10.21415/T5PK6D>

Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child Development Perspectives*, *14*(3), 142–149. https://doi.org/<https://doi.org/10.1111/cdep.12373>

Bergelson, E., Alphen, P. van, Benneti, L., Bunce, J., Casillas, M., Guez, A., … Cristia, A. (in revision). *Child language environments in >2500 daylong recordings across 5 continents*.

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

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*, e12724. <https://doi.org/10.1111/desc.12724>

Bergelson, E., & Swingley, D. (2018). Young infants’ word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, *89*(5), 1567–1576. https://doi.org/<https://doi.org/10.1111/cdev.12888>

Bornstein, M. H., Tal, J., Rahn, C., Galperín, C. Z., Pêcheux, M.-G., Lamour, M., … Tamis-LeMonda, C. S. (1992). Functional analysis of the contents of maternal speech to infants of 5 and 13 months in four cultures: Argentina, France, Japan, and the United States. *Developmental Psychology*, *28*(4), 593–603. https://doi.org/<https://doi.org/10.1037/0012-1649.28.4.593>

Brentari, D., González, C., Seidl, A., & Wilbur, R. (2011). Sensitivity to visual prosodic cues in signers and nonsigners. *Language and Speech*, *54*(1), 49–72. https://doi.org/<https://doi.org/10.1177/0023830910388011>

Broesch, T., & Bryant, G. A. (2015). Prosody in infant-directed speech is similar across western and traditional cultures. *Journal of Cognition and Development*, *16*(1), 31–43.

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. <https://doi.org/10.1111/cdev.12501>

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

Brown, P. (1998). Conversational structure and language acquisition: The role of repetition in Tzeltal adult and child speech. *Journal of Linguistic Anthropology*, *2*, 197–221. <https://doi.org/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. 187–226). Cambridge, UK: Cambridge University Press. <https://doi.org/10.1017/CBO9781139342872.010>

Bruner, J. (1983). *Child’s talk: Learning how to use language*. New York, NY: Norton.

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. <https://doi.org/10.1073/pnas.1309518110>

Casillas, M. (2023). Learning language in vivo. *Child Development Perspectives*, *17*(1), 10–17. https://doi.org/<https://doi.org/10.1111/cdep.12469>

Casillas, M., Bergelson, E., Warlaumont, A. S., Cristia, A., Soderstrom, M., VanDam, M., & Sloetjes, H. (2017). A new workflow for semi-automatized annotations: Tests with long-form naturalistic recordings of children’s language environments. In F. Lacerda, D. House, M. Heldner, J. Gustafson, S. Strömbergsson, & M. Włodarczak (Eds.), *Proceedings of the 18th annual conference of the international speech communication association (INTERSPEECH 2017)* (pp. 2098–2102). Stockholm, Sweden. <https://doi.org/10.21437/Interspeech.2017-1418>

Casillas, M., Brown, P., & Levinson, S. C. (2017). *Tseltal and Yelî Dnye HomeBank corpora*. <https://doi.org/doi:10.21415/T51X12>

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835. https://doi.org/<https://doi.org/10.1111/cdev.13349>

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, *48*, 792–814. <https://doi.org/doi:10.1017/S0305000920000549>

Casillas, M., & Cristia, A. (2019). A step-by-step guide to collecting and analyzing long-format speech environment (LFSE) recordings. *Collabra: Psychology*, *5*(1), 24. https://doi.org/<https://doi.org/10.1525/collabra.209>

Clark, H. H. (1996). *Using Language*. Cambridge University Press. https://doi.org/<https://doi.org/10.1017/CBO9780511620539>

Cooper, R. P., & Aslin, R. N. (1989). The language environment of the young infant: Implications for early perceptual development. *Canadian Journal of Psychology / Revue Canadienne de Psychologie*, *43*(2), 247–265. https://doi.org/<https://doi.org/10.1037/h0084216>

Cristia, A. (2020). Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition. *Developmental Review*, *57*, 100914. https://doi.org/<https://doi.org/10.1016/j.dr.2020.100914>

Cristia, A. (2023). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, *26*(1), e13265. https://doi.org/<https://doi.org/10.1111/desc.13265>

Cristia, A., Ganesh, S., Casillas, M., & Ganapathy, S. (2018). Talker diarization in the wild: The case of child-centered daylong audio-recordings. In N. Yegnanarayana, C. Chandra Sekhar, S. Narayanan, S. Umesh, S. R. M. Prasanna, H. A. Murthy, … P. Kumar Ghosh (Eds.), *Proceedings of the 19th annual conference of the international speech communication association (INTERSPEECH 2018)* (pp. 2583–2587). Hyderabad, India. <https://doi.org/10.21437/Interspeech.2018-2078>

Cristia, A., Gautheron, L., & Colleran, H. (2023). Vocal input and output among infants in a multilingual context: Evidence from long-form recordings in Vanuatu. *Developmental Science*, e13375. https://doi.org/<https://doi.org/10.1111/desc.13375>

Cychosz, M., Villanueva, A., & Weisleder, A. (2020). *Efficient estimation of children’s language exposure in two bilingual communities*. *64*(10), 3843–3866. https://doi.org/<https://doi.org/10.1044/2021_JSLHR-20-00755>

Dailey, S., & Bergelson, E. (2022). Language input to infants of different socioeconomic statuses: A quantitative meta-analysis. *Developmental Science*, *25*(3), e13192. https://doi.org/<https://doi.org/10.1111/desc.13192>

de León, L. (1998). 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. <https://doi.org/10.1002/9781444342901.ch4>

DeAnda, S., Bosch, L., Poulin-Dubois, D., Zesiger, P., & Friend, M. (2016). The language exposure assessment tool: Quantifying language exposure in infants and children. *Journal of Speech, Language, and Hearing Research*, *59*(6), 1346–1356. https://doi.org/<https://doi.org/10.1044/2016_JSLHR-L-15-0234>

Demuth, K., & Mputhi, T. S. (1979). *Introduction to Sesotho: An oral approach (with language tapes)*. Indiana University Linguistics Club. Bloomington, Indiana.

Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development. *Child Development*, *92*(2), 609–625. https://doi.org/<https://doi.org/10.1111/cdev.13511>

Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child Development*, 399–410. https://doi.org/<https://doi.org/10.2307/1130985>

Elmlinger, S. L., Goldstein, M. H., & Casillas, M. (2023). Immature vocalizations simplify the speech of Tseltal Mayan and US caregivers. *Topics in Cognitive Science*, *15*(2), 315–328. https://doi.org/<https://doi.org/10.1111/tops.12632>

Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. *Current Directions in Psychological Science*, *26*(5), 451–457. https://doi.org/<https://doi.org/10.1177/0963721417709087>

Farran, L. K., Lee, C.-C., Yoo, H., & Oller, D. K. (2016). Cross-cultural register differences in infant-directed speech: An initial study. *PloS One*, *11*(3), e0151518. https://doi.org/<https://doi.org/10.1371/journal.pone.0151518>

Fernald, A., Taeschner, T., Dunn, J., Papousek, M., Boysson-Bardies, B. de, & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers’ and fathers’ speech to preverbal infants. *Journal of Child Language*, *16*(3), 477–501. <https://doi.org/doi:10.1017/S0305000900010679>

Foushee, R., Srinivasan, M., & Xu, F. (2021). Self-directed learning by preschoolers in a naturalistic overhearing context. *Cognition*, *206*, 104415. https://doi.org/<https://doi.org/10.1016/j.cognition.2020.104415>

Gaskins, S. (1996). How Mayan parental theories come into play. In S. Harkness & C. M. Super (Eds.), *Parents’ cultural belief systems: Their origins, expressions, and consequences* (pp. 1–183). Guilford Press New York. https://doi.org/<https://www.jstor.org/stable/3196228>

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. (2006). Cultural perspectives on infant–caregiver interaction. In N. J. Enfield & S. C. Levinson (Eds.), *Roots of Human Sociality: Culture, Cognition and Interaction* (pp. 279–298). Oxford: Berg.

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. <https://doi.org/10.1177/1525740110367826>

Hall III, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee recognition in a two-talker masker and a speech-shaped noise masker in adults and children. *Ear and Hearing*, *23*(2), 159–165.

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. <https://doi.org/10.1017/S0140525X10000725>

Hillairet de Boisferon, A., Dupierrix, E., Quinn, P. C., Lœvenbruck, H., Lewkowicz, D. J., Lee, K., & Pascalis, O. (2015). Perception of multisensory gender coherence in 6-and 9-month-old infants. *Infancy*, *20*(6), 661–674. https://doi.org/<https://doi.org/10.1111/infa.12088>

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., … Mehr, S. A. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, *6*, 1545–1556. https://doi.org/<https://doi.org/10.1038/s41562-022-01410-x>

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. <https://doi.org/10.3389/fpsyg.2015.01492>

Holler, J., Alday, P. M., Decuyper, C., Geiger, M., Kendrick, K. H., & Meyer, A. S. (2021). Competition reduces response times in multiparty conversation. *Frontiers in Psychology*, *12*, 693124. https://doi.org/<https://doi.org/10.3389/fpsyg.2021.693124>

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. https://doi.org/<https://doi.org/10.1037/0096-1523.26.5.1570>

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. <https://doi.org/10.1111/infa.12222>

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2022). Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language*, *49*(6), 1173–1192. <https://doi.org/doi:10.1017/S0305000921000623>

Mannle, S., Barton, M., & Tomasello, M. (1992). Two-year-olds’ conversations with their mothers and preschool-aged siblings. *First Language*, *12*(34), 57–71. https://doi.org/<https://doi.org/10.1177/014272379201203404>

Marasli, Z., & Montag, J. L. (2023). Optimizing random sampling of daylong audio. In M. Goldwater, F. Anggoro, B. Hayes, & D. Ong (Eds.), *Proceedings of the 44th Annual Meeting of the Cognitive Science Society* (pp. 87–94). https://doi.org/<https://escholarship.org/uc/item/0478r9pb>

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis. *Psychological Science*, *26*(3), 341–347. https://doi.org/<https://doi.org/10.1177/0956797614562453>

Masataka, N. (1998). Perception of motherese in Japanese sign language by 6-month-old hearing infants. *Developmental Psychology*, *34*(2), 241–246. https://doi.org/<https://doi.org/10.1037/0012-1649.34.2.241>

Masek, L. R., Ramirez, A. G., McMillan, B. T., Hirsh-Pasek, K., & Golinkoff, R. M. (2021). Beyond counting words: A paradigm shift for the study of language acquisition. *Child Development Perspectives*, *15*(4), 274–280. https://doi.org/<https://doi.org/10.1111/cdep.12425>

McDivitt, K., & Soderstrom, M. (2016). *McDivitt HomeBank corpus*. <https://doi.org/doi:10.21415/T5KK6G>

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, *24*(6), e13122. https://doi.org/<https://doi.org/10.1111/desc.13122>

Meylan, S. C., & Bergelson, E. (2022). Learning through processing: Toward an integrated approach to early word learning. *Annual Review of Linguistics*, *8*(1), 77–99. <https://doi.org/10.1146/annurev-linguistics-031220-011146>

Micheletti, M., Barbaro, K. de, Fellows, M. D., Hixon, J. G., Slatcher, R. B., & Pennebaker, J. W. (2020). Optimal sampling strategies for characterizing behavior and affect from ambulatory audio recordings. *Journal of Family Psychology*, *34*(8), 980–990. https://doi.org/<https://doi.org/10.1037/fam0000654>

Montag, J. L. (2020). New insights from daylong audio transcripts of children’s language environments. In S. Denson, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society* (pp. 3005–3011).

Ochs, E., & Kremer-Sadlik, T. (2020). Ethical blind spots in ethnographic and developmental approaches to the language gap debate. *Langage Et Société*, (2), 39–67.

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.

Oshima-Takane, Y., Goodz, E., & Derevensky, J. L. (1996). Birth order effects on early language development: Do secondborn children learn from overheard speech? *Child Development*, *67*(2), 621–634. https://doi.org/<https://doi.org/10.1111/j.1467-8624.1996.tb01755.x>

Piot, L., Havron, N., & Cristia, A. (2022). Socioeconomic status correlates with measures of Language Environment Analysis (LENA) system: A meta-analysis. *Journal of Child Language*, *49*(5), 1037–1051. <https://doi.org/doi:10.1017/S0305000921000441>

Pisani, S., Gautheron, L., & Cristia, A. (in prep). *Long-form recordings: From A to Z*. Retrieved from <https://bookdown.org/alecristia/exelang-book/>

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, *13*(1), 85–100. <https://doi.org/10.1017/S0305000900000313>

R Core Team. (2019). *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. (2017a). Look who’s talking NOW! Parentese speech, social context, and language development across time. *Frontiers in Psychology*, *8*, 1008. https://doi.org/<https://doi.org/10.3389/fpsyg.2017.01008>

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017b). The impact of early social interactions on later language development in Spanish-English bilingual infants. *Child Development*, *88*(4), 1216–1234. https://doi.org/<https://doi.org/10.1111/cdev.12648>

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

Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A., & Casillas, M. (2021). ALICE: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, *53*, 818–835. https://doi.org/<https://doi.org/10.3758/s13428-020-01460-x>

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. <https://doi.org/10.1146/annurev.psych.54.101601.145118>

Rosemberg, C. R., Alam, F., Audisio, C. P., Ramirez, M. L., Garber, L., & Migdalek, M. J. (2020). Nouns and verbs in the linguistic environment of Argentinian toddlers: Socioeconomic and context-related differences. *First Language*, *40*(2), 192–217. https://doi.org/<https://doi.org/10.1177/0142723719901226>

Rosemberg, C. R., Alam, F., Stein, A., Migdalek, M. J., Menti, A., & Ojea, G. (2015–2016). Los entornos lingüísticos de niñas y niños pequeños argentinos/language environments of young Argentinean children (dataset).

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. <https://doi.org/10.1017/S0305000907008343>

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development*, *83*(5), 1762–1774. https://doi.org/<https://doi.org/10.1111/j.1467-8624.2012.01805.x>

Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of Developmental Psychology*, *2*, 201–223. https://doi.org/<https://doi.org/10.1146/annurev-devpsych-042220-121816>

Rowland, C. F., Bidgood, A., Durrant, S., Peter, M., & Pine, J. M. (2018). The Language 0–5 project.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. *Language*, *50*(4), 696–735. https://doi.org/<https://doi.org/10.1353/lan.1974.0010>

Scaff, C., Casillas, M., Stieglitz, J., & Cristia, A. (under review). *Characterization of children’s verbal input in a forager-farmer population using long-form audio recordings and diverse input definitions*.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232. https://doi.org/<https://doi.org/10.1016/0010-0285(89)90008-X>

Shneidman, L. A. (2010). *Language Input and Acquisition in a Mayan Village* (PhD thesis). The University of Chicago.

Shneidman, L. A., Gaskins, S., & Woodward, A. (2016). Child-directed teaching and social learning at 18 months of age: Evidence from Yucatec Mayan and US infants. *Developmental Science*, *19*(3), 372–381. https://doi.org/<https://doi.org/10.1111/desc.12318>

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

Singh, L., Cristia, A., Karasik, L. B., Rajendra, S. J., & Oakes, L. M. (2023). Diversity and representation in infant research: Barriers and bridges toward a globalized science of infant development. *Infancy*, *28*(4), 708–737. https://doi.org/<https://doi.org/10.1111/infa.12545>

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

Snedeker, J., Geren, J., & Shafto, C. L. (2007). Starting over: International adoption as a natural experiment in language development. *Psychological Science*, *18*(1), 79–87. https://doi.org/<https://doi.org/10.1111/j.1467-9280.2007.01852.x>

Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Warlaumont, A. S., Bunce, J., et al. (2021). Developing a cross-cultural annotation system and MetaCorpus for studying infants’ real world language experience. *Collabra: Psychology*, *7*(1), 23445. https://doi.org/<https://doi.org/10.1525/collabra.23445>

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. <https://doi.org/10.1371/journal.pone.0080646>

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019). Language does matter: But there is more to language than vocabulary and directed speech. *Child Development*, *90*(4), 1303–1318. https://doi.org/<https://doi.org/10.1111/cdev.13072>

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. <https://doi.org/10.1177/0142723715596647>

Warlaumont, A. S., Pretzer, G. M., Mendoza, S., & Walle, E. A. (2016). *Warlaumont HomeBank corpus*. <https://doi.org/doi:10.21415/T54S3C>

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. Retrieved from <https://ggplot2.tidyverse.org>

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani, J. Odijk, & D. Tapias (Eds.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (pp. 1556–1559). https://doi.org/<https://aclanthology.org/L06-1>

Wurm, L. H., & Fisicaro, S. A. (2014). What residualizing predictors in regression analyses does (and what it does not do). *Journal of Memory and Language*, *72*, 37–48. https://doi.org/<https://doi.org/10.1016/j.jml.2013.12.003>

1. The typical number of talkers present may vary systematically between populations (see Supplementary Materials Section 2 and Brown, 2011, 2014; Gaskins, 2006; Rosemberg et al., 2020). Therefore the number-of-talkers measure may partly capture cultural differences—not just within-participant activity variation. We suspect that, even for children in talker-dense populations, variation in the number of talkers present impacts how much CDS and ADS is observable at a given moment. We thus add this factor as *separate* from culture in our statistical models. [↑](#footnote-ref-22)
2. Alternatively, one could comprehensively annotate and analyze children’s daylong input (Montag, 2020), but manually annotating CDS at the utterance level in this way is a many-years-long undertaking. [↑](#footnote-ref-24)
3. To simplify the structure of our results, we treat the largest sub-corpus—North American English—as the reference level for language group and the most studied talker type—women—as the reference-level for talker type in the analyses below. We in no way imply with this decision that the North American English data are a global “norm”. Rather, we mean to highlight which results are likely to have gone under-reported in past work. The Supplementary Materials include alternate models with all language groups as reference level to allow for further inspection and full pairwise comparisons between language groups. [↑](#footnote-ref-26)
4. For more information on the caregiving and early language environments of Argentinian Spanish-, Tseltal-, and Yélî Dnye-acquiring children, we refer readers to other work (Brown, 1998, 2011, 2014; Brown & Casillas, in press; Rosemberg et al., 2020) [↑](#footnote-ref-29)
5. Note also that while it is a somewhat common practice to exclude naptime from consideration in analyses of longform audio recording (e.g. Bergelson, Amatuni, et al. (2019); Ramírez-Esparza et al. (2014)), naptime is not a culturally appropriate construct in some of our sampled populations. [↑](#footnote-ref-39)
6. While we do have indicators of maternal education for the recordings used here, we are unconvinced that this indicator is similarly meaningful across such economically and culturally diverse populations and so we do not use it in our analyses. [↑](#footnote-ref-65)