What Do North American Babies Hear? A large-scale cross-corpus analysis

Elika Bergelson[*†1], Marisa Casillas[*‡2], Melanie Soderstrom[3], Amanda Seidl[4], Anne S. Warlaumont[5] and Andrei Amatuni[1]

[1]417 Chapel Dr., Campus Box 90086, Psychology & Neuroscience, Duke University, Durham, NC, 27708, USA

[2]Language Development Department, Max Planck Institute for Psycholinguistics, PO Box 310, 6500 AH, Nijmegen, The Netherlands

[3]Psychology, University of Manitoba, Canada

[4]Speech, Language, and Hearing Sciences, Purdue University, USA

[5]Communication, University of California, Los Angeles, USA

.

———

[*]Joint first authors

[†]elika.bergelson@duke.edu; phone: +001-919-660-8753

[‡]marisa.casillas@mpi.nl; phone: +31-(0)24-3521245

What Do North American Babies Hear? A large-scale cross-corpus analysis

Abstract

A range of demographic variables influence how much speech young children hear. However, because studies have used vastly different sampling methods, quantitative comparison of interlocking demographic effects has been nearly impossible, across or within studies. We harnessed a unique collection of existing naturalistic, day-long recordings from 61 homes across four North American cities to examine language input as a function of age, gender, and maternal education. We analyzed adult speech heard by 3- to 20-month-olds who wore audio recorders for an entire day. We annotated speaker gender and speech register (child-directed or adult-directed) for 10,861 utterances from female and male adults in these recordings. Examining age, gender, and maternal education collectively in this ecologically-valid dataset, we find several key results. First, the speaker gender imbalance in the input is striking: children heard 2–3x more speech from females than males. Second, children in higher-maternal-education homes heard *more* child-directed speech than those in lower-maternal education homes. Finally, our analyses revealed a previously unreported effect: the proportion of child-directed speech in the input increases with age, due to a *decrease* in adult-directed speech with age. This large-scale analysis is an important step forward in collectively examining demographic variables that influence early development, made possible by pooled, comparable, day-long recordings of children's language environments. The audio recordings, annotations, and annotation software are readily available for re-use and re-analysis by other researchers.

Highlights

- We measured adult- and child-directed speech heard by infants from four North American cities, extracted from at-home daylong audio recordings

- We simultaneously modeled the influences of age, gender, and maternal education on the amount of child- and adult-directed speech in children's input (>10,000

utterances)

- Infants heard 2–3x more speech from women than men; maternal education influenced child-directed speech quantity

- Infants heard relatively more child-directed speech as child-age increased, due to a diminishing quantity of adult-directed speech in the input

Introduction

Children's real-world language exposure, the input that forms the basis for language development, consists of both child-directed speech (CDS) and adult-directed speech (ADS). Empirical work suggests that children can learn from both of these sources, although CDS appears to be privileged in many contexts. CDS has particular linguistic and acoustic characteristics that differentiate it from ADS (Cristia, 2013; Soderstrom, 2007), and infants prefer CDS over ADS from a young age (Cooper & Aslin, 1990; Cooper, Abraham, Berman, & Staska, 1997; Dunst, Gorman, & Hamby, 2012; ManyBabies Collaborative, 2017; Segal & Newman, 2015).[1]

The characteristics of CDS, and the caregiver-child interactions that accompany CDS, are hypothesized to play a critical role in language development (Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015). Supporting this, recent findings suggest that greater exposure to CDS, but not ADS, is related to faster lexical processing and larger vocabularies in infants and toddlers from multiple socioeconomic and cultural communities (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Rowe, 2008; Shneidman & Goldin-Meadow, 2012; Shneidman, Arroyo, Levine, & Goldin-Meadow, 2013; Weisleder & Fernald, 2013).

Cultural and demographic factors can influence the amount and style of speech directed to children. For instance, research suggests that in the U.S., measures related to socio-economic status (SES), like maternal education, influence the amount and quality of speech heard by children (Hart & Risley, 1995; Cartmill et al., 2013; Rowe, 2012, 2008; Tulkin & Kagan, 1972). When we look across cultures, we also see that

---

[1]https://psyarxiv.com/s98ab

there can be very large differences in the amount of speech directed to children (Cristia, Dupoux, Gurven, & Stieglitz, 2017; Shneidman & Goldin-Meadow, 2012).

Previous research also suggests that the acoustic and linguistic properties of CDS are affected by gender and age. For example, although both mothers and fathers use CDS, its phonetic properties appear to differ by speaker gender (Fernald, 1989; Gleason, 1975; Shute & Wheldall, 1999; Warren-Leubecker & Bohannon III, 1984). Similarly, the gender of the addressed child may also impact CDS. For example, Kitamara and colleagues find pitch differences in speech directed to boys versus girls, though the differences vary across linguistic communities and with the child's age (Kitamura, Thanavishuth, Burnham, & Luksaneeyanawin, 2001; Kitamura & Burnham, 2003), and may not extend to syntactic features (Phillips, 1973). A range of age-related changes to CDS are also well attested, and interact in complex ways with context and linguistic content (e.g.(K. Johnson, Caskey, Rand, Tucker, & Vohr, 2014; Sherrod, Friedman, Crawley, Drake, & Devieux, 1977; Stern, Spieker, Barnett, & MacKain, 1983)). In sum, previous research suggests that caregivers' CDS varies based on the age of the child, the gender of the child, the gender of the caregiver, and, in addition, can be heavily influenced by culture, SES, and other contextual and individual attributes (see Hoff (2006) for a review).

While the work described above has been fundamental in painting the broad strokes of language input to young children, it remains difficult to compare the effects of these demographic variables across studies due to highly variable sampling techniques, data collection methods, and outcome measures. To date, the vast majority of the data on CDS is from relatively short, constrained recordings, predominantly of mother-infant interactions, and often recorded in a lab.

For several reasons, these recording contexts may not capture the patterns of speech that infants typically hear in their daily lives, nor how the input changes as infants develop. First, infants interact with adults other than their mothers. While mothers may modally be young infants' primary caregivers, interactions with fathers, and other caregivers also impact child development, (e.g., Bergelson and Aslin (2017);

Laing and Bergelson (under review); Pancsofar and Vernon-Feagans (2006); Pancsofar, Vernon-Feagans, Investigators, et al. (2010); Shannon, Tamis-LeMonda, London, and Cabrera (2002)).

Quantifying input from both male and female caregivers can help clarify these complementary pathways of influence. Relatedly, in most parts of the world (including North America), few children grow up as only children (United Nations & Social Affairs, 2017). Given that speech from siblings and speech in triadic contexts varies from that in mother-child dyadic interactions (Oshima-Takane & Robbins, 2003), and that birth order and sibling quantity affect language development (Laing & Bergelson, 2017; Oshima-Takane, Goodz, & Derevensky, 1996), sampling from this more complex social context is important for ecological validity.

Second, short interactions between mother and child do not reflect the full range of everyday interactions between caregivers and their children. Indeed, short, structured dyadic interactions seem to overestimate the quantity of speech directed toward children (Bergelson, Amatuni, Dailey, Koorathota, & Tor, accepted pending minor revisions; Tamis-LeMonda, Kuchirko, Luo, Escobar, & Bornstein, 2017). Indeed, even in more naturalistic settings, sensitivity to video recording devices and observers can influence the quantity of speech produced (e.g., Shneidman and Goldin-Meadow (2012)). Finally, sampling at a single age or at a few discrete time-points may obscure whether changes observed over developmental time are gradual, abrupt (Adolph, Robinson, Young, & Gill-Alvarez, 2008), or potentially confounded due to divergent methodology across studies on different age groups.

Daylong Recordings

Recently, there has been an upswing in work that seeks to provide a more ecologically valid option for recording the speech children hear on a day-long timescale. This is due in large part to emerging technologies like LENA (Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011), which provide an unobtrusive long-form recording method supplemented by automated speech analyses.

Such recordings have been used to examine child and caregiver speech to both typically developing children, and those with language delays and disorders, using both manual and algorithm-augmented speech analysis (e.g., Bergelson and Aslin (2017); Dykstra et al. (2012); Gilkerson, Richards, and Topping (2017); Soderstrom and Wittebolle (2013); Suskind et al. (2013); Weisleder and Fernald (2013); Warlaumont, Richards, Gilkerson, and Oller (2014)). Initial studies based on these data have revealed much about the role of children's input and speech volubility on language outcomes. However, automated approaches to classifying the speech children hear as ADS or CDS have been limited (Inoue, Nakagawa, Kondou, Koga, & Shinohara, 2011; Schuster, Pancoast, Ganjoo, Frank, & Jurafsky, 2014; Vosoughi & Roy, 2012), largely due to insufficient quantities of tagged, diverse data on which classifiers can be trained (cf. Schuller et al. (2017)). Thus, studies using these naturalistic, daylong recordings tend to focus on language that happens to occur in the presence of children rather than the putatively more important speech that is directed *toward* children (i.e., CDS).

Some notable exceptions include Weisleder and Fernald (2013) who hand-classified 5-minute sections from LENA recordings of 19- and 24-month-olds according to whether the adult speech contained within each section was primarily between adults or primarily addressed to the child. Taking a similar approach, Ramírez-Esparza, García-Sierra, and Kuhl (2014, 2017) hand-coded 30-second sections within LENA recordings of 11-, 14-, and 33-month olds and coded each 30-second section according to whether it contained multiple adults, and whether the speech to children actually employed a CDS or ADS register. The results of these studies suggest that the quantity of child-directed speech is associated with later vocabulary knowledge, among other findings. However, comparing these studies is difficult, given large differences in their sampling approach. Additionally, neither study analyzed data from children under 11 months of age, thus missing a crucial period of early language development given that much phonological learning (e.g., tuning to language-specific phonemes, language-specific phonotactics, etc.) and even early word learning takes place before this point in development (Bergelson & Swingley, 2012; Seidl, Tincoff, Baker, & Cristia,

2015; Tincoff & Jusczyk, 1999; E. K. Johnson, 2016). These studies also did not consider the relationship between gender or children's exposure to adult speech.

The primary objective of the present study was to examine the quantity and proportion of CDS and ADS that North American children from four cities hear over the first two years of life as a function of child age, child and caregiver gender, and maternal education. While existing research addresses these dimensions separately in smaller samples, the present approach is, to our knowledge, the first to (1) simultaneously examine the effects of these demographic variables, (2) sample in a continuous cross-sectional way from infancy to toddlerhood, and (3) use naturalistic home recordings of young children with their families. Furthermore, we also make the audio data, the labels, and the labeling software readily available in a shared repository for future analyses by other researchers.

We pooled existing corpora of daylong recordings and created a unified set of annotations of the adult speech contained within them. This approach yields ecologically valid measures of ADS and CDS in the speech children hear and allows us to examine the robustness of findings that have been obtained with a diverse array of sampling and coding methods in child language research over the past 50 years.

While the primary goal of the present paper is to examine how demographic variables influence infants' input, the speech-tagging upon which the analyses are based was initially undertaken to create a dataset for speech technology validation, both for the LENA-generated adult gender tags (used by some as ground-truth, e.g., K. Johnson et al. (2014), cf. VanDam and Silbert (2016)), and for the further development of CDS/ADS classification algorithms Schuster et al. (2014); Schuller et al. (2017). Improvement in such algorithms would open the door to analysis of the ADS and CDS already present in the thousands of hours of daylong recordings that are currently collected but not accessible due to the time and financial burdens of completing ADS/CDS coding manually.

Methods

Daylong Recording Selection and Preprocessing

This dataset was created by sub-sampling daylong audio recordings from four North American corpora that come from a larger repository of real-world child language recordings, HomeBank ((VanDam et al., 2016) `homebank.talkbank.org`): Bergelson (Bergelson, 2016), McDivitt (McDivitt & Soderstrom, 2016), VanDam (VanDam, 2016), and Warlaumont (Warlaumont & Pretzer, 2016). All families whose data are used for the present analyses consented to data collection and sharing with authorized researchers via individual IRB protocols for each corpus (University of Rochester, University of Manitoba, Washington State University, and University of California Merced). The secondary analyses conducted here were further approved by the IRB at the authors' institutions, as needed.

All recordings were collected via a LENA audio recorder, which was worn by children in a chest-pocket in specialized clothing (Greenwood et al., 2011). We selected one recording from each of 61 children (29 female) who were typically-developing and primarily heard English at home. We sampled as uniformly as possible between 0 and 2 years across the combined corpora (Figure 1). The recordings had a mean duration of 13.25 hours (median = 13.08; range = 2.67–16.00; two recordings were shorter than 8 hours). Though the corpora came from different labs, we used a unified annotation system to allow for interoperability, pooled coding, and analysis. Table 1 shows demographic information from each of the sub-corpora used.

The recordings were first processed by LENA's proprietary algorithms, which parse the daylong file into utterances with onsets and offsets (hereafter 'clips') and assign a speaker tag to each clip from a closed set of 15 alternatives (e.g., 'Female-Adult-Near', 'Target-Child-Near'). The software also delineates 'conversational blocks', among other features (Greenwood et al., 2011; Oller et al., 2010). Conversation blocks contain at least one sequence of vocalizations separated by silences shorter than 5 seconds; they vary in their duration and in the density of vocalizations found within them.

Table 1

Summary of demographic information from each LENA sub-corpus used from Homebank. Univ. Ed. = university education.

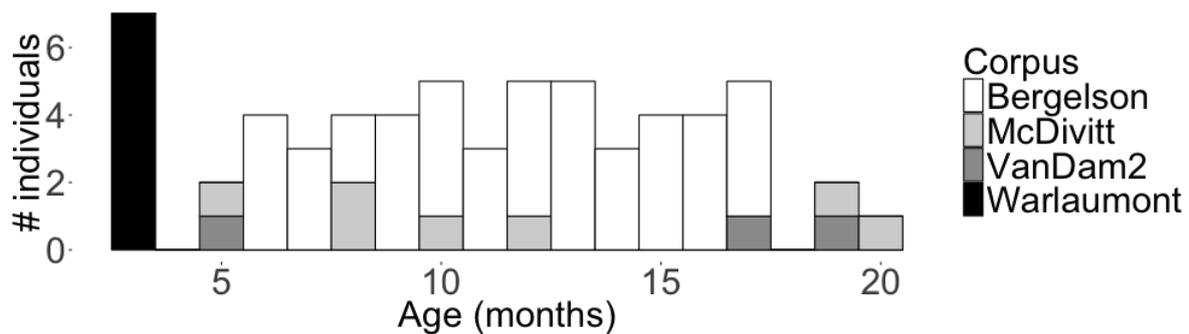| Reference	Location (state)	N (gender)	N w/ older siblings	Age (mo.)	N caregivers w/ univ. ed.
VanDam	Washington	3 (2M)	1	5–19	3
Bergelson	New York	44 (23M)	22	6–17	32
McDivitt	Manitoba	7 (2M)	0	5–20	0
Warlaumont	California	7 (5M)	4	3	7
All	Range	61	27	3–20	42

Figure 1. Age and corpus distribution for the 61 children's recordings included in the present sample.

Speech Clip Annotation and Data Preprocessing

To target adult speech, we randomly selected 20 LENA conversational blocks from each recording (1,220 blocks in total), with the condition that each block have at least 10 Female-Adult-Near (FAN) or Male-Adult-Near (MAN) clips ($n = 12,684$ MAN and FAN clips in total), as determined by the LENA system. This sampling approach let us leverage the LENA algorithm output to ask: of the times in these recordings when adult speech occurs, how is the speech divided among adult male and female speakers and in which register (ADS or CDS) are they speaking? Note this sampling approach does not allow us to measure raw rates of speech or speech types over the course of the day, but it does allow us to compare these rates across our variables of interest. These 1,220 blocks (comprising 43,568 clips in total) were then extracted from the daylong
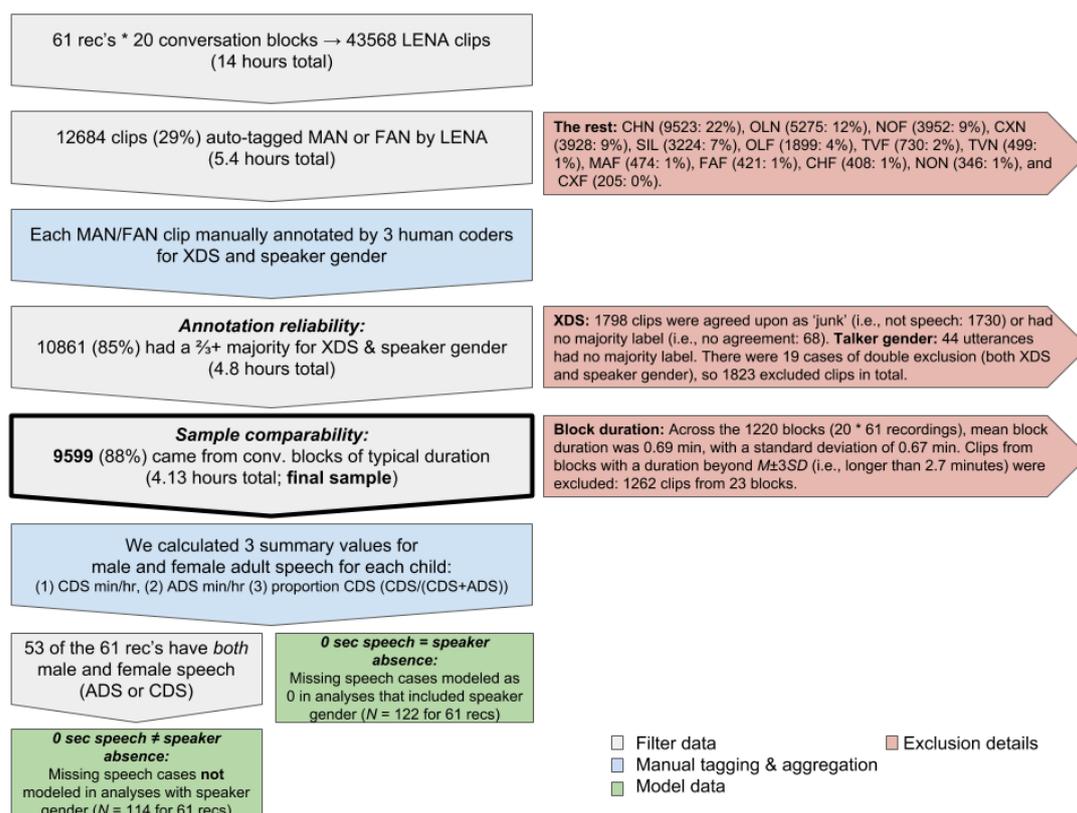
Figure 2. Clip selection process resulting in final sample of 9,599 clips tagged for speaker gender and addressee. XDS = CDS or ADS, MAN = Male Adult Near, FAN = Female Adult Near. The remaining LENA-generated speaker tags (not used for analyses here) are explained in Figure S1 in the Supplemental Information.

audio file and split into their component clips.[2] See Figure 2.

Custom client and server software[3] was written to randomly distribute the blocks to annotators across three labs. Each annotator was first trained on a practice dataset not included in the present analysis. Each FAN and MAN clip was coded by three unique RAs. The client/server system was structured as a RESTful JSON API, allowing us to coordinate both the allocation and ingestion of data across all the participating

---

[2]One could consider other sampling parameters (i.e., number of conversational blocks and minimum number of adult clips. As noted above, these particular parameters were chosen with speech technology extensions in mind. We provide all code and data, and invite interested researchers to explore different sampling and analyses approaches as they see fit: `https://github.com/marisacasillas/NorthAmericanChildren-ADSvsCDS`.

[3]https://github.com/SeedlingsBabylab/idslabel

labs in real time. This architecture allowed individual labs to monitor their progress by querying the back-end servers through their client interface, view and download their own past responses, and add/delete accounts as needed. Separating the front end, which handled audio playback and annotation, from the server back end gave us the flexibility to make updates to both independently, maintaining the flow of annotation and minimizing the frequency of mandatory software upgrades as our code-base evolved.

Across all blocks, RAs tagged 12,684 MAN and FAN speech clips for speaker gender (male/female) and addressee (child/adult) using primarily acoustic-phonetic information (annotation guidelines are summarized in the Online Supporting Information (SI)), with full documentation at (`https://osf.io/d9ac4/`.) Although RAs were permitted to use context to resolve ambiguous cases, they were instructed to make their judgments based on whether the CDS register was being used (i.e., whether it "sounds like" CDS), even if the target was not a child (e.g., a fellow adult or a pet). Annotators achieved high reliability in differentiating male/female (Krippendorff's alpha = .96) and CDS/ADS (Krippendorff's alpha = .82).

Each of the clips was then assigned a 'true' value for speaker gender and speech register if at least two of the three annotators converged on a decision for each value (gender: Male/Female/Junk; register: CDS/ADS/Junk). 'Junk' was used for non-speech and ambiguous speech clips. Clips that did not have a majority true value (only 68 clips for gender and 44 for register), or clips with a majority 'Junk' designation ($n = 1,798$) were removed from the analysis. These exclusionary criteria sometimes overlapped, resulting in 1,823 excluded clips overall. Finally, at the block level, we excluded blocks longer than $\pm 3\text{SD}$ from the mean block length ($n = 23$; 1.8% of blocks). Thus, the final dataset included 9,599 clips produced by adults in 1,197 blocks from 61 recordings. The included blocks sum to 12.5 hours of audio and, in addition to the male and female speech we coded, include (non-coded) clips classified by LENA as overlapping speech, child speech, electronics, or silence. In sum, in the present analysis we analyzed the 4.13 hours of adult speech in these blocks that was found by LENA and verified by human listeners (76% of all FAN and MAN speech identified and 33% of

the total audio time). The audio clips are available to HomeBank members on the

HomeBank Special Projects webpage, as "The IDSLabel Dataset",

`http://media.talkbank.org/HomeBank/Password/Projects/IDSLabel/`. See Figure

2 for an overview of the pipeline for data annotation and filtering.

Results

We separate the results below into analyses of CDS quantity, ADS quantity, and

CDS proportion (i.e., CDS/(CDS+ADS)). For CDS quantity, we ask how much

exposure children had to speech that sounded child-directed. For ADS quantity we ask

how much exposure children had to speech that sounded adult-directed. For CDS

proportion, we ask about the relative amount of CDS children hear. For each of these

dependent variables, we look at both language input overall and then also separately

analyze female versus male adult input.

In each set of analyses, we use linear mixed-effects regression, with random effects

for corpus (all models) and participant (models with >1 datapoint per child). We did

not *a priori* expect predictors to differentially affect CDS or ADS, so our models did

not include added random slopes (see SI for versions with maximal random slopes;

(Barr, Levy, Scheepers, & Tily, 2013)). Given that these analyses are exploratory, we

used an incremental model-building process to systematically test for effects of child

age, child gender, speaker gender, maternal education, and their interactions on each of

our three language input measures (CDS quantity, ADS quantity, and CDS proportion).

In each model, predictors were only included when their addition significantly improved

model fit, according to a likelihood ratio test comparing minimally different models

with and without that predictor. Using this pairwise-testing method we first checked for

the addition of simple predictors, then two-, and finally three-way interactions (see SI

for details). Child age was centered and modeled as a continuous variable (range =

3–20 months, $M = 10.8(4.7)$). Both child and speaker gender were modeled as binary

variables.

Eight children did not hear speech from male speakers in the annotated data at all

(CDS or ADS). When analyzing male and female ADS and CDS patterns below, we cannot infer what male speakers would have done in these eight cases. We therefore removed them from the analyses in which we investigate the influence of speaker gender on quantity of ADS and CDS. Note that we still analyzed female speech patterns from these eight children but simply do not make any conclusions regarding male speech in those eight children's environments.[4]

For maternal education, we split the data into three categories: no-BA, BA, or advanced degree ($n = 19, 16, 26$, respectively). We also considered the number of older siblings for each child, classifying households into three levels: 0 older siblings, 1 older siblings, and 2+ older siblings ($n = 33, 17, 11$).

For CDS and ADS quantity, from which CDS proportion is derived, we compute the input in minutes per hour for each child, thus normalizing across aggregated block length (cumulative ADS + CDS by child: 0.85–9.68 minutes of speech, after the outlier exclusions described above (See Figure 2).

CDS Quantity

On average, children in this dataset heard 11.36 minutes of CDS per hour ($SD = 4.24$). Using our incremental model-building approach to predict CDS quantity, the best model only included maternal education as a predictor, with a random effect of sub-corpus. In this model, every additional level of maternal education (no-BA, BA, advanced degree) resulted in an average of 1.3 more minutes per hour of CDS in a child's input ($B = 1.34$, $SE = .62$, $t= 2.16$; see Figure 3). Neither child age nor child gender improved model fit by model comparison (all $p > .05$).

Dividing these data by speaker gender, infants heard 2–3 times more CDS from women than from men ($M_{female} = 8.6(4.8)$, $M_{male} = 3.2(1.8)$; Figure 4). Indeed, in analogous models estimating CDS quantity by speaker gender (now with an added random effect of child), we find a robust main effect: males produced significantly less CDS than females did ($B = -5.43$, $SE = .83$, $t = -6.56$). Thus, examining CDS

---

[4]See SI for models that include these additional eight datapoints, assigning them a '0' value for male input. The results are very similar to those we present here.
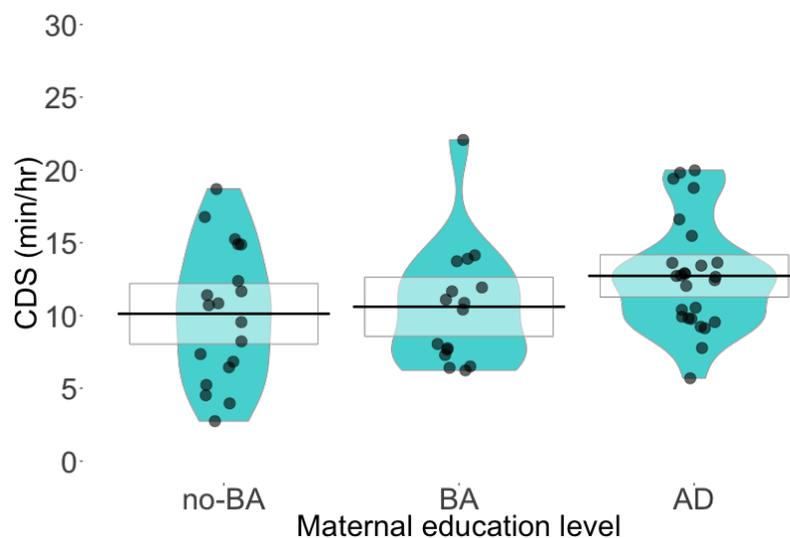
Figure 3. Average CDS minutes per hour, grouped by maternal education level (no-BA, BA, advanced degree). Each datapoint represents the average for one child. The boxes overlaying each distribution indicate 95% confidence intervals.

quantity, we find a strong influence of speaker gender and, when gender is pooled, a more modest influence of maternal education. Notably, none of child age, child gender, or number of older siblings significantly contributed to either model of CDS quantity.

ADS Quantity

Turning to speech directed to adults, infants heard 7.3 minutes of ADS per hour on average ($SD = 6.4$) across all adult speakers in their input. Our model selection process resulted in a model with only one main effect—child age—in addition to a random effect of corpus. The model's estimates are that older children overheard less ADS than younger children ($B = -.76$, $SE = .15$, $t = -5.23$): for each month of child age (0–20 months), infants heard ∼45 fewer seconds of ADS.

Separating ADS quantity by female and male speakers, we again find a robust effect of speaker gender: infants heard significantly less ADS from males ($M = 2.46$ minutes per hour, $SD = 2.63$) than from females ($M = 5.26$, $SD = 5.32$; $B = -2.85$, $SE = .69$, t = -4.15), just as they did with CDS quantity. Alongside speaker gender, the effect of child age remained significant ($B = -.64$, $SE = .10$, $t = -6.33$): for every
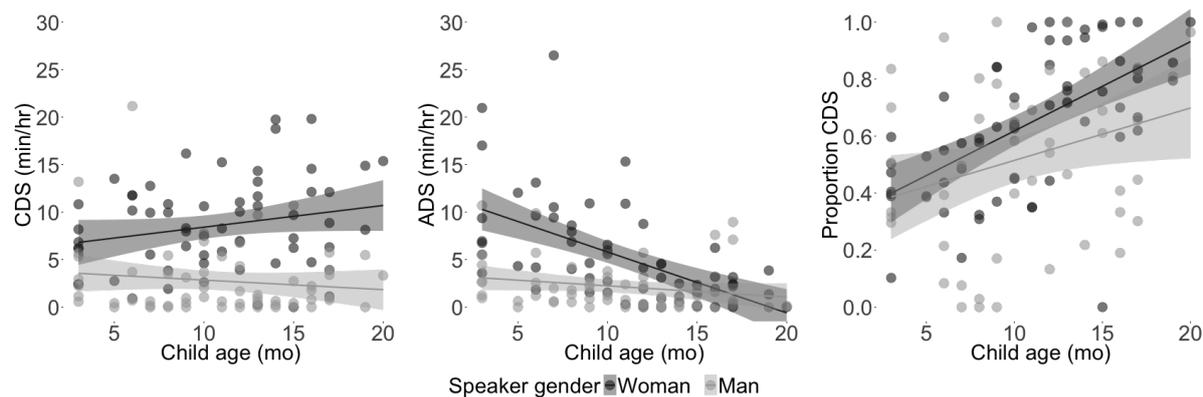
Figure 4. Child age and speaker gender effects on CDS minutes per hour (left), ADS minutes per hour (middle), and proportion CDS (right) for all 61 children. Each child is represented by up to two datapoints: one for speech from male adults and one for speech from female adults. Shaded bands indicate 95% CI.

additional month of child age, infants heard ∼38 fewer seconds of ADS. Finally, the model showed a significant interaction of child age and speaker gender ($B = .54$, $SE = .15$, $t = 3.71$) such that the decrease in ADS per month was smaller for male speakers than for female speakers (Figure 4). Thus, we consistently find strong speaker gender effects on both CDS and ADS quantity. However, we only find effects related to child age on ADS quantity, and only find a minor effect of maternal education on CDS quantity. Neither child gender nor number of older siblings significantly contributed to CDS or ADS quantities. We next turn to an examination of the *proportion* of CDS in our dataset, a metric that permits comparison with previous research more directly.

CDS Proportion

An alternative way of describing children's speech environments is to consider the relative quantity of their input that is adapted to them (i.e., CDS in relation to total CDS and ADS input). In the present dataset, 65% of the speech infants heard from adults sounded child directed ($SD = .22$). This proportion varied widely, with one infant hearing only 17% CDS while another exclusively heard CDS (100%). In our final set of analyses, we again consider the influence of demographic variables, only this time on the *proportion* of CDS infants hear (i.e., minutes of CDS/(minutes of CDS +

minutes of ADS)).

Using the same incremental model-building process as before, we find that, as with ADS, child age significantly improved model fit. Specifically, older children heard a larger proportion of CDS than younger children ($B = .027$, $SE = .005$, $t = 5.30$; Figure 4). Put otherwise, the prevalence of CDS increases at a rate of >2% per month of child age. Consequently, the youngest infants in the sample proportionally hear ∼40% less CDS than the oldest infants.

Separating out speech from female and male speakers in this proportional analysis, we find the same main effects we saw for ADS quantity by speaker gender: there are significant main effects of child age and speaker gender. Older children heard a larger proportion of CDS than younger children ($B = .025$, $SE = .005$, $t = 4.93$), and children heard 11% less CDS from men than from women ($B = -.11$, $SE = .04$, $t = -2.53$). Unlike in our ADS quantity model, however, child age and speaker gender did not interact significantly. Finally, as in the CDS and ADS quantity models, neither child gender nor number of older siblings contributed significantly to model fit for proportion CDS.

Thus, the results for CDS proportion are in line with the results for CDS and ADS quantity: CDS proportion *increases* with child age. When we separate male and female speech, we also see that women produce a higher proportion of CDS than men do.

Discussion

Taken together, our analyses of CDS and ADS in this large naturalistic dataset converge toward several key results: First, we found that while CDS quantity remained largely stable with child age, ADS quantity decreased with child age. Concomitantly, the proportion of CDS increased with child age. Second, infants heard over double the amount of speech from women as they did from men. Third, while effects of maternal education were found for CDS quantity, neither child gender nor number of siblings accounted for significant variance in any analysis. We take up each of these points in turn, but first note a few key differences between our current approach and much of the previous literature.

Our approach is to combine automated methods with manual annotation, to sample from day-long audio recordings drawn from young children's daily lives. The main benefits of this approach are that it allows us to pull our relatively small set of manually-annotated clips from a larger and more representative sample of language input than many previous studies, and to examine both the quantity of speech (i.e., in time units), and the relative proportions of speech (e.g., proportion of CDS in the input). The main drawbacks of this approach are that we have neither transcriptions of what was said, nor exhaustive information about who the other caretakers and other children in every utterance in each recording are. While recognizing the important differences between speech input quantity and quality (e.g., Cartmill et al. (2013); Hirsh-Pasek et al. (2015); Hurtado, Marchman, and Fernald (2008); Rowe (2012)), in this study we limit ourselves to quantity measures as a first step toward harnessing the information in these day-long recordings.

Proportional CDS Increase, Quantitative ADS Decrease

Our estimates of CDS proportion are similar to those found in other studies. For example, Shneidman et al. (2013) finds 69% CDS in North American homes of 14- to 42-month-olds while we find 65% on average across 3–20 months. However, to our knowledge, ours is the first research to report an ADS *decrease* over development. Indeed, this is a clear instance of the complementary nature of examining both time- and proportion-based properties of the speech input. We consider several possible explanations for this pattern.

For instance, as self-locomotion and sitting ability increases, infants may scoot or toddle away from non-child-centric talk, instead sitting and playing independently (e.g., with toys they can now pick up and operate themselves) and thus overhearing less ADS. Research on learning from overhearing suggests that, while 1.5–2-year-olds initially have difficulty learning from overheard speech, they improve by later toddlerhood (Akhtar, Jipson, & Callanan, 2001; Foushee, Griffiths, & Srinivasan, 2016; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011). The children in our dataset (0–20 month-olds) may

therefore struggle to pick up on information from adult-directed speech, which could be consistent with a scooting-away related drop-off in ADS in the present data.

One way to examine this possibility in the current dataset is to see if there is a boost in electronic noise, silence, overlapping speech, or 'far' speech for older children. However, an initial examination of the prevalence of *all* LENA-generated speaker tags in the present dataset (i.e., 43,586 tagged clips from all 1,220 conversational blocks analyzed above), shows that the only tag that increases with age is the target child's (see SI Figures 1 and 2). This initial analysis thus suggests that neither increasing non-speech noise nor increasing distance from adult caregivers can straightforwardly account for the age-related decrease in ADS, or that, if they do, the increasing distance is accompanied by an increased tendency for children to engage in solo vocal or language play.

Another possibility is that as infants themselves begin to produce speech, they are seen as more viable members of a social interaction, and are thus more often included in adult talk. We might then expect a related boost in CDS quantity. However, instead we find that CDS quantity stays stable with age (see Figure 4). Perhaps increased child speech leads to more adult attention to child vocalizations, with adult listening to the child instead of engaging in adult-adult conversation.

A related possibility is that, as infants become older, they may hear less talk between adults since their caregivers may become more likely to divide caregiving time for their increasingly independent child. While the current dataset lacks the intensive level of manual annotation that would be necessary to track which individuals come in close proximity to the infant in each recording, future work could fruitfully examine such variables in both North American homes and in other cultural contexts as a function of age.

Alternatively, it is possible that some aspect of LENA's (proprietary) algorithm may have led to a spurious finding that ADS decreases with age. By design, we only annotated LENA's female-adult-near (FAN) and male-adult-near (MAN) clips within conversational blocks, which were stipulated to contain at least 10 clips, at least 2 of

which were MAN or FAN. If, for instance, LENA's speaker classification accuracy decreases with age, we may have systematically included less adult speech in the clips heard by older children. To test this, we examined the correlation between child age and the proportion of clips manually classified as 'Junk'. We find no correlation between the two ($r = $ -0.003, $p = 0.996$), suggesting that LENA's speaker classification accuracy is not likely to be the source of this effect. A replication of the effect using a different utterance detection and/or sampling strategy would strengthen this conclusion.

Robust Gender Imbalance in the Input

Our second key finding was that, across our demographic variables, children heard over 2.5x more CDS minutes per hour, and over 2x more ADS minutes per hour from women compared to from men; proportionally, females produced 11% more CDS than males. Given the relatively universal prevalence of female caregivers in the first few years of life, this result may be unsurprising to most readers. However, the advance we make in the current study is quantifying this input difference across addressee, age, child gender, and maternal education, in unobserved naturalistic interactions. We fully expect these figures to vary along with differences in early caregiving strategies (e.g., multiple caregivers, child caregivers, or primary male caregivers). These particular data exemplify how caregiving practices—in this case, North American households in which female caregivers predominate—affect the source of linguistic information in a child's environment. Indeed, it is notable that in every model that examined female and male adult input separately, the speaker gender effect was robust, and numerically larger than all other effects that were statistically justified for inclusion in the models.

The prevalence of female input during early childhood has implications for understanding and targeting the naturally occurring dyadic interactions that North American babies experience, and the linguistic, cognitive, and social learning episodes that emerge from them. For instance, in light of these results, theories of phonological development may consider more heavily that infants' native language phonemic representations do not sample from the full range of speaker variability, but are instead

dominated by adult female speakers' phonemes, which themselves are more CDS-heavy than the (minority) input from male adults (cf. Bergelson and Swingley (2017); Martin et al. (2015)). Indeed work with slightly older children has found differences in phonetic realizations of sociolinguistic variants in CDS that differ from those in ADS (Foulkes, Docherty, & Watt, 2005). Further, the finding that most speech to children comes from women may serve to explain female-driven learning biases and preferences in the literature. For example, greater experience with female voices and faces in close proximity may explain why children more readily match vocal gender with facial gender for women than men (Hillairet de Boisferon et al., 2015).

Child Gender and Maternal Education Effects

Our third main finding was that child gender effects did not improve model fit in any of our models, while maternal education effects were only appreciable for CDS quantity, and only when considering all adult speakers collectively, rather than in gender-separated adult input. While previous work has found differences in input based on child gender using other metrics (Kitamura et al., 2001; K. Johnson et al., 2014), we find no evidence of this in our dataset.

In terms of the maternal education effects, we find evidence in line with previous research showing that children whose mothers have lower education hear less speech when we consider all speakers, but not when we separately examine input from female and male adults. This may be because maternal education effects in this sample are relatively small, and thus pooling speaker genders simply increases the amount of data we are analyzing. Furthermore, while maternal education is considered a strong proxy for SES (Bornstein, Hahn, Suwalsky, & Haynes, 2003), SES itself is a complex construct that the present data do not have the granularity to fully examine. Indeed, even our three maternal education categories (<B.A., B.A., >B.A.) leave something to be desired; unfortunately the use of different educational scales across sub-corpora limited the viability of a more gradient approach.

While SES effects on child language input have been found to be robust across

many samples (e.g., Hart and Risley (1995); Hoff (2003)), our current results suggest that when controlling for larger sources of variance (i.e., speaker gender and child age), the effect of maternal education within the (somewhat limited) range examined here is comparatively much smaller. Indeed, we did not select sub-corpora on the basis of diversity in maternal education, and variation in maternal education was not symmetrically spread across our sub-corpora, though sub-corpus was included as a random effect, to prevent an undue statistical influence of any given sub-corpora's characteristics. Future work that maximizes variability in SES in the analyzed participant sample may find relatively larger effects than what we report here.

Relatedly, while previous research found no correlation between the quantity of ADS and CDS in a solely low-SES (i.e., low maternal-education) sample (Weisleder & Fernald, 2013), we do find a positive correlation between the quantity of ADS and the quantity of CDS (Spearman's $rho = .33$, p $= .0089$). There are large methodological differences between the current study and that by Weisleder and Fernald (2013). The families included in the current study had children between ages 3 and 20 months and were primarily middle class (with the exception of the McDivitt corpus, which sampled from a population of low-SES young mothers) while Weisleder and Fernald's input data were collected with 19-month-olds in low-SES Spanish-speaking homes. Additionally, we classify CDS and ADS at the level of the clip while they do so at the level of five-minute chunks. Given these differences between the studies, it is unclear whether why we find a correlation between ADS and CDS and they did not; if both results are reliable, it would suggest that the constellation of factors that influence talk in children's homes may exert differentiated force as a function of maternal education.

Limitations and Future Directions

It is challenging to directly compare the present results with previous work due, in large part, to methodological differences. However, we can compare the LENA-generated adult word counts (AWC) and child vocalization counts (CVC) for the 59/61 recordings in our dataset that were >8 hours with published LENA results, as a

check of potential generalizability (Greenwood et al., 2011; Gilkerson et al., 2017; Soderstrom & Wittebolle, 2013; Zimmerman et al., 2009). While the average AWC count in our corpora was slightly higher than in other papers, both the average AWC and average CVC were within one standard deviation of this previously published work (see SI Table S2).

Looking back on prior work on the quantity of CDS, we found that in a subset of studies where it's possible to quantify the amount of language input to children under age 3, recording duration ranged from 45 minutes to all day, and outcome measures ranged from words per hour (derived from manual transcription), to automated LENA-based Adult Word Counts. While some studies have examined all input to the child, many others have only examined CDS, and none, to our knowledge, have examined ADS separately (SI Table S1 shows the breakdown of studies of each).

An important direction for future studies is to work further with pooled datasets like this one (e.g., those stored on Homebank (VanDam et al., 2016)) to investigate the comparability of different outcome measures and what they tell us about demographic contributions to language development.

One limitation of the present work is that we relied on LENA's automated speaker-tagging to determine which speech clips to annotate (i.e., starting with Male-Adult-Near and Female-Adult-Near clips). This allowed us to analyze a larger sample than would have been possible had we manually segmented the samples. Although LENA output has been vetted for reliability of speaker tagging (Greenwood et al., 2011), and is often used 'as is' (K. Johnson et al., 2014), relying on this output likely added noise to our analyses. We are unable to compare manual annotation of speech register with an automated measure since one is not yet available, but we can compare our gender tags with those automatically generated by LENA. In so doing, we find that that the manual majority tag for speaker gender (which indicates agreement from 2+ of 3 human coders) diverged from the LENA gender tag 22.1% of the time (18.1% of the time for females and 33% for males; e.g., LENA-labeled as 'Female-Adult-Near' speech that was actually male speech or non-speech noise).

The LENA algorithm's mislabelings also varied systematically with speech register: LENA was more likely to mistakenly report male speakers as female when they were using CDS (error rate: 3.8% for ADS and 9.5% for CDS) and more likely to mistakenly report female speakers as male when they are using ADS (error rate: 33.8% for ADS and 21.8% for CDS). This comparison helps us understand LENA's false positive rate but, from our current dataset, it is not possible to estimate the miss rate: adult speech that that was not tagged by LENA as Male- or Female-Adult-Near and therefore not included in our analyses. Similarly, it is possible that with a fully manual segmentation of the input, rather than LENA's automatic segmentation, a different pattern of results would emerge. This is an active area of work in our labs. That said, only including clips where 2+/3 raters agreed on the gender and addressee, and the high levels of reliability achieved, increases confidence in the current findings. Finally, while this sample reflects input heard by over five dozen children from four North American cities, it remains a limited reflection of the full range of demographics in North America and beyond. This too is an important topic for further research, and one that will grow along with databases of home audio recordings collected from different languages and cultures.

Conclusion

Taken together, our results provide first steps towards understanding the distribution of CDS in naturalistic conditions, across maternal education levels and infancy within North American families. Future work is needed to assess the generalizability of these results globally, hopefully with the help of improved computational tools. All the data from the present study are readily available for reanalysis by other researchers and for use in developing automatic labeling methods and we encourage such data re-use. This work also provides an example for how shared datasets with interoperable classification schema can provide more robust analyses than any single lab's data alone; this is an important step forward in addressing replicability and reproducibility in psychology (ManyBabies Collaborative, 2017).

Returning to our initial question: what do North American babies hear? Over a

large, naturalistic, developmental sample, we find that they hear most of their input from female caregivers, and that infants in higher-maternal education families hear more CDS from adults than those in lower-maternal education families. While speech from adults does not appear to be modulated by infant gender in this dataset, we find that, across the board, the speech children hear does vary with their age; young children hear input in a speech register that is increasingly tailored to them as they become more active participants in caregiver interactions over the first two years of life.

References

Adolph, K. E., Robinson, S. R., Young, J. W., & Gill-Alvarez, F. (2008). What is the shape of developmental change? *Psychological Review*, *115*(3), 527–543.

Akhtar, N., Jipson, J., & Callanan, M. A. (2001). Learning words through overhearing. *Child Development*, *72*(2), 416–430.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bergelson, E. (2016). *Bergelson Seedlings HomeBank corpus.* (doi:10.21415/T5PK6D)

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (accepted pending minor revisions). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*.

Bergelson, E., & Aslin, R. N. (2017). Nature and origins of the lexicon in 6-mo-olds. *Proceedings of the National Academy of Sciences*, *114*(49), 12916–12921.

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*(9), 3253–3258.

Bergelson, E., & Swingley, D. (2017). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*.

Bornstein, M. H., Hahn, C.-S., Suwalsky, J. T., & Haynes, O. M. (2003). The hollingshead four-factor index of social status and the socioeconomic index of occupations. *Socioeconomic status, parenting, and child development*, 29–81.

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.

Cooper, R. P., Abraham, J., Berman, S., & Staska, M. (1997). The development of infants' preference for motherese. *Infant Behavior and Development*, *20*(4), 477–488.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, *61*(5), 1584–1595.

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

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2017). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, XX–XX. Retrieved from `http://dx.doi.org/10.1111/cdev.12974` doi: 10.1111/cdev.12974

Dunst, C., Gorman, E., & Hamby, D. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning*, *5*(1), 1–13.

Dykstra, J., Sabatos-DeVito, M. G., Irvin, D., Boyd, B., Hume, K., & Odom, S. L. (2012). Using the Language Environment Analysis (LENA) system in preschool classrooms with children with autism spectrum disorders. *Autism*, *17*, 582–594.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510.

Foulkes, P., Docherty, G. J., & Watt, D. (2005). Phonological variation in child-directed speech. *Language*, *81*(1), 177–206.

Foushee, R., Griffiths, T. L., & Srinivasan, M. (2016). Lexical complexity of child-directed and overheard speech: Implications for learning. *Proceedings of the 38th Annual Conference of the Cognitive Science Society*.

Gilkerson, J., Richards, J. A., & Topping, K. (2017). Evaluation of a lena-based online intervention for parents of young children. *Journal of Early Intervention*, 1053815117718490.

Gleason, J. B. (1975). Fathers and other strangers: Men's speech to young children. *Developmental Psycholinguistics: Theory and Applications*, 289–297.

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby) talk to me: The social context of infant-directed speech and its effects on early language acquisition. *Current Directions in Psychological Science*, *24*(5), 339–344.

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*, 83-92.

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

Hillairet de Boisferon, A., Dupierrix, E., Quinn, P., Loevenbruck, H., Lewkowicz, D., Lee, K., & Pascalis, O. (2015). Perception of multisensory gender coherence in 6- and 9-month-old infants. *Infancy*, *20*, 661–674.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Pace, R. M. G. A., Yust, P. K. S., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071-1083. Retrieved from `https://doi.org/10.1177/0956797615581493` (PMID: 26048887) doi: 10.1177/0956797615581493

Hoff, E. (2003). Causes and consequences of ses-related differences in parent-to-child speech.

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

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

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

Inoue, T., Nakagawa, R., Kondou, M., Koga, T., & Shinohara, K. (2011). Discrimination between mothers' infant-and adult-directed speech using hidden Markov models. *Neuroscience Research*, *70*(1), 62–70.

Johnson, E. K. (2016). Constructing a proto-lexicon: an integrative view of infant language development. *Annual Review of Linguistics*, *2*, 391–412.

Johnson, K., Caskey, M., Rand, K., Tucker, R., & Vohr, B. (2014). Gender differences in adult-infant communication in the first months of life. *Pediatrics*, *134*(6),

e1603–e1610.

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85–110.

Kitamura, C., Thanavishuth, C., Burnham, D., & Luksaneeyanawin, S. (2001). Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language. *Infant Behavior and Development*, *24*(4), 372–392.

Laing, C., & Bergelson, E. (2017). More siblings means lower input quality in early language development. *Proceedings of the 39th Annual Conference of the Cognitive Science Society*.

Laing, C., & Bergelson, E. (under review). The effect of mothers' work schedule on 17-month-olds' productive vocabulary.

Ma, W., Golinkoff, R., Houston, D., & Hirsh-Pasek, K. (2011). Word learning in infant- and adult-directed speech. *Language Learning and Development*, *7*, 185–201.

ManyBabies Collaborative. (2017). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*. (Accepted pending data collection)

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers speak less clearly to infants than to adults. *Psychological Science*, *26*(3), 341-347. Retrieved from `https://doi.org/10.1177/0956797614562453` (PMID: 25630443) doi: 10.1177/0956797614562453

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

Oller, D., Niyogi, P., Gray, S., Richards, J., Gilkerson, J., Xu, D., . . . Warren, S. (2010). Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. *Proceedings of the National Academy of Sciences*, *107*(30), 13354–13359.

Oshima-Takane, Y., Goodz, E., & Derevensky, J. (1996). Birth order effects on early

language development: Do secondborn children learn from overheard speech? *Child Development*, *67*, 621–634.

Oshima-Takane, Y., & Robbins, M. (2003). Linguistic environment of secondborn children. *First Language*, *23*, 21–40.

Pancsofar, N., & Vernon-Feagans, L. (2006). Mother and father language input to young children: Contributions to later language development. *Journal of Applied Developmental Psychology*, *27*(6), 571–587.

Pancsofar, N., Vernon-Feagans, L., Investigators, F. L. P., et al. (2010). Fathers' early contributions to children's language development in families from low-income rural communities. *Early Childhood Research Quarterly*, *25*(4), 450–463.

Phillips, J. R. (1973). Syntax and vocabulary of mothers' speech to young children: Age and sex comparisons. *Child Development*, 182–185.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2014). Look who's talking: speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, *17*, 880-891.

Ramírez-Esparza, N., García-Sierra, A., & Kuhl, P. K. (2017). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontiers in Psychology*, *8*, 1008.

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.

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.

Schuller, B., Steidl, S., Batliner, A., Bergelson, E., Krajewski, J., Janott, C., . . . others (2017). The INTERSPEECH 2017 computational paralinguistics challenge: Addressee, cold & snoring. In *Computational Paralinguistics Challenge (ComParE), Interspeech 2017*.

Schuster, S., Pancoast, S., Ganjoo, M., Frank, M. C., & Jurafsky, D. (2014).

Speaker-independent detection of child-directed speech. In *Spoken Language Technology Workshop (SLT), 2014 IEEE* (pp. 366–371).

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

Seidl, A., Tincoff, R., Baker, C., & Cristia, A. (2015). Why the body comes first: effects of experimenter touch on infants' word finding. *Developmental Science*, *18*(1), 155–164.

Shannon, J. D., Tamis-LeMonda, C. S., London, K., & Cabrera, N. (2002). Beyond rough and tumble: Low-income fathers' interactions and children's cognitive development at 24 months. *Parenting: Science and Practice*, *2*(2), 77–104.

Sherrod, K. B., Friedman, S., Crawley, S., Drake, D., & Devieux, J. (1977). Maternal language to prelinguistic infants: syntactic aspects. *Child Development*, 1662–1665.

Shneidman, L. A., Arroyo, M. E., Levine, S. C., & Goldin-Meadow, S. (2013). What counts as effective input for word learning? *Journal of Child Language*, *40*(3), 672–686.

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.

Shute, B., & Wheldall, K. (1999). Fundamental frequency and temporal modifications in the speech of British fathers to their children. *Educational Psychology*, *19*(2), 221–233.

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

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

Stern, D. N., Spieker, S., Barnett, R., & MacKain, K. (1983). The prosody of maternal

speech: Infant age and context related changes. *Journal of Child Language*, *10*(1), 1–15.

Suskind, D., Leffel, K. R., Hernandez, M. W., Sapolich, S. G., Suskind, E., Kirkham, E., & Meehan, P. (2013). An exploratory study of "quantitative linguistic feedback": Effect of LENA feedback on adult language production. *Communication Disorders Quarterly*, *34*, 1–11.

Tamis-LeMonda, C., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, *XX*, XX–XX.

Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*(2), 172–175.

Tulkin, S. R., & Kagan, J. (1972). Mother-child interaction in the first year of life. *Child Development*, 31–41.

United Nations, D. o. E., & Social Affairs, P. D. (2017). *World Population Prospects: The 2017 Revision.* New York: United Nations.

VanDam, M. (2016). *VanDam2 HomeBank corpus.*

VanDam, M., & Silbert, N. H. (2016). Fidelity of automatic speech processing for adult and child talker classifications. *PloS One*, *11*(8), e0160588.

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

Vosoughi, S., & Roy, D. K. (2012). A longitudinal study of prosodic exaggeration in child-directed speech.

Warlaumont, A. S., & Pretzer, G. M. (2016). *Warlaumont HomeBank corpus.* (doi:10.21415/T54S3C)

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

Warren-Leubecker, A., & Bohannon III, J. N. (1984). Intonation patterns in

child-directed speech: Mother-father differences. *Child Development*, 1379–1385.

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

Zimmerman, F. J., Gilkerson, J., Richards, J. A., Christakis, D. A., Xu, D., Gray, S., & Yapanel, U. (2009). Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics*, *124*(1), 342–349.
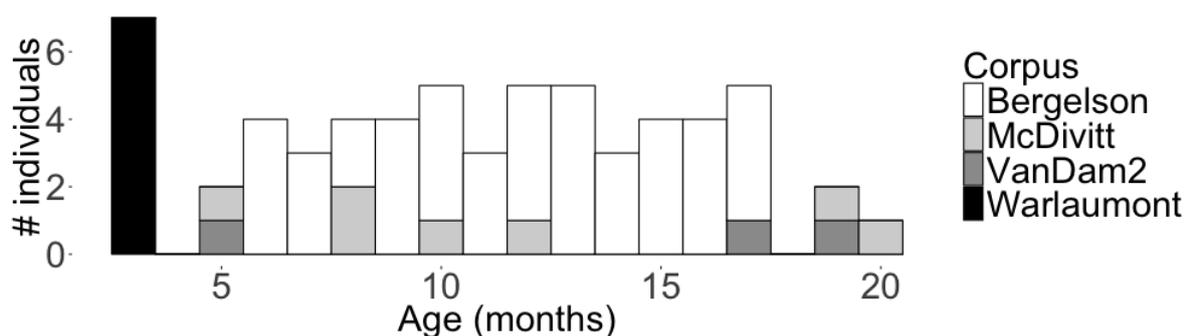
Figure 1. Age and corpus distribution for the 61 children's recordings included in the present sample. (repeated from page 9)
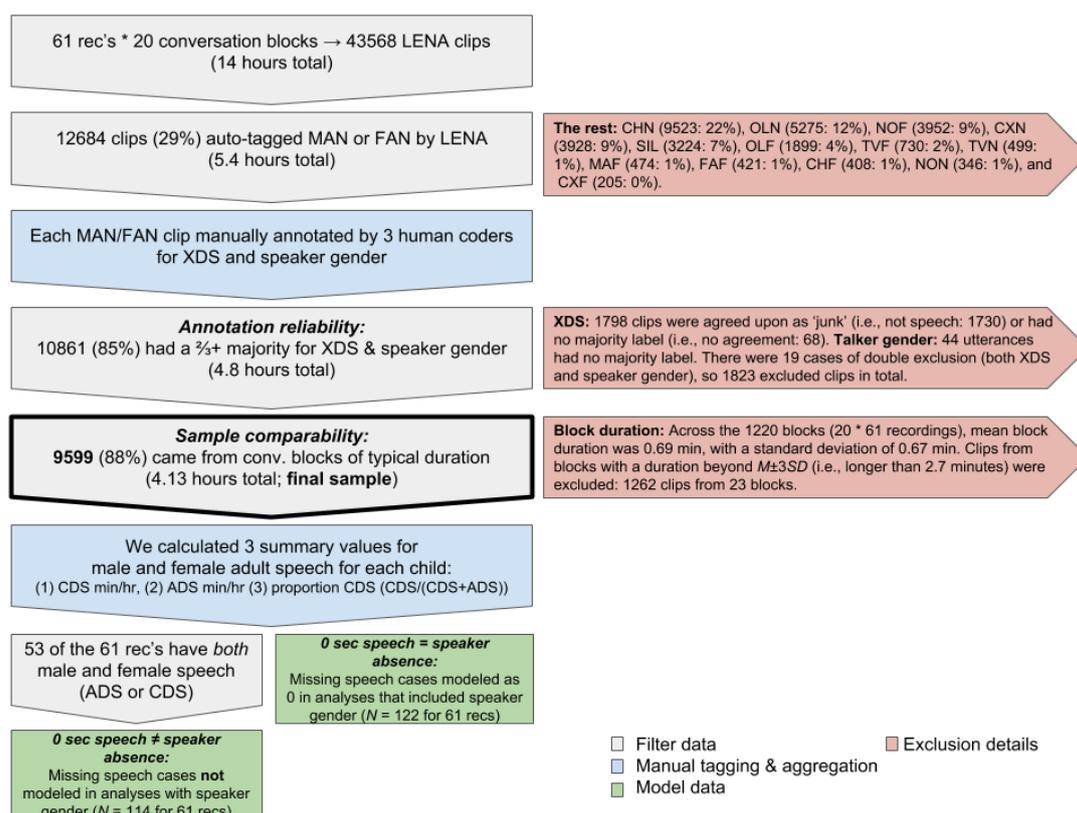

Figure 2. Clip selection process resulting in final sample of 9,599 clips tagged for speaker gender and addressee. XDS = CDS or ADS, MAN = Male Adult Near, FAN = Female Adult Near. The remaining LENA-generated speaker tags (not used for analyses here) are explained in Figure S1 in the SI. (repeated from page 10)
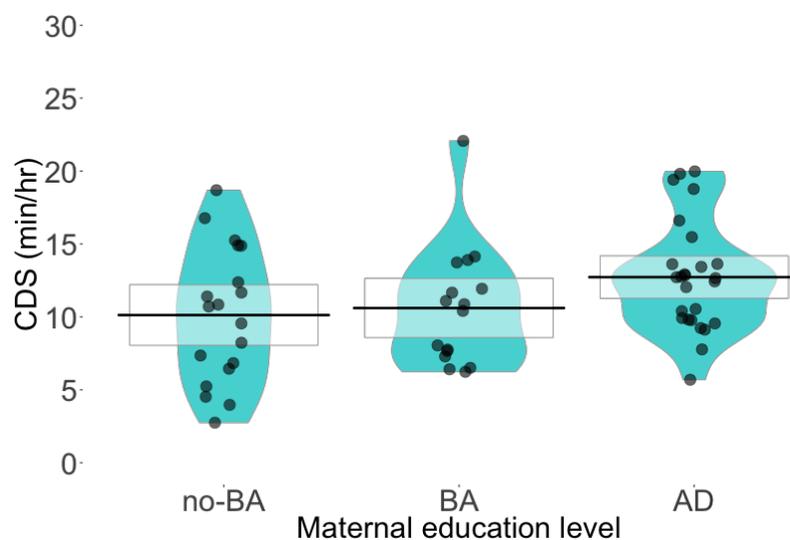
Figure 3. Average CDS minutes per hour, grouped by maternal education level (no-BA, BA, advanced degree). Each datapoint represents the average for one child. The boxes overlaying each distribution indicate 95% confidence intervals. (repeated from page 14)
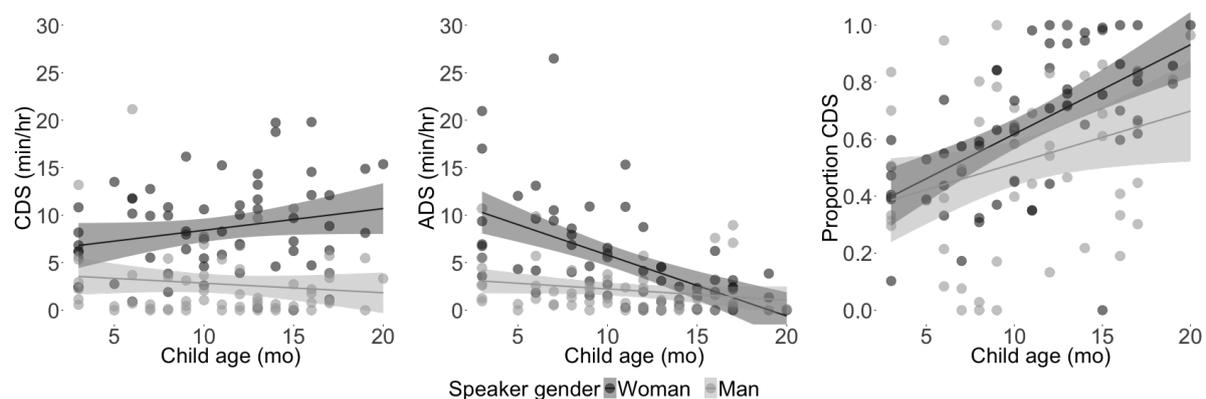

Figure 4. Child age and speaker gender effects on CDS minutes per hour (left), ADS minutes per hour (middle), and proportion CDS (right) for all 61 children. Each child is represented by up to two datapoints: one for speech from male adults and one for speech from female adults. Shaded bands indicate 95% CI. (repeated from page 15)