A cross-cultural examination of young children's everyday language experiences

John Bunce[1,2], Melanie Soderstrom[2], Elika Bergelson[3], Celia Rosemberg[4], Alejandra Stein[4], Florencia Alam[4], Maia Julieta Migdalek[4], & Marisa Casillas[5,6]

[1] California State University, East Bay

[2] University of Manitoba

[3] Duke 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, 563 869-2016-0003) and by 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

Abstract

We present an exploratory cross-cultural analysis of the quantity of target-child-directed speech and adult-directed speech to young children learning 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-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 at any given moment 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 societies.

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

Word count: 10,206—11,085 (including tables and captions)

A cross-cultural 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, 2014, 2017a, 2017b). 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 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 understudied cultural contexts, where other language phenomena (e.g., rhetorical practices) may prove more relevant (Ochs & Kremer-Sadlik, 2020; Sperry, Miller, & Sperry, 2015).

CROSS-CULTURAL EVERYDAY LANGUAGE EXPERIENCES

Recent cross-linguistic and cross-cultural work on typically developing children supports the idea that there is 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, 2023). Relatedly, Casillas and colleagues (Casillas et al., 2020, 2021) found surprisingly similar, and 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)). 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 Alam, Ramirez, and Migdalek (2021), Cristia, Gautheron, and Colleran (2023), and Loukatou, Scaff, Demuth, Cristia, and Havron (2022). 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.

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, Ramírez-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 appears 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 reported to be 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 young children observe (passively or actively), CDS produced by other children, and multimodal and multiparty interactions (Alam et al., 2021; Casillas et al., 2020, 2021; Cristia et al., 2018, 2023; de León, 1998; de León & García-Sánchez, 2021; Hou, 2024; Loukatou et al., 2022; Scaff et al., 2023).

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-linguistic group influence the quantities of the directly addressed and overhearable adult speech children encounter in five distinct settings. Before we dive into 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 speech 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; finally, we describe prior approaches to estimating these input sources 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. In the first, these two terms ("CDS" and "ADS") are used to denote the intended addressee, i.e. child vs. adult, respectively. In the second approach, these terms denote the speech *register* or other characteristics of the speech, regardless of actual addressee. That is, speech that contains the prosodic, lexical, grammatical, and affective characteristics typically associated with speech to children is classified as child-directed speech, regardless of who was being spoken to. In the present study, we will measure linguistic input quantities based on the first approach: 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 vital to consider when constructing comparative theories of child language development (e.g., Bornstein et al., 1992; Broesch, T, Rochat, Olah, Broesch, J. & Henrich, 2016; Brown, 2014; de León & García-Sánchez, 2021; Masek, Ramirez, McMillan, Hirsh-Pasek, & Golinkoff, 2021; Ochs & Schieffelin, 1984; Pye, 2017), input quantities are ideal for roughly comparing the linguistic material children encounter in their daily lives. Moreover, input quantity estimates that are centered specifically on *directed* vs. *non-directed* speech can capture some aspects of input "quality". Addressees have an advantage comprehending conversational talk addressed to them over talk addressed to others, precisely because the conversational talk in question is tailored specifically to the addressees' immediate comprehension (Bell, 1984; Foushee, Srinivasan, & Xu, 2021; Schober & Clark, 1989). Thus, general (and likely universal) mechanisms of human coordination (Clark, 1996) predict that child-addressed speech is a referentially clearer linguistic signal for the child learner than adult-directed speech.

In the present study, we compare adult-directed speech (ADS) quantities and target-child directed speech (TCDS) quantities, the latter being speech addressed *specifically* to the child under study, rather than to another nearby child. These measures represent two, qualitatively distinct sources of linguistic input; 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 child rearing (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, regarding socioeconomic group, 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 cf. Bergelson et al., 2023).

Regarding cultural group, some prior work found no evidence for difference in baseline TCDS rate between Tseltal- and Yélî Dnye-speaking children under age three, despite clear ethnographic evidence that adults in these two communities take very different approaches to talking to infants and young children ("non-child-centric" vs. "child-centric" input environments; Brown, 2011, 2014; Brown & Casillas, in press; Casillas et al., 2020, 2021). In contrast, Shneidman and Goldin-Meadow (2012) *did* observe clear differences in US and Yucatec Mayan children's input quantities, with the US children under age three hearing significantly more directed input. This evidence concords with Cristia's (2023) characterization of the primary split in input quantities being in rural versus urban populations, rather than culture-to-culture differences. Complementing this work on input quantity, studies of input quality consistently show clear 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, 2023; Stein et al., 2021; Vogt et al., 2015).

In the meso part of the spectrum, children's age and available interactants may also shape input quantities. Regarding age, prior work does not consistently demonstrate evidence of change in CDS quantity with child age but does demonstrate age-related change for other input sources, including ADS and non-canonical CDS (Bergelson, Casillas, et al., 2019; Casillas et al., 2020, 2021; Ramírez-Esparza et al., 2017b; see also Shneidman and Goldin-Meadow, 2012). Regarding available interactants, prior work points to a greater availability of CDS from adults compared to children—and, among adults, from 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 the high prevalence of peer-produced CDS (Alam et al., 2021; Casillas et al., 2020, 2021; Cristia et al., 2023; Loukatou et al., 2022;; Scaff et al., 2023).

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/awake, stationary/in motion), and more. Soderstrom and colleagues (Soderstrom & Wittebolle, 2013; Soderstrom et al., 2018) found that linguistic input rates systematically varied depending on the activity context and number of adults present in Canadian daylong recordings (see also Casillas et al. (2020, 2021), Greenwood, Thiemann-Bourque, Walker, Buzhardt, and Gilkerson (2011), and Rosemberg et al. (2020, 2023)). Though we will not have information about activity context in the present work, we will at least be able to account for the number of individual talkers present. 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 smaller, simultaneous conversations, multiplying the amount of observable talk (conversational “schism” see, e.g., Sacks, Schegloff, and Jefferson (1978) 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 interactional contexts that involve denser family participation (e.g., in overlapping, co-present subgroups) versus contexts where smaller groups of individuals are on their own.[1]

Extracting CDS and ADS from daylong recordings

Ecologically valid estimates of speech 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 (2021) 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 (2014, 2017a, 2017b) 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 audio-clip level to the utterance level, Bergelson, Casillas, and colleagues (2019) extracted individual utterances using LENA's automated utterance annotations

[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 random within-participant 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 cultural group in our statistical models.

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 interactional contexts across their day (including typical "down" time), we must take random or periodic samples of the language environment (Casillas & Cristia, 2019; see also Alam et al., 2021; Rosemberg et al., 2020, 2023; Stein et al., 2021) 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]

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

[2] Alternatively, one could comprehensively annotate and analyze children's daylong input (Montag, 2020), but manually annotating input at the utterance level in this way is a many-years-long undertaking (e.g. at a representative work-to-recording ratio of 60 minutes:1 minute, that would be roughly 43,839 hours for the current sample of 69 children; approximately 23 full-time work years).

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.[3] 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 using OCDS in the Supplementary Materials, where we combine it

[3] While the data collectively represent a linguistically and culturally diverse array of early language experiences, one might still consider the present data to be using a convenience sample of pre-existing, available recordings.

with TCDS to generate parallel analyses of all-CDS (TCDS + OCDS), with similar results to what is reported below.

Exploratory Hypotheses. Following from the findings summarized above, the specific aims of our analysis were to examine how TCDS and ADS varied across age, language group, talker type, and number of talkers in a highly naturalistic and culturally diverse set of daylong recordings. To the prior literature, the present work adds an apples-to-apples comparative view on these effects, given that each of the included corpora recorded, sampled, and annotated children's input in highly similar ways (Soderstrom et al., 2021). Before analysis, we established a specific set of exploratory hypotheses—with corresponding regression formulae—regarding TCDS and ADS. We term them "exploratory" here because of the present sample size—while the sample is large relative to prior comparable work, each corpus includes data from only 9–10 children (Table 1). These hypotheses were slightly different for TCDS and ADS, based on prior work (see Tables 2 and 3 for detailed overviews).

Target-child-directed speech hypotheses. Based on the work cited above, we expected TCDS rate to vary across language groups (e.g., to be higher 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 produce more talk.

Adult-directed speech hypotheses. We expected for ADS rate to vary across language groups (e.g., to be lower 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.

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 (see also Bergelson et al. (2023) for a LENA-based, comparative mega-analysis of nearby adult input). We examine these factors in order to identify axes of consistency and variation across the multi-corpus sample. While similar analyses have been previously conducted on two individual corpora (Tseltal and Yélî Dnye; Casillas et al., 2020; 2021), the current study offers a first view into baseline TCDS and ADS rates at home in North American English, UK English, and Argentinian Spanish while simultaneously providing a comparative perspective on how each of the proposed factors—group, child age, and interactants—influences children's input rates.

To reiterate, we do not attempt to examine all possible interactions, and instead take a hypothesis-driven approach to analysis.[4] Importantly, while we identify key points of

[4] To simplify the structure of our results, we treat the largest 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.

theoretically relevant 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.[5] 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 duration, target child ages (see Supplementary Materials Section 3), and other demographic characteristics (see Table 1 for an overview).

[5] 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).

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.

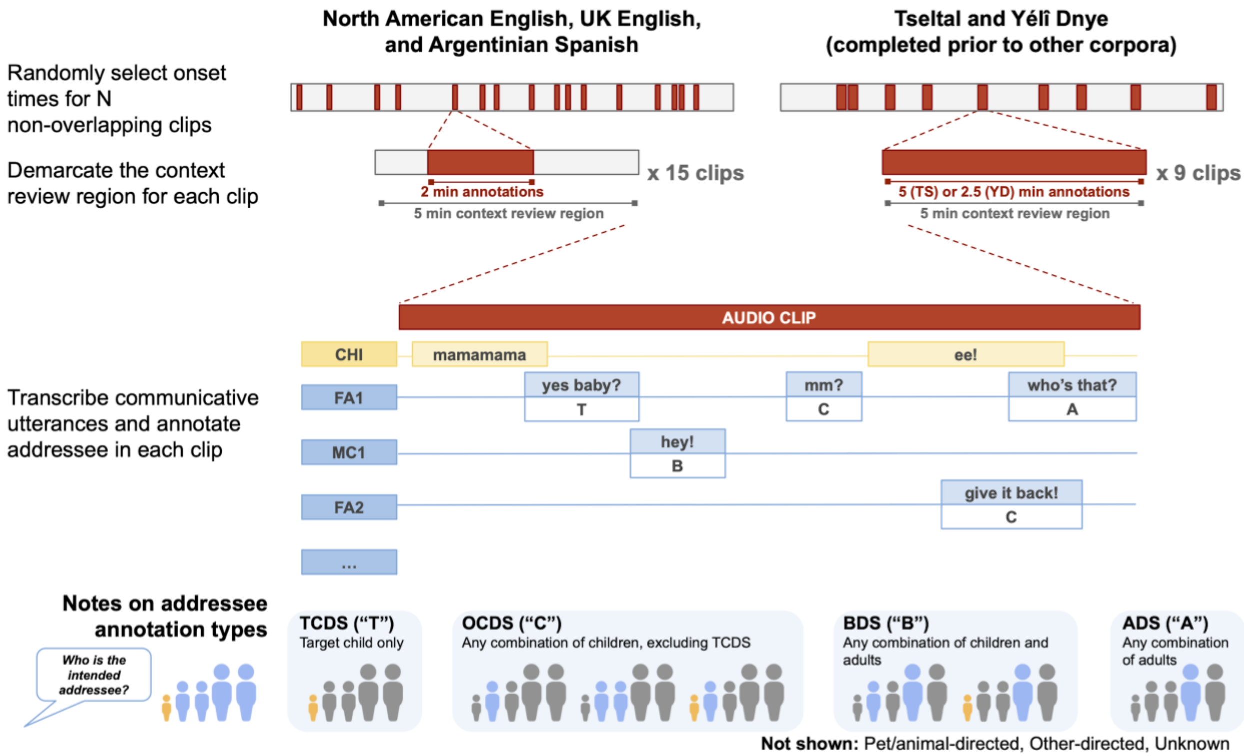

Figure 1. Summary of clip selection and annotation method across corpora.

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 three-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.

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 a 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* in seven categories—exclusively target child, non-target child, adult, mixed-age, animal, other, 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 CDS having 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 were not always native speakers of 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 1: *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	N	Recording Lengths
Bergelson	North American English	11.2 mo (7–17)	4.2 (3–7)	Northeast (US)	10 (4 males)	802.8 min (665–960)
McDivitt Winnipeg	North American English	12.4 mo (2–32)	3.3 (2–5)	Winnipeg Manitoba (Canada)	9 (4 males)	532.6 min (118–837)
Warlaumont	North American English	6.3 mo (3–9)	Not available	Central Valley California (US)	10 (5 males)	815.8 min (612–960)
LuCiD Lang0-5	UK English	20 mo (11–31)	3.6 (3–5)	North West (England)	10 (5 males)	931.5 min (845–960)
Rosemberg	Argentinian Spanish	17.1 mo (9–27)	5.5 (3–12)	Buenos Aires (Argentina)	10 (6 males)	242 min (125–303)
Casillas Tseltal	Tseltal (Mayan)	16.1 mo (2–36)	7 (4–14)	Chiapas (Mexico)	10 (5 males)	554.5 min (492–578)
Casillas Yélî Dnye	Yélî Dnye (Papuan)	14.2 mo (2–36)	6.5 (3–11)	Milne Bay Province (Papua New Guinea)	10 (5 males)	488.2 min (430–555)
Total		*13.9 mo (2–36)*	*5.1 (2–14)*		*69 (34 males)*	*625.2 min (118–960)*

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[6] 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** (indicated by “ziformula” in the model formulae), which uses a logistic

---

[6] 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.

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 dummy-coded factorial variables, which limited 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 were 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, *pairwise* comparisons between language groups may also be of interest to readers. For those curious about 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., four additional 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 2: *Predictions for TCDS analysis. Asterisk indicates previously observed with daylong child language data (Casillas et al., 2020, 2021; Scaff et al., 2023). 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 (null result)
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	Yes, but for Yélî Dnye only
Talker type x Corpus	More TCDS from children in Argentinian Spanish, Tseltal, and Yélî Dnye; no specific predictions regarding men	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; no specific predictions regarding men	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 3: *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, but a significant outcome depends on reference level choice
Talker type	Women > men, children*	Yes, but a significant outcome depends on reference level choice
# 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—UK English only
Target child age x Corpus	Bigger ADS decrease in Argentinian Spanish, Tseltal, and Yélî Dnye	No
Zero-inflation model		
Target child age	More zero-ADS clips for older children	(Not tested 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 mean 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.

As a reminder, we report results from the models of TCDS and ADS 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.

Table 4: *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

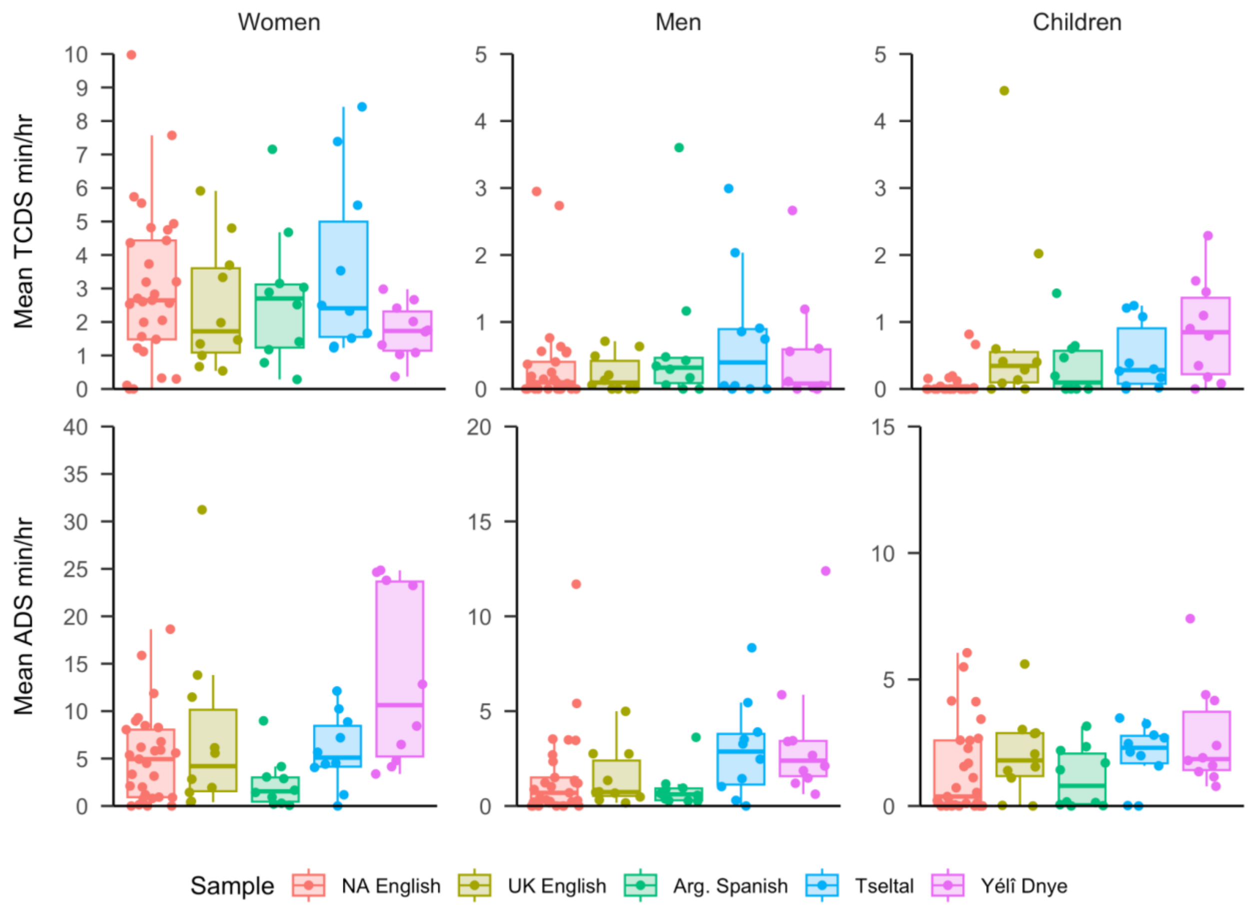

Figure 2. Mean by-recording rates of TCDS (above) and ADS (below) min/hr rates across language groups and talker types. For example, the upper-leftmost datapoint shows a recording with an average of 10 minutes per hour of TCDS from women talkers in North American English. Left-to-right order of language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

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 = 0.33, SE = 0.04, z = 7.62, p < 0.001), 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).

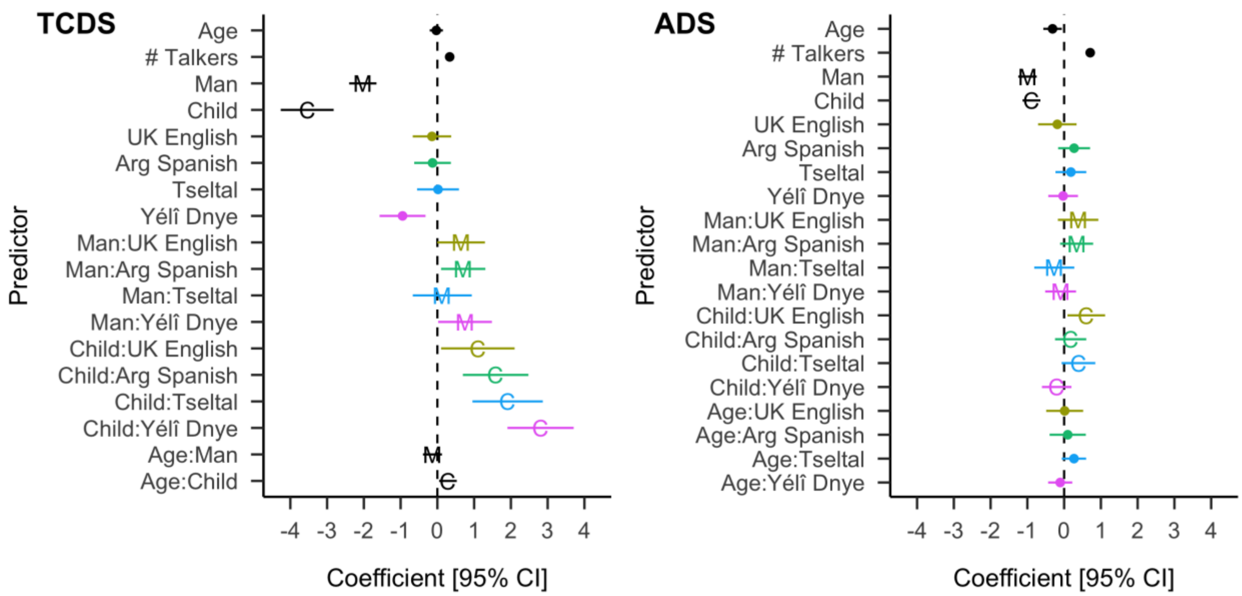

Figure 3. Coefficients with 95% confidence intervals from the count models of TCDS (left) and ADS (right) for all included fixed effects in the model with NA English and women set as the reference levels for language group and talker type, respectively. Intervals not overlapping with zero indicate significance. 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).

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 a subset of 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, 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, as opposed to direct changes to their linguistic input (see Bergelson, 2020 for a review). Rather than attributing development to changes in the input, this theoretical approach looks instead to growth 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 a preliminary but important cross-linguistic datapoint: this basic idea may hold 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. The rate of TCDS produced by men was significantly higher in Argentinian Spanish and Yélî Dnye compared to North American English. The rate of TCDS produced by children was significantly higher 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. We are cautious in interpreting this lack of evidence for age-related change within speaker types given our limited sample size.

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., 1978; Stein et al., 2021), we found that more talkers leads to more input, both for TCDS and ADS. This effect is due in part to the simple fact that, all else being equal, the presence of more talkers leads to more talk. The presence of more talkers increases competition for the floor

(Holler et al., 2021; Sacks et al., 1978). 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 interactants to break off into smaller conversations (e.g., two, two-person conversations), potentially doubling the observable talk via overlapping conversation in the input (Sacks et al., 1978). Future work might selectively examine subsets of daylong data to more precisely characterize how interactionally driven factors such as the number of talkers present accounts for fluctuations in a child's linguistic input rate, both within and across language groups. Doing so will likely require more transcribed interactions than the current dataset offers.

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 individually addressed child 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), Foushee et al. (2021), 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 (e.g., de León, 1998; de León & Gárcia-Sánchez, 2021), 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)). As we discuss in more detail below, an important addendum to this discussion is that systematic cross-cultural differences in multi-party interaction can *also* be understood as cultural—not solely situation-specific.

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 lower rates of TCDS and higher rates of ADS compared to North American English-acquiring children. By and large, our data show little evidence for 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. Beyond the concern of statistical power—which is relevant with our relatively small samples—it is essential to think again about where differences between groups could come from. In particular, it's worth re-examining how to understand the effect of the number of talkers present: could this be group-specific behavior or not?

We observed that many of the 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[7]. 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 group-specific vs. not. 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

[7] I.e., formula = min.p.hr ~ child.age + lang.grp + (1 | child.id), ziformula = ~ child.age + lang.grp.

more discussion of these issues). Put differently, variation in the number of talkers present can signal of culture-specific routines, practices, and interactional contexts. 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 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, et al., 2016; de León, 1998; Kuchirko, Tafuro, & Tamis-LeMonda, 2018; Montag, 2020; Rowe, 2012; Rosemberg et al., 2020, 2023). 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 post-industrial 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 and in transcribers and transcription due to the logistical constraints in doing annotation across linguistically and culturally diverse corpora—this minor variation is inevitable in comparative work on naturalistic interaction. We carefully considered these minor deviations, and have no reason to believe that they 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. Given the diversity of populations in our metacorpus, random sampling was also 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 yields accurate average word count estimates, with varying but symmetrical rates of error depending on clip duration (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); see also Räsänen, Seshadri, Lavechin, Cristia, and Casillas (2021) for evidence from the specific 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.

Indeed, while we may have sampled sufficiently within recordings to create stable estimates for each child, the present analyses would be more powerful if done over more recordings, with a greater number of cultural and linguistic contexts, and/or with a more systematic or theory-driven selection of cultural and linguistic contexts that exist in today's world. These are persistent problems in the field of developmental science (e.g., Kosie & Lew-Williams, 2022; Oakes, 2017; Singh et al., 2023) and so, as usual, null effects should be taken as preliminary.

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 visible in quantitative measures of children's language environment in ways that allow us to identify axes of cross-cultural variation that are relevant for developing generalizable theories of language learning. By looking deeper at the local context for each dataset, we could better understand variation within each. For example, the construct of "socioeconomic status" is so different between these communities that it is hard to imagine a meaningful way of directly comparing between groups. Instead, within-population analyses that take into account individual and collective power within social hierarchies and relevant local institutions seem much more likely to shed light on socioeconomic effects across these diverse groups.[8] We thus strongly urge readers to take caution in generalizing our results (or those of other researchers) beyond the current data, to new populations.

[8] 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.

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., 2023). 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 playing 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 (2023).

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 our preliminary evidence suggests that it decreases across age. Language group also impacts who input is likely to come from, especially when it comes to *directed* input from other children, which is more common in some groups than others. That said, 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 and linguistic 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.

Alam, F., Ramírez, L., & Migdalek, M. (2021) Other children's words in the linguistic environment of infants and young children from distinct social groups in Argentina (Las palabras de otros niños en el entorno lingüístico de bebés y niños pequeños de distintos grupos sociales de Argentina). *Journal for the Study of Education and Development*, *44*(2), 269-302, DOI: 10.1080/02103702.2021.1888489

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.

Bergelson, E. (2016). Bergelson SEEDLingS HomeBank corpus. *Doi*, *10*, T5PK6D.

Bergelson, E. (2020). The comprehension boost in early word learning: Older infants are better learners. *Child Development Perspectives*, *14*(3), 142–149.

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., Soderstrom, M., Schwarz, I., Rowland, C. F., Ramirez-Esparza, N., Hamrick, L., … Cristia, A. (2023). Everyday language input and production in 1001 children from 6 continents. *PNAS*, *120*(52), e2300671120.

Bergelson, E., & Swingley, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, *89*(5), 1567–1576.

Bornstein, M. H., Tal, J., Rahn, C., Galperin, C. Z., Pecheux, 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.

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.

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*.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *91*(5), 1819–1835.

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, *48*, 792–814.

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/10.1525/collabra.209

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Cooper, R. P., & Aslin, R. N. (1989). The language environment of the young infant: Implications for early perceptual development. *Canadian Journal of Psychology*, *43*(2), 247–265.

Cristia, A. (2020). Language input and outcome variation as a test of theory plausibility: The case of early phonological acquisition. *Developmental Review*, *57*, 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.

Cristia, A., Ganesh, S., Casillas, M., & Ganapathy, S. (2018). Talker diarization in the wild: The case of child-centered daylong audio-recordings. 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 anuatu. *Developmental Science*, e13375.

Cychosz, M., Villanueva, A., & Weisleder, A. (2020). *Efficient estimation of children's language exposure in two bilingual communities*.

Dailey, S., & Bergelson, E. (2022). Language input to infants of different socioeconomic statuses: A quantitative meta-analysis. *Developmental Science*, *25*(3), e13192.

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

de León, L. & García-Sánchez, I. M. (2021). Language Socialization at the Intersection of the Local and the Global: The Contested Trajectories of Input and Communicative Competence. *Annual Review of Linguistics*, *7*, 421–448. https://doi.org/10.1146/annurev-linguistics-011619-030538

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.

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

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.

Dunn, J., & Shatz, M. (1989). Becoming a conversationalist despite (or because of) having an older sibling. *Child Development*, 399–410.

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.

Erickson, L. C., & Newman, R. S. (2017). Influences of background noise on infants and children. *Current Directions in Psychological Science*, *26*(5), 451–457.

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.

Foushee, R., Srinivasan, M., & Xu, F. (2021). Self-directed learning by preschoolers in a naturalistic overhearing context. *Cognition*, *206*, 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.

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.

Hilton, C. B., Moser, C. J., Bertolo, M., Lee-Rubin, H., Amir, D., Bainbridge, C. M., et al.others. (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behaviour*, 1–12.

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

Hou, L. (2024). Giving oranges and puppies: Children's production of directional verbs in an emerging sign language from Oaxaca. *First Language*, *0*, 1–22. https://doi.org/10.1177/01427237231221886

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.

Kosie, J. & Lew-Williams, C. (2022). Open science considerations for descriptive research in developmental science. *Infant and Child Development*, e2377. https://doi.org/10.1002/icd.2377

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.

Mannle, S., Barton, M., & Tomasello, M. (1992). Two-year-olds' conversations with their mothers and preschool-aged siblings. *First Language*, *12*(34), 57–71.

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*.

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.

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.

McDivitt, K., & Soderstrom, M. (2016). *McDivitt HomeBank corpus*.

Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, *24*(6), e13122.

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*.

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).

Oakes, L. M. (2017). Sample size, statistical power, and false conclusions in infant looking-time research. *Infancy*, *22*(4), 436–469. https://doi.org/10.1111/infa.12186

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.

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.

Pisani, S., Gautheron, L., & Cristia, A. (2021). *Long-form recordings: From A to Z*. Retrieved from https://bookdown.org/alecristia/exelang-book/. DOI: 10.5281/zenodo.6685828

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. (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*, 880–891. https://doi.org/10.1111/desc.12172

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/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.

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.

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.

Rosemberg, C. R., Alam, F., Ramirez, M. L., Ibañez, M. I. (2023). Activity Contexts and Child-Directed Speech in Socioeconomically Diverse Argentinian Households. *International Journal of Early Childhood*, *55*, pp. 1–25. https://doi.org/10.1007/s13158-022-00345-8

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

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.

Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of Developmental Psychology*, *2*, 201–223.

Rowland, C. F., Bidgood, A., Durrant, S., Peter, M., & Pine, J. M. (2018). The language 0-5 project. *Unpublished Manuscript, University of Liverpool. Doi*, *10*.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7–55). Elsevier.

Scaff, C., Casillas, M., Stieglitz, J., & Cristia, A. (2023). Characterization of children's verbal input in a forager-farmer population using long-form audio recordings and diverse input definitions. *Infancy*, *EarlyView*, 1–20.

Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, *21*(2), 211–232.

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.

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*.

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.

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*.

Soderstrom, M., Grauer, E., Dufault, B., & McDivitt, K. (2018). Influences of number of adults and adult: child ratios on the quantity of adult language input across childcare settings. *First Language*, *38*(6), 563-581.

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., Miller, P., & Sperry, L. (2015). Is there really a word gap. *American Anthropological Association Annual Meeting, Denver, CO*.

Stein, A., Menti, A. B., & Rosemberg, C. R. (2021) Socioeconomic status differences in the linguistic environment: a study with Spanish-speaking populations in Argentina, *Early Years*, *43*(1), 31-45, DOI: 10.1080/09575146.2021.1904383

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*.

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. *Proceedings of the Fifth International Conference on Language Resources and Evaluation*, 1556–1559.

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

Supplementary Materials: A cross-cultural examination of young children's everyday language experiences

Before describing the contents of these Supplementary Materials, we remind readers that our analyses are centered on the dependent variables of minutes-per-hour rate of target-child-directed speech (TCDS), other-child-directed speech (OCDS), all child-directed speech (CDS, derived from TCDS + OCDS), and adult-directed speech (ADS). Please see the main manuscript for reasoning and details. Below we first briefly describe the contents of each sub-section:

Section 1. For those interested in knowing about the *total quantity* of child-directed speech in these children's environments (i.e., all input that is designed for children addressees; TCDS + ODS), the first section includes analyses of all child-directed speech (CDS) that parallel what is reported in the main text for target-child-directed speech (TCDS).

Section 2. The second section gives expanded analyses on the number of talkers present. The main manuscript demonstrates strong effects of the number of talkers present in a given clip and motivates the inclusion of number of talkers in the primary statistical models. This section gives more descriptive information about number of talkers typical in each language group and preliminarily explores how differences in typical number of talkers present may account for main-text patterns in different input source types across language groups and target child age.

Section 3. The third section shows the distribution of target child age by corpus and delves further into discussion of the lack of simple age effects in the primary analyses.

Section 4. The fourth section uses a set of alternative models of the three dependent variables—TCDS, CDS, and ADS—to examine cross-group differences in these input sources that are naïve to effects of number of talkers and talker type, effects which may partly reflect cultural patterns.

Section 5. The fifth section breaks up the North American English language group into individual corpora for those interested in examining potential differences in future work.

Section 6. The sixth section gives tabular regression outputs for the full binomial mixed-effects regression models of TCDS, CDS, and ADS from which the main-text and Section 1 of these Supplementary Materials are reported. Also provided are the full suite of alternative models for TCDS and ADS, in which we run one model each for all possible reference levels of language group.

Section 7. The seventh section shows a marginal means plot of model-estimated rates of TCDS and ADS rates across language group and age given that the main-text plot illustrates raw data with no age effects.

Section 8. The eighth and final section shows confusion matrices for addressee-type annotations (e.g., target-child versus other-child status of an utterance) overall and for each contributed corpus individually.

As in the main text for this study, 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 anonymized data are available at URL_MASKED_FOR_REVIEW.

Table 1: *Average input rates per clip across participants for each corpus. Parentheses following the mean indicate the median and range across participants. OCDS indicates rate of input directed to non-target-child children; CDS sums rates of TCDS and OCDS.*

Language	TCDS rate	OCDS rate	ADS rate	Mean proportion TCDS	Mean proportion any CDS
NA English	3.49 (3.24; 0-10.12)	1.56 (0.48; 0-8.98)	8.06 6.86; 0-19.32)	0.27	0.39
UK English	3.69 (3.72; 1.22-7.15)	1.74 (1.05; 0-6.45)	4.38 (4.42; 0.6-9.59)	0.38	0.55
Arg. Spanish	4.77 (3.19; 1.4-9.38)	2.5 (2.48; 0-5.49)	10.83 (10.24; 1.59-23.93)	0.26	0.40
Tseltal	3.54 (3.94; 0.83-6.55)	4.48 (4.74; 0-8.68)	11.08 8.35; 2.78-33.08)	0.19	0.42
Yélî Dnye	3.13 (2.95; 1.58-6.26)	12.77 (13.57; 4.61-20.9)	19.87 (17.1; 7.25-38.54)	0.09	0.44

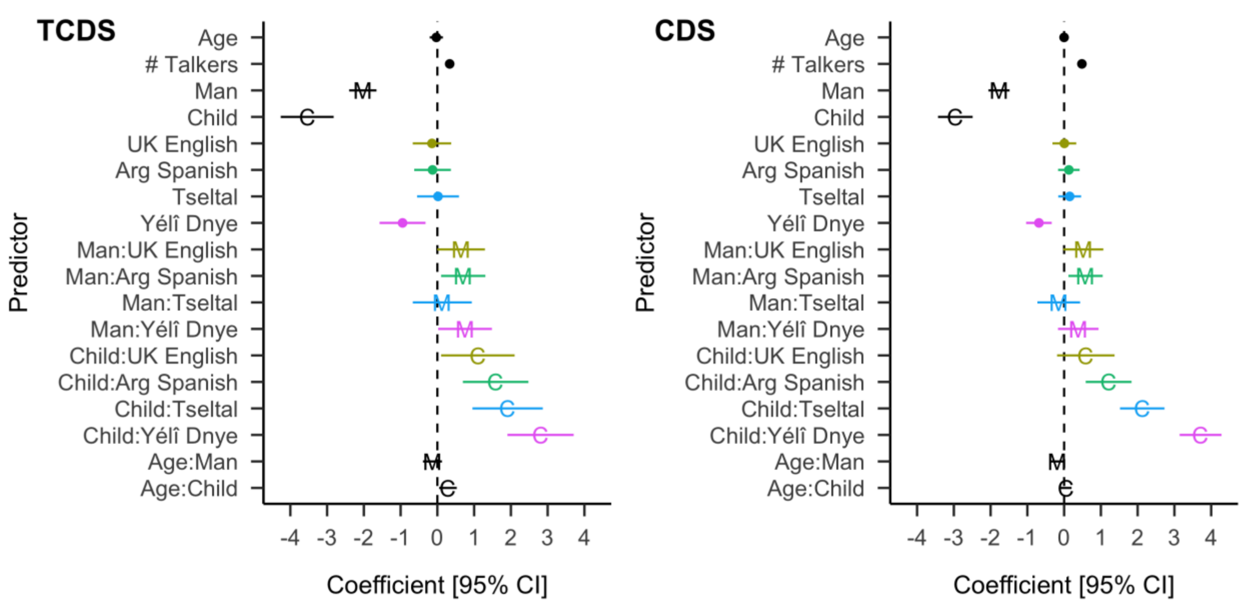

Figure 1. Coefficients and 95% confidence intervals for the count models of TCDS (left) and CDS (i.e. TCDS + OCDS; right) for all included fixed effects. This figure differs from the similar one in the main text, which by contrast features TCDS and ADS. Color indicates population (North American English is the modeled reference level), 'C' and 'M' indicate effects related to

child- and man-produced speech, respectively (woman-produced speech is set as the model reference level).

All child-directed speech (CDS)

The analysis of directed linguistic input in the main text focuses on TCDS; that is, input that is exclusively directed to the target child. And yet other types of child-directed input in the environment may also contain the linguistic and communicative features that are associated with language learning. We here analyze *all* CDS in the recordings using the same factors as we did for TCDS in the main text. "CDS" here includes all utterances directed to the target child, plus all other observable child-directed speech in the audio recording clips, including input directed at groups of children that may or may not include the target child. In other words "CDS" is here all hearable utterances that are directed to a child in the recording; comparable to what is measured in Bergelson et al. (2019). Therefore this measure of CDS includes all input designed for a child listener within earshot of the target child wearing the recorder. Keep in mind, however, that much of this input is likely to have been addressed to children of a different age than the target child, to children at a far distance to the target child, or even occasionally to children in a different language than what is typically used for the target child. We gloss over these issues here, as we do not have the annotations to tease each of them apart. Our aim instead is to provide a parallel statistical analysis of CDS to that of TCDS reported in the main text.

On average, across all language groups children were exposed to 7.43 minutes of CDS per hour across audio clips (median = 6.18), with wide variation between children (range = 0.37–24.11). Our model of CDS rate was nearly identical to that used for TCDS rate in the main text: It included target child age, talker type, the number of talkers present in the clip, and language group, with two additional two-way interactions (talker type by language group and child age by talker type) and random intercepts by child, adding only child age in the zero-inflation model component. The only difference from the main-text model of TCDS was that we did not include language group as a predictor in the zero-inflation component because its inclusion caused model non-convergence issues. As a reminder, there was no significant effect of language group in the zero-inflation model component of the main-text TCDS model. This fact, together with the qualitatively similar pattern of findings for CDS in the present model suggests that the pattern of findings reported are robust to this small difference in model structure (N = 2745, log-likelihood = -3,724.17, overdispersion estimate = 9.72, formula = CDS.min.p.hr ~ child.age + talker.type + num.tlkrs.in.clip + lg.grp + talker.type:lg.grp + child.age:talker.type + (1 | child.id), ziformula = ~ child.age). The results are qualitatively highly similar to the TCDS model presented in the main text. The coefficients and 95% confidence intervals for all fixed effects in the CDS count model are shown in Figure 1, side by side with the same plot from the TCDS model, which is replicated from the main text.

CDS input rate significantly differed by talker type, number of talkers present, language group, and the interaction between talker type and language group. As with TCDS rate, CDS rate was significantly lower for men compared to women (B = -1.77, SE = 0.15, z = -12.06, $p < 0.001$) and for children compared to women (B = -2.96, SE = 0.24, z = -12.31, $p < 0.001$). CDS rate was, like TCDS rate, also significantly higher when there were more talkers present (B = 0.48, SE =

0.03, $z = 15.82$, $p < 0.001$). As with TCDS, rates of CDS in Yélî Dnye were significantly lower compared to North American English ($B = -0.69$, $SE = 0.18$, $z = -3.86$, $p < 0.001$), with no significant differences between North American English and the other language groups (all p's ≥ 0.3).

Interactions between talker type and language group were overall similar, with some small differences. Men were previously found to produce significantly more TCDS in the Argentinian Spanish and Yélî Dnye samples compared to North American English. When this measure is changed to CDS, the difference only remains apparent for Argentinian Spanish compared to North American English ($B = 0.58$, $SE = 0.24$, $z = 2.45$, $p = 0.01$), though the Yélî Dnye data still point in the same direction ($B = 0.38$, $SE = 0.28$, $z = 1.37$, $p = 0.17$). Children were previously found to produce significantly more TCDS in all four of the non-North American English samples compared to North American English. When this measure is changed to CDS, the difference remains apparent for all cases except UK English, which still goes in the same direction (UK English: $B = 0.59$, $SE = 0.40$, $z = 1.48$, $p = 0.14$; Argentinian Spanish: $B = 1.21$, $SE = 0.32$, $z = 3.81$, $p < 0.001$; Tseltal: $B = 2.13$, $SE = 0.31$, $z = 6.89$, $p < 0.001$; Yélî Dnye: $B = 3.71$, $SE = 0.29$, $z = 12.78$, $p < 0.001$).

Interactions between talker type and age differed between TCDS and CDS. Whereas the previous analysis suggested that child-produced, but not man-produced, TCDS grows more with age compared to woman-produced TCDS, there are no significant differences across age by talker type with the input measure of CDS (Men $p = 0.07$; Children $p = 0.5$).

Like the model of TCDS rate, the zero-inflation regression component for CDS did not suggest any additional evidence for effects of child age ($p = 0.73$).

Typical numbers of talkers by corpus

Our primary statistical models account for the number of talkers present in a clip, with the idea that the presence of more talkers leads to more talk in the clip. This is trivially true in the sense that a talker in the clip isn't "counted" unless they talk at least once. The model effects of TCDS, CDS, and ADS, all show very strong effects of number of talkers in the clip, and suggest that the presence of others has a more-than-minimal effect on how much input children encounter, particularly for ADS. Further, we anticipated that some of the differences apparent between language groups are actually due to the greater or lesser number of people typically present around children. For example, we suspected that the organization of households and the number of children per household would lead to greater presence of both adults and children in the Yélî Dnye, Tseltal, and Argentinian Spanish recordings. Without controlling for the number of talkers present in our statistical models, it would have been impossible to tell what portion of cross-group differences in input rates is simply due to number of present talkers versus other differences in culture, language, and daily life. Our main-text results thus reflect estimates of cultural difference controlling for number of talkers as a separate and significant factor.

However, it is also the case that number of talkers may systematically differ between language groups in a way that approximates important cultural differences. We here analyze the number of talkers detected in clips across language groups and age, both overall and by talker type. Our aim is to illustrate the scale of cross-group differences in potential available interlocutors, which likely reflects differences in household size, household organization, and child caregiving practices. We only superficially characterize these differences here, leaving it to future work to more deeply engage with how these patterns reflect population-specific practices.

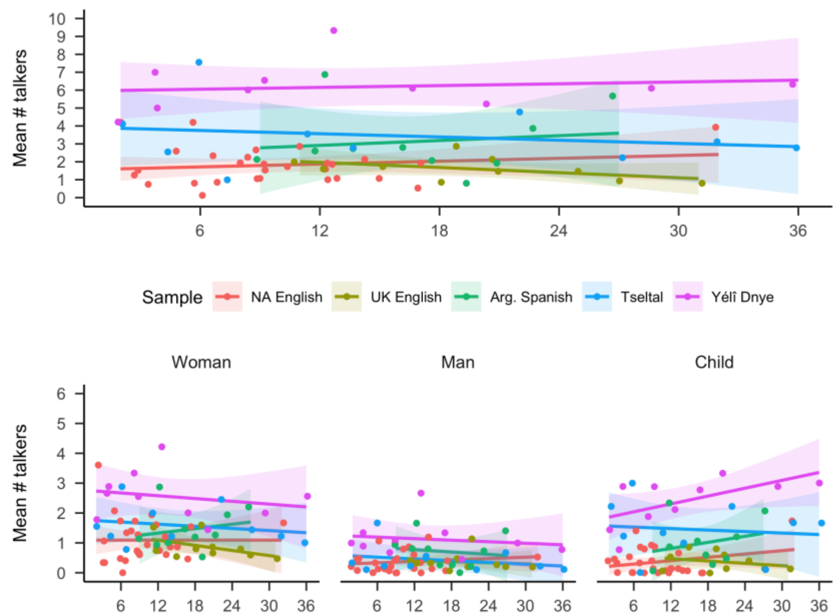

Figure 2. Number of talkers present across language groups, talker types, and target child age. Each datapoint represents the mean from one recording.

On average, and in addition to any vocalizations by the target child, a given audio clip included at least one utterance from 3.24 other talkers (median = 2.96; range over all clips = 0–19). By talker type, those other interlocutors included an average of 1.51 women, 0.59 men, and 1.11 children (medians are 1.37, 0.37, and 0.97, respectively). However, these averages obscure significant cross-group variation, which is apparent in Figure 2. In particular, the Yélî Dnye recordings show much higher rates of other talker presence compared to the other language groups, with averages of 2.54 women, 1.12 men, and 2.40 children. Compare to North American English with averages of 1.09 women, 0.36 men, and 0.36 children and UK English, with 0.86 women, 0.36 men, and 0.37 children. The Tseltal and Argentinian Spanish communities fall somewhere between the Yélî Dnye and English-speaking groups, with the Tseltal group showing averages of 1.58 women, 0.42 men, and 1.44 children and the Argentinian Spanish group averages of 1.45 women, 0.71 men, and 0.99 children.

Overall presence of other talkers looks similar across age, though we observe a slight downward trend in the number of women contributing input and a slight uptick in children contributing input in some groups. We do not statistically analyze these data given that the measure relies on the inferred number of talkers present rather than the actual number, which would require video data or time-sampled annotations (e.g., Cristia, Dupoux, Gurven, & Stieglitz, 2017). Thereby the current measure gives insight into effects of household and routine by language group, but not adequately to make well-substantiated claims at present.

(Non-/)effects of target child age by corpus

The distribution of child age across corpora (Figure 3) varied somewhat due to the fact that each corpus was collected at a different time and for different purposes by the contributing researchers, long before the present study was initiated. If we use simple linear regression to analyze the sampled 69 recordings to test whether age differs across corpora, we do find that children are significantly younger in the North American English language groups compared to all the other corpora except Yélî Dnye (age.in.months ~ lg.grp; $p < .05$ for UK English, Argentinian Spanish, and Tseltal and $p = 0.15$ for Yélî Dnye).

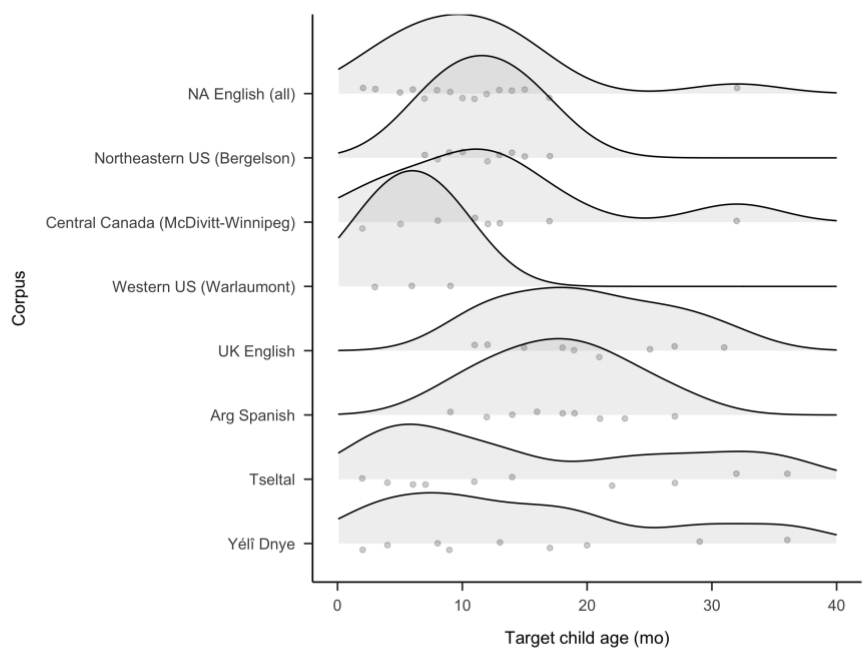

Figure 3. Target child age for each corpus. Each datapoint represents the mean from one recording.

In principle, age differences between language groups, even if present, are not necessarily a problem for the present study—in lay terms, each mixed-effects regression accounts for all modeled dimensions of each datapoint (including age, corpus, number of talkers in the clip, etc. in addition to a random effect of child) when estimating the direction and significance of impact that each predictor has on the dependent variable. That said, the relatively small sample size here combined with variance in age distribution between language groups could mean that there are true age effects present in the data that we cannot detect under present circumstances (i.e., supposing increased data would substantially change the linear fits of age in the model). As a reminder, we found no evidence for an overall effect of target child age (neither a decrease nor an increase) in the primary models of child-directed speech (i.e., TCDS in the main text and CDS here in the Supplementary Materials). We do find a significant decrease in ADS associated with child age. In our view, the primary concern is then whether we are missing overall effects of child age on TCDS and CDS. However, considering the much better age coverage of our four other language groups—which do not support overall age effects—and that the findings are in line with prior work on North American English showing no increase in CDS with age (Bergelson et al., 2019), we are satisfied with the current dataset and analysis.

We here visualize the effect of child age on the dependent variables of interest in each corpus (Figure 4) so that the interested reader can glean some informal and qualitative impression of potential differences that might be detected if more data were to be added in future work. Please note that any apparent visual differences here, as in the main text figures and tables, do not have the benefit of random-effects controls that are applied in our analysis via the use of statistical models.

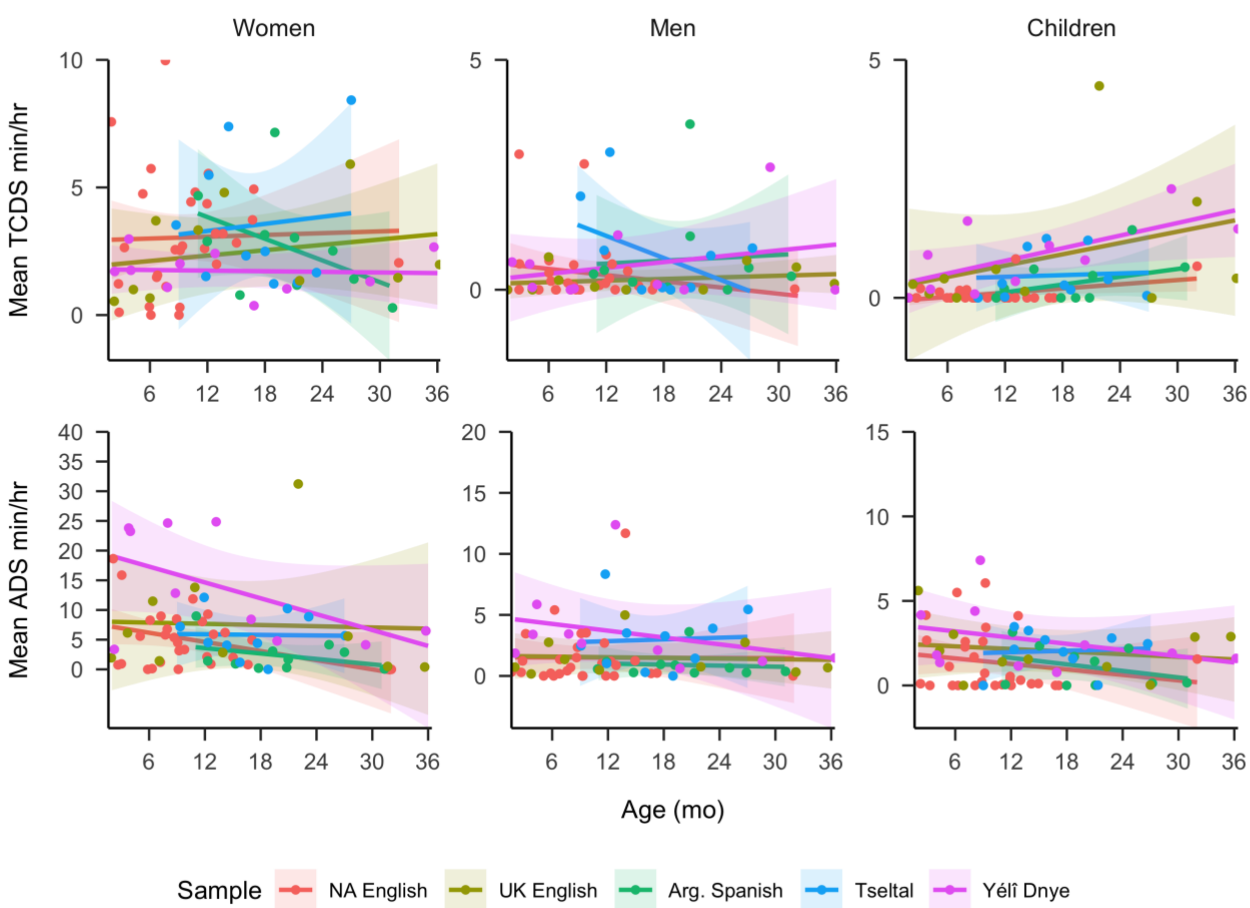

Figure 4. TCDS (above) and ADS (below) min/hr rates across language groups and talker types across target child age. Each datapoint represents the mean from one recording. This figure is similar to Figure 2 in the main text, but now additionall displays the data by child age. As is

apparent, age effects are minimal for TCDS across language groups and talker type whereas there is a general decrease in ADS across language group and talker type.

Simple models of age and cross-group difference in TCDS, CDS, and ADS rates

Up until now we have analyzed cross-corpus and age-based differences in TCDS, CDS, and ADS rate while *also* accounting for other factors that may drive variation in input rate. These factors include: the number of talkers known to be present in a given clip and the different talker types who produce this talk (male and female adults and non-target children). There are arguments for and against including these factors in our model of cross-cultural differences, depending on one's theoretical goals.

By including these factors in the model, as we have in the main-text models of TCDS and ADS and in the model of CDS above, we can gain a more detailed perspective on the shared features that drive variation in input rate between and within language groups. For example, by adding in the number of talkers in a clip to our model, we can account for the fact that the presence of more people generally leads to more talk, regardless of the child's developmental context—indeed, we find that this effect drives variation in general and thereby affects children regardless of whether they grow up in North America, the UK, Argentina, Chiapas, or Rossel Island. A similar case can be made for talker types: the fact that women are more likely to produce all three types of input than men or children illustrates a general finding that cross-cuts the language groups, though our main models show that this effect of talker type is slightly different from context to context. By modeling these effects, we can make fairly detailed predictions about the input a child is likely to hear in a given clip (e.g., we can predict how much ADS a Tseltal-acquiring child at age 12 months with 4 other voices present will hear, and how likely it is to come from a woman vsersus a child versus a man).

However, a valid alternative perspective is that these cross-corpus differences in the type and number of talkers are reflective of the children's broader cultural and linguistic milieu and therefore variance due to these factors should not be separately accounted for in the model if the end goal is to obtain a general picture of the differences in children's language experiences across these communities. Consider number of talkers present in the clip.

As shown in Supplementary Materials Section 2 above, there is systematic variation across our language groups in the number of present talkers: for example, Yélî Dnye-acquiring children are surrounded by substantially more talkers than children in the other groups. There may be two ways of looking at Yélî children's experience of TCDS: (1) All else being equal, Yélî Dnye-acquiring children hear significantly less TCDS compared to North American English, *but* the situation is not equal; because they have so many more people present than the North American English case, their overall TCDS input experienced is the same as (if not more than) what is heard by North American English-acquiring children or (2) considering children's overall linguistic environment, Yélî Dnye-acquiring children hear approximately the same rates of TCDS as North American English-acquiring children, if not more.

As the reader can tell, the first interpretation provides greater nuance, but more importantly, it speaks to useful avenues forward in understanding consistent and observable levers of cross-cultural difference (e.g., number of talkers present as a proxy for household size and composition or everyday routines; types of talker input as a proxy for alloparenting practices and (non-)overlap in work versus home settings). It can, however, obscure overall differences that are apparent when all these cultural effects add up in an individual child's experience.

In this analysis, therefore, we replicate our models of TCDS, CDS, and ADS, only now removing predictors relating to number of talkers present and type of talker. Therefore each count model only includes effects of child age (in months; centered and standardized) and language group (North American English/UK English/Argentinian Spanish/Tseltal/Yélî Dnye), and the zero-inflation component includes the same two predictors, with a random effect of child (formula = XDS.min.p.hr ~ child.age + lg.grp + (1 | child.id), ziformula = ~ child.age + lg.grp).

Target-child-directed speech (TCDS)

The count model of the simpler regression of TCDS (N = 915, log-likelihood = -2,031.92, overdispersion estimate = 6.67) showed no effects of child age or language group (all p's > .09). The zero-inflation component similarly showed no evidence for significant effects of age or language group (all p's ≥ .15).

All child-directed speech (CDS)

The count model of the simpler regression of CDS (N = 915, log-likelihood = -2,532.78, overdispersion estimate = 8.38) showed no effect of child age ($B = 0.08$, $SE = 0.07$, $z = 1.18$, $p = 0.24$) but a significant effect of language group: CDS rates were significantly higher for Yélî Dnye-acquiring children compared to North American English-acquiring children ($B = 0.80$, $SE = 0.18$, $z = 4.43$, $p < 0.001$). No other language group showed significant difference in the rate of CDS compared to North American English (all p's ≥ 0.2). However, in this case the zero-inflation component showed that both Tseltal and Yélî Dnye were significantly less likely than North American English to have clips with zero CDS (Tseltal: $B = -1.35$, $SE = 0.67$, $z = -2.02$, $p = 0.04$; Yélî Dnye: $B = -2.54$, $SE = 0.93$, $z = -2.71$, $p < 0.01$; UK English $p = .49$; Arg. Spanish $p = .99$). Put differently, the combined outcomes of the model components show that zero-CDS clips were significantly more likely for North American English-acquiring children than Tseltal and Yélî Dnye-acquiring children and that, for clips with some non-zero amount of CDS, the rate of CDS is significantly higher for Yélî Dnye-acquiring children than North American English-acquiring children. The zero-inflation component gave no evidence for an effect of child age ($p = 0.31$).

Adult-directed speech (ADS)

The count model of the simpler regression of ADS (N = 915, log-likelihood = -2,517.12, overdispersion estimate = 14.51) showed significant effects of both child age and language group. ADS decreased with child age ($B = -0.23$, $SE = 0.11$, $z = -2.16$, $p = 0.03$). ADS rates were also significantly higher in Yélî Dnye compared to North American English ($B = 0.75$, $SE = 0.29$, $z = 2.62$, $p < 0.01$). No other language group showed significant difference in the rate of ADS compared to North American English (all $p > 0.19$). As with CDS, the zero-inflation model component revealed further structure in the data: zero-ADS clips were significantly less likely in Yélî Dnye data compared to North American English (Yélî Dnye: $B = -2.46$, $SE = 0.79$, $z = -3.13$, $p < 0.01$; Tseltal: $B = -0.67$, $SE = 0.38$, $z = -1.79$, $p = 0.07$; UK English $p = .31$; Arg. Spanish $p = .20$). Again then, the combined output of the model shows that zero-ADS clips were significantly more likely for North American English-acquiring children than Yélî Dnye-acquiring children and that, for clips with non-zero amounts of ADS, the rate of ADS is significantly higher for Yélî Dnye-acquiring children than North American English-acquiring children. Consistent with the other models, there was no evidence of an age effect in the zero-inflation model component ($p = 0.85$).

Pulling these results together with those reported in the main text (TCDS, ADS) and above (CDS), two primary points are worth highlighting. First, Yélî Dnye looks very different when number of talkers is removed from the model—it has equivalent overall rates of TCDS, higher rates of CDS and ADS, and is less likely to have a zero-CDS or zero-ADS clip compared to North American English. This pattern falls in line with the main-text results and the fact that there are simply more people present in the language environment of Yélî Dnye-acquiring kids compared to the other language groups included here. Second, in this simplified analysis approach we lose sight of the critical and cross-group effects that account for fluctuations in talker presence and types of talkers present that we know, from the primary analyses, have a significant impact on the data.

Individual North American English data

In the main-text analyses we pool together the North American English datasets. We here present the primary descriptive figure from the main text, but with data broken out by individual corpus and not by language group.

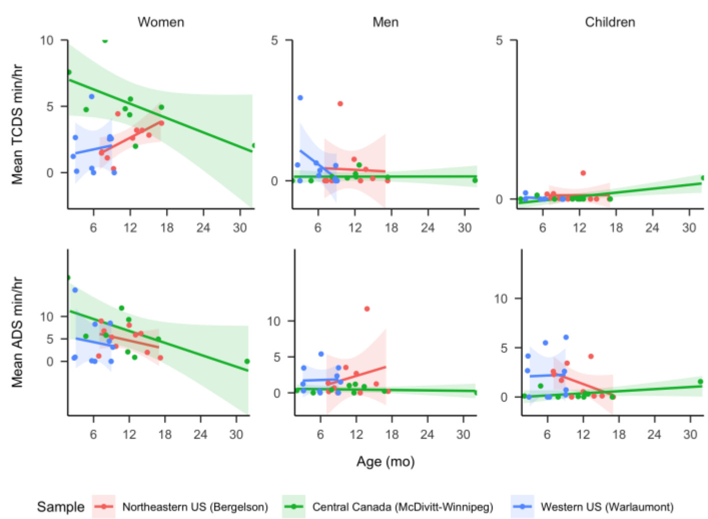

Figure 5. TCDS (above) and ADS (below) min/hr rates across individual North American English corpora and talker types across target child age. Each datapoint represents the mean from one recording.

Full model output for TCDS, CDS, and ADS

The full zero-inflated negative binomial mixed-effects regression output tables for TCDS, CDS, and ADS rate are presented below. Along with the output tables of TCDS and ADS we show the alternative models with other reference levels for language group.

Table 2: *Full output of the zero-inflated negative binomial mixed-effects regression of TCDS min/hr.*

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	(Intercept)	1.16	0.185	6.283	< 0.001	NA English
count	tchiyr.std	-0.03	0.092	-0.312	0.755	NA English
count	SpkrTypeMan	-2.03	0.19	-10.693	< 0.001	NA English
count	SpkrTypeChild	-3.54	0.367	-9.641	< 0.001	NA English
count	nsk.std	0.33	0.044	7.619	< 0.001	NA English
count	group_corpNEUK_English	-0.15	0.267	-0.551	0.582	NA English
count	group_corpNEArg_Spanish	-0.13	0.254	-0.517	0.605	NA English
count	group_corpNETseltal	0.02	0.291	0.06	0.952	NA English
count	group_corpNEYeli_Dnye	-0.95	0.319	-2.971	0.003	NA English
count	SpkrTypeMan:group_corpNEUK_English	0.64	0.331	1.949	0.051	NA English
count	SpkrTypeChild:group_corpNEUK_English	1.10	0.51	2.158	0.031	NA English
count	SpkrTypeMan:group_corpNEArg_Spanish	0.70	0.307	2.29	0.022	NA English
count	SpkrTypeChild:group_corpNEArg_Spanish	1.58	0.455	3.484	< 0.001	NA English
count	SpkrTypeMan:group_corpNETseltal	0.13	0.409	0.326	0.745	NA English
count	SpkrTypeChild:group_corpNETseltal	1.91	0.488	3.915	< 0.001	NA English
count	SpkrTypeMan:group_corpNEYeli_Dnye	0.75	0.373	2.017	0.044	NA English
count	SpkrTypeChild:group_corpNEYeli_Dnye	2.81	0.46	6.109	< 0.001	NA English
count	tchiyr.std:SpkrTypeMan	-0.13	0.132	-1.013	0.311	NA English
count	tchiyr.std:SpkrTypeChild	0.29	0.122	2.348	0.019	NA English
count	sd__(Intercept)	0.36				NA English
zero-inflation	(Intercept)	-2.67	1.938	-1.377	0.168	NA English
zero-inflation	tchiyr.std	-1.54	1.397	-1.103	0.27	NA English
zero-inflation	group_corpNEUK_English	-14.31	1941.172	-0.007	0.994	NA English
zero-inflation	group_corpNEArg_Spanish	-1.85	9.44	-0.196	0.845	NA English
zero-inflation	group_corpNETseltal	0.42	0.913	0.465	0.642	NA English
zero-inflation	group_corpNEYeli_Dnye	-1.45	3.607	-0.401	0.689	NA English

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	(Intercept)	1.04	0.191	5.473	< 0.001	UK English
count	tchiyr.std	-0.03	0.093	-0.275	0.783	UK English
count	SpkrTypeMan	-1.38	0.261	-5.312	< 0.001	UK English
count	SpkrTypeChild	-2.44	0.349	-6.97	< 0.001	UK English
count	nsk.std	0.34	0.043	7.791	< 0.001	UK English
count	group_corpUKNA_English	0.08	0.217	0.353	0.724	UK English
count	group_corpUKArg_Spanish	0.04	0.242	0.166	0.868	UK English
count	group_corpUKTseltal	0.03	0.265	0.096	0.923	UK English
count	group_corpUKYeli_Dnye	-0.74	0.288	-2.585	0.01	UK English
count	SpkrTypeMan:group_corpUKNA_English	-0.64	0.332	-1.941	0.052	UK English
count	SpkrTypeChild:group_corpUKNA_English	-1.11	0.51	-2.169	0.03	UK English
count	SpkrTypeMan:group_corpUKArg_Spanish	0.03	0.335	0.095	0.924	UK English
count	SpkrTypeChild:group_corpUKArg_Spanish	0.46	0.427	1.086	0.278	UK English
count	SpkrTypeMan:group_corpUKTseltal	-0.50	0.436	-1.145	0.252	UK English
count	SpkrTypeChild:group_corpUKTseltal	0.85	0.448	1.905	0.057	UK English
count	SpkrTypeMan:group_corpUKYeli_Dnye	0.12	0.414	0.296	0.767	UK English
count	SpkrTypeChild:group_corpUKYeli_Dnye	1.71	0.431	3.964	< 0.001	UK English
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	UK English
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	UK English
count	sd__(Intercept)	0.38				UK English
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.068	UK English
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	UK English
count	(Intercept)	1.08	0.193	5.608	< 0.001	Argentinian Spanish
count	tchiyr.std	-0.03	0.093	-0.275	0.783	Argentinian Spanish
count	SpkrTypeMan	-1.35	0.229	-5.907	< 0.001	Argentinian Spanish
count	SpkrTypeChild	-1.97	0.266	-7.403	< 0.001	Argentinian Spanish
count	nsk.std	0.34	0.043	7.791	< 0.001	Argentinian Spanish
count	group_corpASNA_English	0.04	0.204	0.178	0.859	Argentinian Spanish
count	group_corpASUK_English	-0.04	0.242	-0.166	0.868	Argentinian Spanish
count	group_corpASTseltal	-0.02	0.262	-0.056	0.956	Argentinian Spanish
count	group_corpASYeli_Dnye	-0.78	0.277	-2.827	0.005	Argentinian Spanish
count	SpkrTypeMan:group_corpASNA_English	-0.68	0.299	-2.26	0.024	Argentinian Spanish
count	SpkrTypeChild:group_corpASNA_English	-1.57	0.452	-3.468	< 0.001	Argentinian Spanish
count	SpkrTypeMan:group_corpASUK_English	-0.03	0.335	-0.095	0.925	Argentinian Spanish
count	SpkrTypeChild:group_corpASUK_English	-0.46	0.427	-1.086	0.278	Argentinian Spanish
count	SpkrTypeMan:group_corpASTseltal	-0.53	0.417	-1.272	0.203	Argentinian Spanish

SM: CROSS-CULTURAL EVERYDAY LANGUAGE EXPERIENCES

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	SpkrTypeChild:group_corpASTseltal	0.39	0.392	0.997	0.319	Argentinian Spanish
count	SpkrTypeMan:group_corpASYeli_Dnye	0.09	0.395	0.23	0.818	Argentinian Spanish
count	SpkrTypeChild:group_corpASYeli_Dnye	1.24	0.372	3.343	< 0.001	Argentinian Spanish
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	Argentinian Spanish
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	Argentinian Spanish
count	sd__(Intercept)	0.38				Argentinian Spanish
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.069	Argentinian Spanish
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	Argentinian Spanish
count	(Intercept)	1.07	0.222	4.814	< 0.001	Tseltal
count	tchiyr.std	-0.03	0.093	-0.275	0.783	Tseltal
count	SpkrTypeMan	-1.88	0.356	-5.286	< 0.001	Tseltal
count	SpkrTypeChild	-1.58	0.31	-5.108	< 0.001	Tseltal
count	nsk.std	0.34	0.043	7.791	< 0.001	Tseltal
count	group_corpTSNA_English	0.05	0.235	0.217	0.828	Tseltal
count	group_corpTSUK_English	-0.03	0.265	-0.096	0.923	Tseltal
count	group_corpTSArg_Spanish	0.02	0.262	0.056	0.956	Tseltal
count	group_corpTSYeli_Dnye	-0.77	0.29	-2.652	0.008	Tseltal
count	SpkrTypeMan:group_corpTSNA_English	-0.14	0.404	-0.358	0.72	Tseltal
count	SpkrTypeChild:group_corpTSNA_English	-1.96	0.483	-4.059	< 0.001	Tseltal
count	SpkrTypeMan:group_corpTSUK_English	0.50	0.436	1.145	0.252	Tseltal
count	SpkrTypeChild:group_corpTSUK_English	-0.85	0.448	-1.905	0.057	Tseltal
count	SpkrTypeMan:group_corpTSArg_Spanish	0.53	0.417	1.272	0.203	Tseltal
count	SpkrTypeChild:group_corpTSArg_Spanish	-0.39	0.392	-0.997	0.319	Tseltal
count	SpkrTypeMan:group_corpTSYeli_Dnye	0.62	0.481	1.291	0.197	Tseltal
count	SpkrTypeChild:group_corpTSYeli_Dnye	0.85	0.397	2.146	0.032	Tseltal
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	Tseltal
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	Tseltal
count	sd__(Intercept)	0.38				Tseltal
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.068	Tseltal
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	Tseltal
count	(Intercept)	0.30	0.237	1.263	0.207	Yélî Dnye
count	tchiyr.std	-0.03	0.093	-0.275	0.783	Yélî Dnye
count	SpkrTypeMan	-1.26	0.326	-3.872	< 0.001	Yélî Dnye
count	SpkrTypeChild	-0.73	0.275	-2.649	0.008	Yélî Dnye
count	nsk.std	0.34	0.043	7.791	< 0.001	Yélî Dnye
count	group_corpYDNA_English	0.82	0.253	3.239	0.001	Yélî Dnye

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	group_corpYDUK_English	0.74	0.288	2.585	0.01	Yélî Dnye
count	group_corpYDArg_Spanish	0.78	0.277	2.827	0.005	Yélî Dnye
count	group_corpYDTseltal	0.77	0.29	2.652	0.008	Yélî Dnye
count	SpkrTypeMan:group_corpYDNA_English	-0.77	0.377	-2.032	0.042	Yélî Dnye
count	SpkrTypeChild:group_corpYDNA_English	-2.81	0.463	-6.077	< 0.001	Yélî Dnye
count	SpkrTypeMan:group_corpYDUK_English	-0.12	0.414	-0.296	0.767	Yélî Dnye
count	SpkrTypeChild:group_corpYDUK_English	-1.71	0.431	-3.964	< 0.001	Yélî Dnye
count	SpkrTypeMan:group_corpYDArg_Spanish	-0.09	0.395	-0.23	0.818	Yélî Dnye
count	SpkrTypeChild:group_corpYDArg_Spanish	-1.24	0.372	-3.343	< 0.001	Yélî Dnye
count	SpkrTypeMan:group_corpYDTseltal	-0.62	0.481	-1.291	0.197	Yélî Dnye
count	SpkrTypeChild:group_corpYDTseltal	-0.85	0.397	-2.146	0.032	Yélî Dnye
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	Yélî Dnye
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	Yélî Dnye
count	sd__(Intercept)	0.38				Yélî Dnye
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.069	Yélî Dnye
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	Yélî Dnye

Table 3: *Full output of the zero-inflated negative binomial mixed-effects regression of CDS min/hr.*

Model component	Term	Estimate (B)	Std. Error	t-value	p-value
count	(Intercept)	1.37	0.086	16.008	< 0.001
count	tchiyr.std	0.00	0.057	-0.015	0.988
count	SpkrTypeMan	-1.77	0.147	-12.063	< 0.001
count	SpkrTypeChild	-2.96	0.24	-12.309	< 0.001
count	nsk.std	0.48	0.031	15.821	< 0.001
count	group_corpNEUK_English	0.01	0.165	0.039	0.969
count	group_corpNEArg_Spanish	0.13	0.15	0.862	0.389
count	group_corpNETseltal	0.15	0.159	0.954	0.34
count	group_corpNEYeli_Dnye	-0.69	0.178	-3.855	< 0.001
count	SpkrTypeMan:group_corpNEUK_English	0.52	0.278	1.872	0.061
count	SpkrTypeChild:group_corpNEUK_English	0.59	0.399	1.48	0.139

Model component	Term	Estimate (B)	Std. Error	t-value	p-value
count	SpkrTypeMan:group_corp NEArg_Spanish	0.58	0.238	2.454	0.014
count	SpkrTypeChild:group_corp NEArg_Spanish	1.21	0.318	3.809	< 0.001
count	SpkrTypeMan:group_corp NETseltal	-0.15	0.296	-0.494	0.621
count	SpkrTypeChild:group_corp NETseltal	2.12	0.309	6.885	< 0.001
count	SpkrTypeMan:group_corp NEYeli_Dnye	0.38	0.281	1.368	0.171
count	SpkrTypeChild:group_corp NEYeli_Dnye	3.71	0.29	12.782	< 0.001
count	tchiyr.std:SpkrTypeMan	-0.18	0.102	-1.792	0.073
count	tchiyr.std:SpkrTypeChild	0.05	0.077	0.677	0.498
count	sd__(Intercept)	0.23			
zero-inflation	(Intercept)	-3.91	0.846	-4.629	< 0.001
zero-inflation	tchiyr.std	0.17	0.505	0.342	0.733

Table 4: *Full output of the zero-inflated negative binomial mixed-effects regression of ADS min/hr.*

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	(Intercept)	1.16	0.185	6.283	< 0.001	NA English
count	tchiyr.std	-0.03	0.092	-0.312	0.755	NA English
count	SpkrTypeMan	-2.03	0.19	-10.693	< 0.001	NA English
count	SpkrTypeChild	-3.54	0.367	-9.641	< 0.001	NA English
count	nsk.std	0.33	0.044	7.619	< 0.001	NA English
count	group_corpNEUK_English	-0.15	0.267	-0.551	0.582	NA English
count	group_corpNEArg_Spanish	-0.13	0.254	-0.517	0.605	NA English
count	group_corpNETseltal	0.02	0.291	0.06	0.952	NA English
count	group_corpNEYeli_Dnye	-0.95	0.319	-2.971	0.003	NA English
count	SpkrTypeMan:group_corp NEUK_English	0.64	0.331	1.949	0.051	NA English
count	SpkrTypeChild:group_corpNEUK_English	1.10	0.51	2.158	0.031	NA English
count	SpkrTypeMan:group_corp NEArg_Spanish	0.70	0.307	2.29	0.022	NA English
count	SpkrTypeChild:group_corpNEArg_Spanish	1.58	0.455	3.484	< 0.001	NA English
count	SpkrTypeMan:group_corp NETseltal	0.13	0.409	0.326	0.745	NA English
count	SpkrTypeChild:group_corpNETseltal	1.91	0.488	3.915	< 0.001	NA English
count	SpkrTypeMan:group_corp NEYeli_Dnye	0.75	0.373	2.017	0.044	NA English

SM: CROSS-CULTURAL EVERYDAY LANGUAGE EXPERIENCES

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	SpkrTypeChild:group_corpNEYeli_Dnye	2.81	0.46	6.109	< 0.001	NA English
count	tchiyr.std:SpkrTypeMan	-0.13	0.132	-1.013	0.311	NA English
count	tchiyr.std:SpkrTypeChild	0.29	0.122	2.348	0.019	NA English
count	sd__(Intercept)	0.36				NA English
zero-inflation	(Intercept)	-2.67	1.938	-1.377	0.168	NA English
zero-inflation	tchiyr.std	-1.54	1.397	-1.103	0.27	NA English
zero-inflation	group_corpNEUK_English	-14.31	1941.172	-0.007	0.994	NA English
zero-inflation	group_corpNEArg_Spanish	-1.85	9.44	-0.196	0.845	NA English
zero-inflation	group_corpNETseltal	0.42	0.913	0.465	0.642	NA English
zero-inflation	group_corpNEYeli_Dnye	-1.45	3.607	-0.401	0.689	NA English
count	(Intercept)	1.04	0.191	5.473	< 0.001	UK English
count	tchiyr.std	-0.03	0.093	-0.275	0.783	UK English
count	SpkrTypeMan	-1.38	0.261	-5.312	< 0.001	UK English
count	SpkrTypeChild	-2.44	0.349	-6.97	< 0.001	UK English
count	nsk.std	0.34	0.043	7.791	< 0.001	UK English
count	group_corpUKNA_English	0.08	0.217	0.353	0.724	UK English
count	group_corpUKArg_Spanish	0.04	0.242	0.166	0.868	UK English
count	group_corpUKTseltal	0.03	0.265	0.096	0.923	UK English
count	group_corpUKYeli_Dnye	-0.74	0.288	-2.585	0.01	UK English
count	SpkrTypeMan:group_corpUKNA_English	-0.64	0.332	-1.941	0.052	UK English
count	SpkrTypeChild:group_corpUKNA_English	-1.11	0.51	-2.169	0.03	UK English
count	SpkrTypeMan:group_corpUKArg_Spanish	0.03	0.335	0.095	0.924	UK English
count	SpkrTypeChild:group_corpUKArg_Spanish	0.46	0.427	1.086	0.278	UK English
count	SpkrTypeMan:group_corpUKTseltal	-0.50	0.436	-1.145	0.252	UK English
count	SpkrTypeChild:group_corpUKTseltal	0.85	0.448	1.905	0.057	UK English
count	SpkrTypeMan:group_corpUKYeli_Dnye	0.12	0.414	0.296	0.767	UK English
count	SpkrTypeChild:group_corpUKYeli_Dnye	1.71	0.431	3.964	< 0.001	UK English
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	UK English
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	UK English
count	sd__(Intercept)	0.38				UK English
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.068	UK English
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	UK English
count	(Intercept)	1.08	0.193	5.608	< 0.001	Argentinian Spanish
count	tchiyr.std	-0.03	0.093	-0.275	0.783	Argentinian Spanish
count	SpkrTypeMan	-1.35	0.229	-5.907	< 0.001	Argentinian Spanish

SM: CROSS-CULTURAL EVERYDAY LANGUAGE EXPERIENCES

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	SpkrTypeChild	-1.97	0.266	-7.403	< 0.001	Argentinian Spanish
count	nsk.std	0.34	0.043	7.791	< 0.001	Argentinian Spanish
count	group_corpASNA_English	0.04	0.204	0.178	0.859	Argentinian Spanish
count	group_corpASUK_English	-0.04	0.242	-0.166	0.868	Argentinian Spanish
count	group_corpASTseltal	-0.02	0.262	-0.056	0.956	Argentinian Spanish
count	group_corpASYeli_Dnye	-0.78	0.277	-2.827	0.005	Argentinian Spanish
count	SpkrTypeMan:group_corpASNA_English	-0.68	0.299	-2.26	0.024	Argentinian Spanish
count	SpkrTypeChild:group_corpASNA_English	-1.57	0.452	-3.468	< 0.001	Argentinian Spanish
count	SpkrTypeMan:group_corpASUK_English	-0.03	0.335	-0.095	0.925	Argentinian Spanish
count	SpkrTypeChild:group_corpASUK_English	-0.46	0.427	-1.086	0.278	Argentinian Spanish
count	SpkrTypeMan:group_corpASTseltal	-0.53	0.417	-1.272	0.203	Argentinian Spanish
count	SpkrTypeChild:group_corpASTseltal	0.39	0.392	0.997	0.319	Argentinian Spanish
count	SpkrTypeMan:group_corpASYeli_Dnye	0.09	0.395	0.23	0.818	Argentinian Spanish
count	SpkrTypeChild:group_corpASYeli_Dnye	1.24	0.372	3.343	< 0.001	Argentinian Spanish
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	Argentinian Spanish
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	Argentinian Spanish
count	sd__(Intercept)	0.38				Argentinian Spanish
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.069	Argentinian Spanish
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	Argentinian Spanish
count	(Intercept)	1.07	0.222	4.814	< 0.001	Tseltal
count	tchiyr.std	-0.03	0.093	-0.275	0.783	Tseltal
count	SpkrTypeMan	-1.88	0.356	-5.286	< 0.001	Tseltal
count	SpkrTypeChild	-1.58	0.31	-5.108	< 0.001	Tseltal
count	nsk.std	0.34	0.043	7.791	< 0.001	Tseltal
count	group_corpTSNA_English	0.05	0.235	0.217	0.828	Tseltal
count	group_corpTSUK_English	-0.03	0.265	-0.096	0.923	Tseltal
count	group_corpTSArg_Spanish	0.02	0.262	0.056	0.956	Tseltal
count	group_corpTSYeli_Dnye	-0.77	0.29	-2.652	0.008	Tseltal
count	SpkrTypeMan:group_corpTSNA_English	-0.14	0.404	-0.358	0.72	Tseltal
count	SpkrTypeChild:group_corpTSNA_English	-1.96	0.483	-4.059	< 0.001	Tseltal

SM: CROSS-CULTURAL EVERYDAY LANGUAGE EXPERIENCES

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	SpkrTypeMan:group_corpTSUK_English	0.50	0.436	1.145	0.252	Tseltal
count	SpkrTypeChild:group_corpTSUK_English	-0.85	0.448	-1.905	0.057	Tseltal
count	SpkrTypeMan:group_corpTSArg_Spanish	0.53	0.417	1.272	0.203	Tseltal
count	SpkrTypeChild:group_corpTSArg_Spanish	-0.39	0.392	-0.997	0.319	Tseltal
count	SpkrTypeMan:group_corpTSYeli_Dnye	0.62	0.481	1.291	0.197	Tseltal
count	SpkrTypeChild:group_corpTSYeli_Dnye	0.85	0.397	2.146	0.032	Tseltal
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	Tseltal
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	Tseltal
count	sd__(Intercept)	0.38				Tseltal
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.068	Tseltal
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	Tseltal
count	(Intercept)	0.30	0.237	1.263	0.207	Yélî Dnye
count	tchiyr.std	-0.03	0.093	-0.275	0.783	Yélî Dnye
count	SpkrTypeMan	-1.26	0.326	-3.872	< 0.001	Yélî Dnye
count	SpkrTypeChild	-0.73	0.275	-2.649	0.008	Yélî Dnye
count	nsk.std	0.34	0.043	7.791	< 0.001	Yélî Dnye
count	group_corpYDNA_English	0.82	0.253	3.239	0.001	Yélî Dnye
count	group_corpYDUK_English	0.74	0.288	2.585	0.01	Yélî Dnye
count	group_corpYDArg_Spanish	0.78	0.277	2.827	0.005	Yélî Dnye
count	group_corpYDTseltal	0.77	0.29	2.652	0.008	Yélî Dnye
count	SpkrTypeMan:group_corpYDNA_English	-0.77	0.377	-2.032	0.042	Yélî Dnye
count	SpkrTypeChild:group_corpYDNA_English	-2.81	0.463	-6.077	< 0.001	Yélî Dnye
count	SpkrTypeMan:group_corpYDUK_English	-0.12	0.414	-0.296	0.767	Yélî Dnye
count	SpkrTypeChild:group_corpYDUK_English	-1.71	0.431	-3.964	< 0.001	Yélî Dnye
count	SpkrTypeMan:group_corpYDArg_Spanish	-0.09	0.395	-0.23	0.818	Yélî Dnye
count	SpkrTypeChild:group_corpYDArg_Spanish	-1.24	0.372	-3.343	< 0.001	Yélî Dnye
count	SpkrTypeMan:group_corpYDTseltal	-0.62	0.481	-1.291	0.197	Yélî Dnye
count	SpkrTypeChild:group_corpYDTseltal	-0.85	0.397	-2.146	0.032	Yélî Dnye
count	tchiyr.std:SpkrTypeMan	-0.14	0.131	-1.039	0.299	Yélî Dnye
count	tchiyr.std:SpkrTypeChild	0.27	0.123	2.219	0.026	Yélî Dnye
count	sd__(Intercept)	0.38				Yélî Dnye
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.069	Yélî Dnye
zero-inflation	tchiyr.std	-1.62	1.281	-1.261	0.207	Yélî Dnye

Marginal means estimates of TCDS and ADS

In the main manuscript the plotted data of TCDS and ADS across language groups are shown based on the actual data. However, our analyses take into account a number of factors that are not visible in that main-text plot. Here we show marginal mean plots from the models of ADS and TCDS, displaying estimated values of each over child age, and talker type, for each language group.

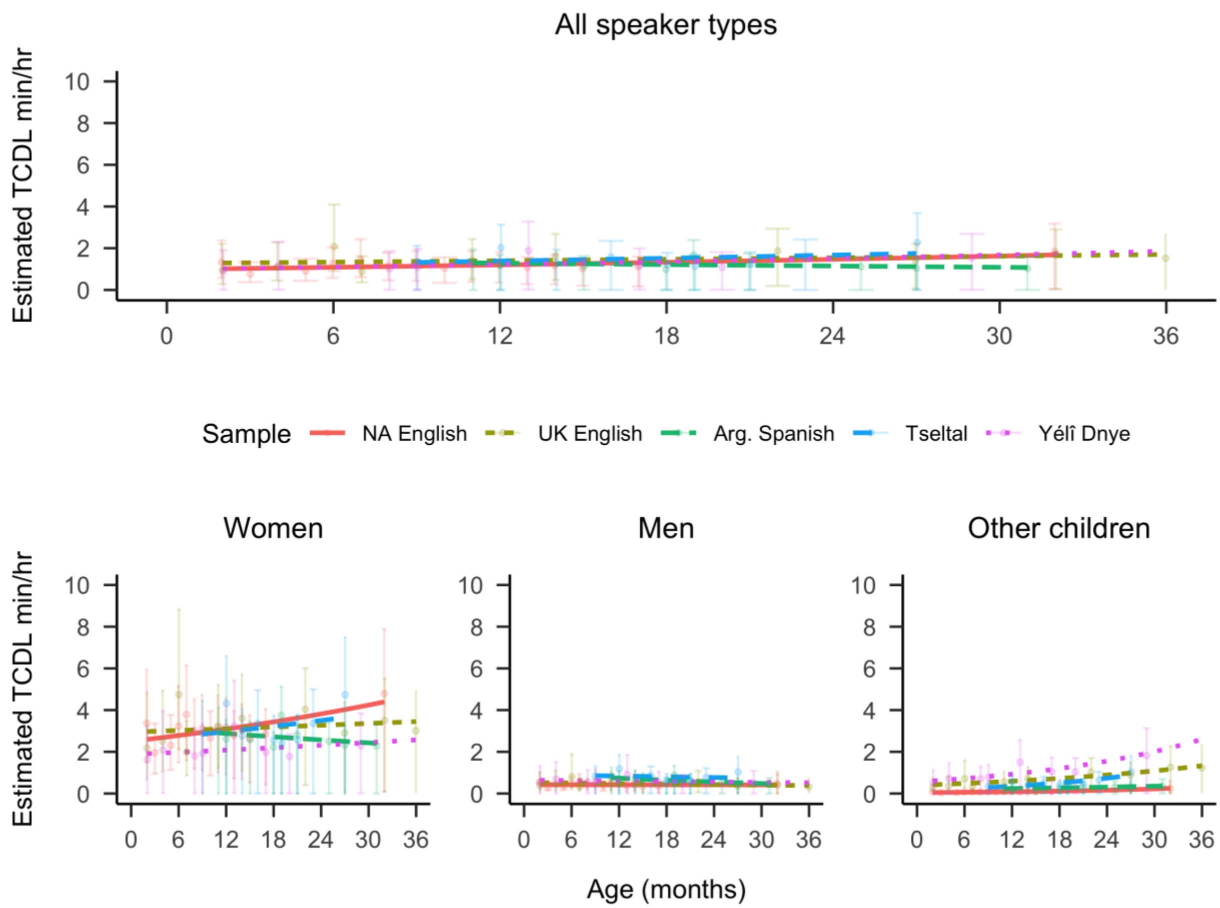

Figure 6. Model estimated rates of TCDS min/hr by target child age and corpus, shown all together in the top panel and split by talker type in the bottom panel. Each datapoint represents the mean from one recording.

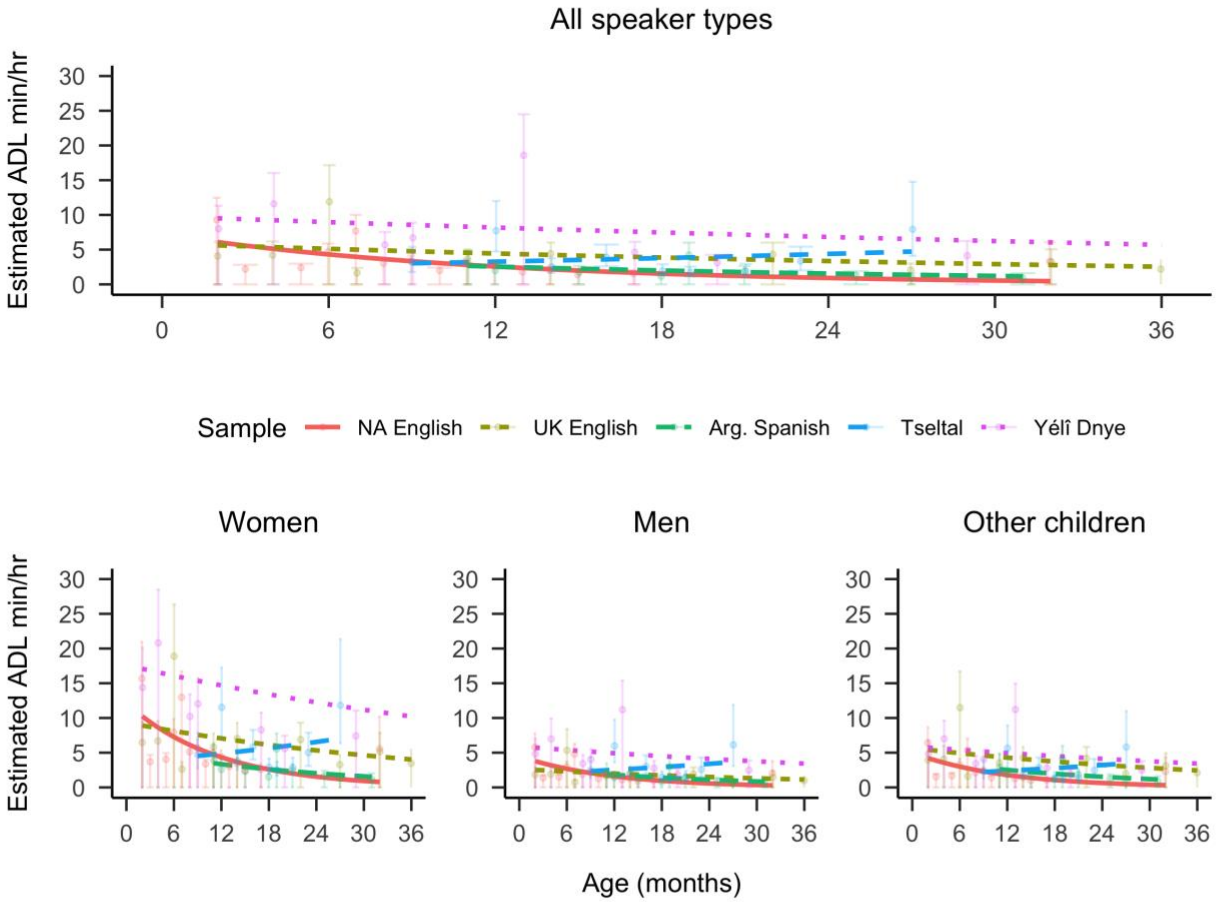

Figure 7. Model estimated rates of ADS min/hr by target child age and corpus, shown all together in the top panel and split by talker type in the bottom panel. Each datapoint represents the mean from one recording.

Confusion among addressee annotation types

The agreement scores for addressee ranged from slight to substantial across corpora. We here give further information about how these agreement scores are derived and distributed in our dataset. We show below a confusion matrix displaying the first versus second annotator's impressions of addressee across all corpora. To deal with the fact that we have continuous data (not, e.g., ratings at the clip level, as in Bergelson et al. (2019) or Weisleder and Fernald (2013)) we compute reliability over very short audio frames in the transcriptions. If the annotation from Human 1 matches the annotation from Human 2 during a given frame, then it is a match. Otherwise, it is a mismatch. So if, for example, Human 1 annotates a 1000 msec stretch as addressee "T" and Human 2 annotates an overlapping 500 msec stretch as addressee "T", they only agree for half of the frames. As the reader can see, this means that if one annotator misses an utterance, or if there is a lot of overlap in talkers (for which we cannot straightforwardly compute an addressee value for a frame), agreement declines.

The actual addressee categories in our reliability analysis are "T" (target child), "C" (other child), "A" (adult), or "O" (other, including all other types and unsure). The "Other" category shown in the confusion matrix below includes cases where one annotator's frame included silence or overlapping talk. Each row sums to 100% (note that cell estimates are rounded). As is apparent in the figure, for the cases when the two annotators agree that there is a single talker producing talk, confusion is relatively low and the correct category assignment is reliably visible against the other choices (i.e., note the apparent darker diagonal across the matrix). Confusion between "T" and "C" is slightly higher than the other cases of addressee type, but not by much. Most cases of disagreement concern cases of silence or overlapping speech, an error type that ultimately comes from the process of speech segmentation and is then inherited in an assessment of addressee agreement.

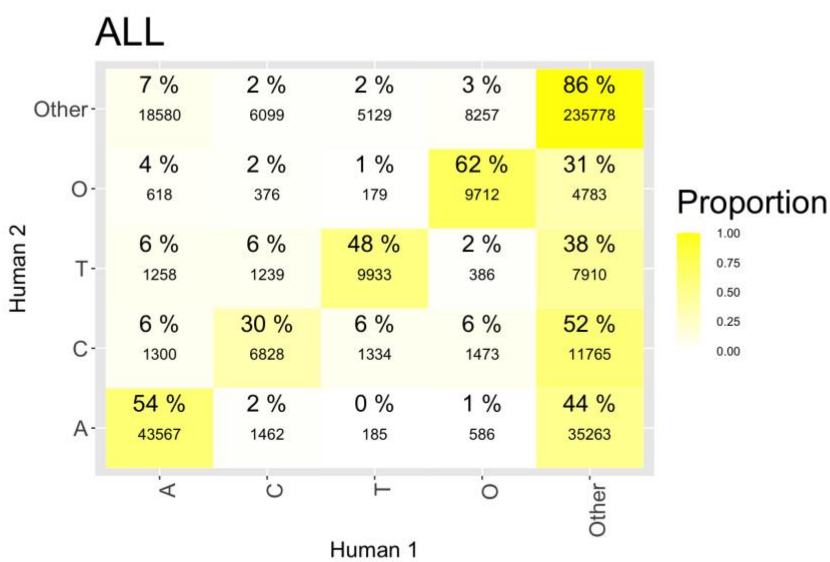

Figure 8. Confusion matrix across all corpora for addressee annotations between the two human annotators.

However, this is the sum over *all* of the data that we have. We had anticipated that reliability would be substantially worse in the Argentinian Spanish, Tseltal, and Yélî Dnye data because the reliability annotators ("Human 2") did not fluently speak the language variety they were working on for reliability—if they had any knowledge at all of that language. Only the English-based corpora always had native-speaking annotators. Indeed, the reliability patterns are less clear for the non-English corpora, likely due to this language barrier, which prevents the annotator from taking all context into account in their decisions. This pattern is visible in the same matrix, but now created for each individual corpus. We take this set of findings as indirect evidence in favor of the only *partial* recoverability of register and addressee-specific features across unrelated languages (see also Soderstrom et al., 2021).

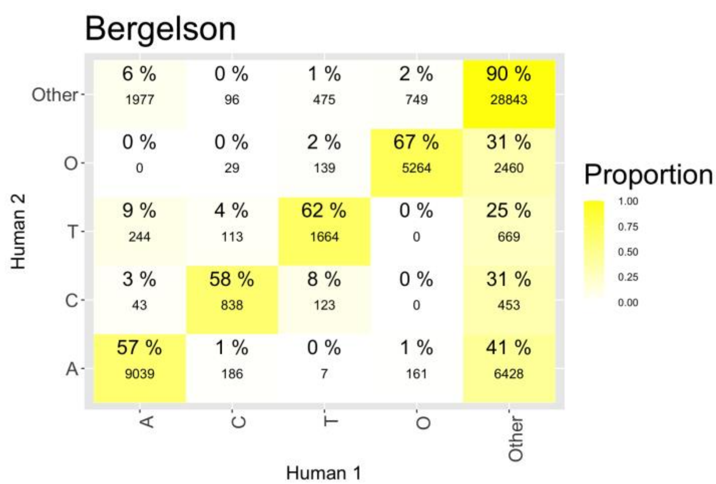

Figure 9. Confusion matrix for the Bergelson (US English) corpus for addressee annotations between the two human annotators.

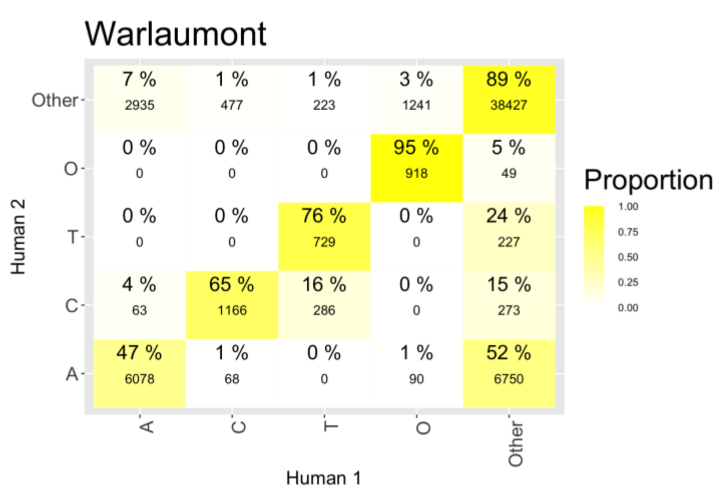

Figure 10. Confusion matrix for the Warlaumont (US English) corpus for addressee annotations between the two human annotators.

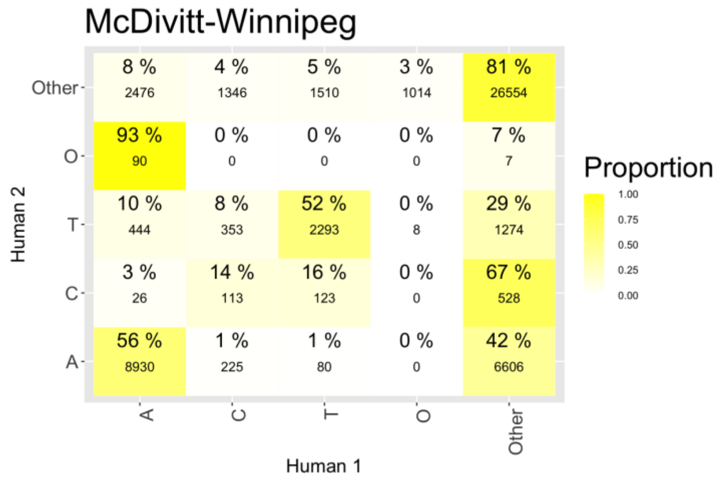

Figure 11. Confusion matrix for the McDivitt/Winnipeg (Canadian English) corpus for addressee annotations between the two human annotators.

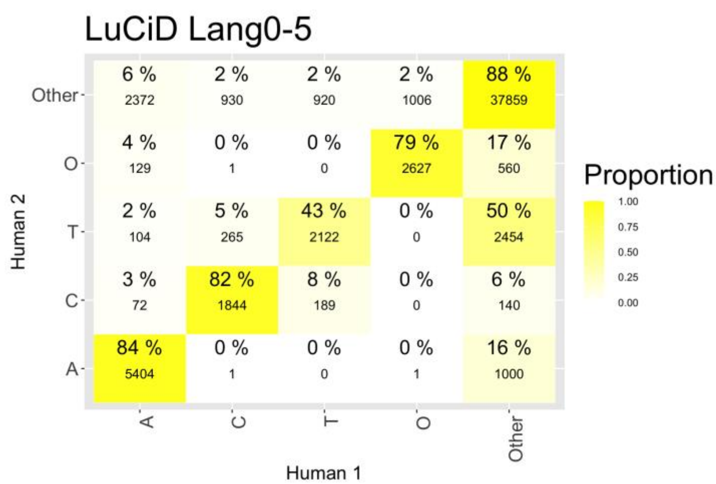

Figure 12. Confusion matrix for the LuCiD Lang0–5 (UK English) corpus for addressee annotations between the two human annotators.

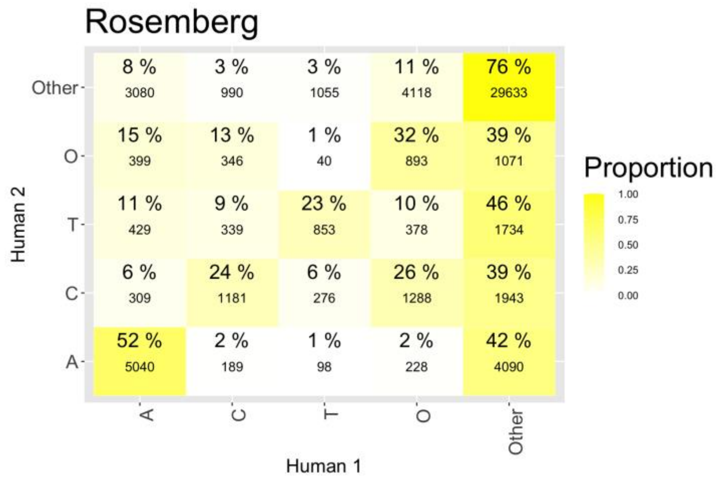

Figure 13. Confusion matrix for the Rosemberg (Argentinian Spanish) corpus for addressee annotations between the two human annotators. Reliability annotators had non-native familiarity with a different variety of Spanish.

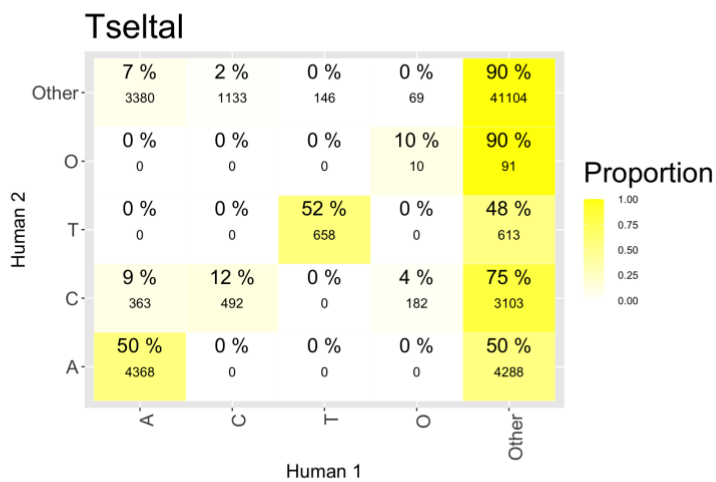

Figure 14. Confusion matrix for the Tseltal corpus for addressee annotations between the two human annotators. Reliability annotators were completely unfamiliar with the language but there are occasional words forms used in Spanish (e.g., borrowings) that may have been familiar to the annotators.

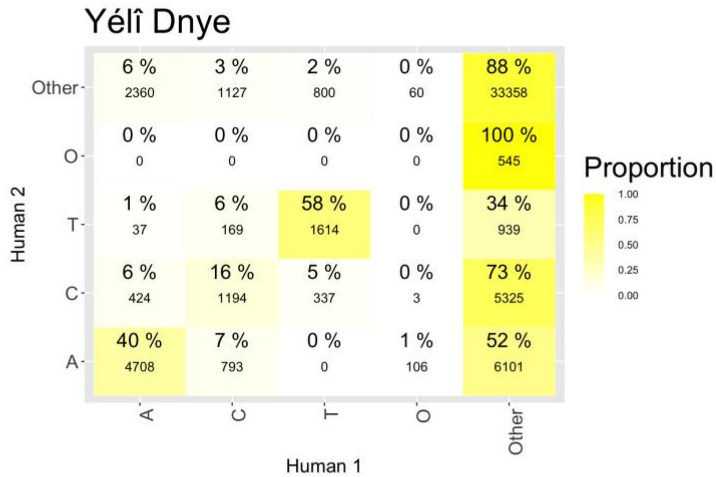

Figure 15. Confusion matrix for the Yélî Dnye corpus for addressee annotations between the two human annotators. Reliability annotators were completely unfamiliar with the language.

References

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

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

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, *90*(3), 759–773. https://doi.org/https://doi.org/10.1111/cdev.12974

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/

Soderstrom, M., Casillas, M., Gornik, M., Bouchard, A., MacEwan, S., Shokrkon, A., & Bunce, J. (2021). English-speaking adults' labeling of child-and adult-directed speech across languages and its relationship to perception of affect. *Frontiers in Psychology*, *12*, 708887. https://doi.org/https://doi.org/10.3389/fpsyg.2021.708887

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