Supplementary Materials for:
A cross-linguistic 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] Harvard University
[4] Centro Interdisciplinario de Investigaciones en Psicología Matemática y Experimental - CONICET
[5] University of Chicago
[6] Max Planck Institute for Psycholinguistics

Author note

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

Correspondence concerning this article should be addressed to John Bunce, California State University East Bay, Department of Human Development and Women's Studies, 25800 Carlos Bee Blvd, Hayward, CA 94542 USA or Marisa Casillas, University of Chicago, 1101 E. 58th Street, Chicago, IL, 60615 USA. E-mail: john.bunce@csueastbay.edu or mcasillas@uchicago.edu

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, calculated as TCDS + OCDS), and adult-directed speech (ADS). Please see the main manuscript for reasoning and further details. Below we first briefly describe the contents of each sub-section of these Supplementary Materials:

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 the number of talkers in the primary statistical models. This section gives more descriptive information about the number of talkers typical in each language group and preliminarily explores how differences in the 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 the number of talkers and talker type, effects which may partly reflect patterns deriving from cultural-linguistic group routines, practices, and lifestyle differences.

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

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 https://osf.io/pysth/.

Table SM1: *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 (TCDS + 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

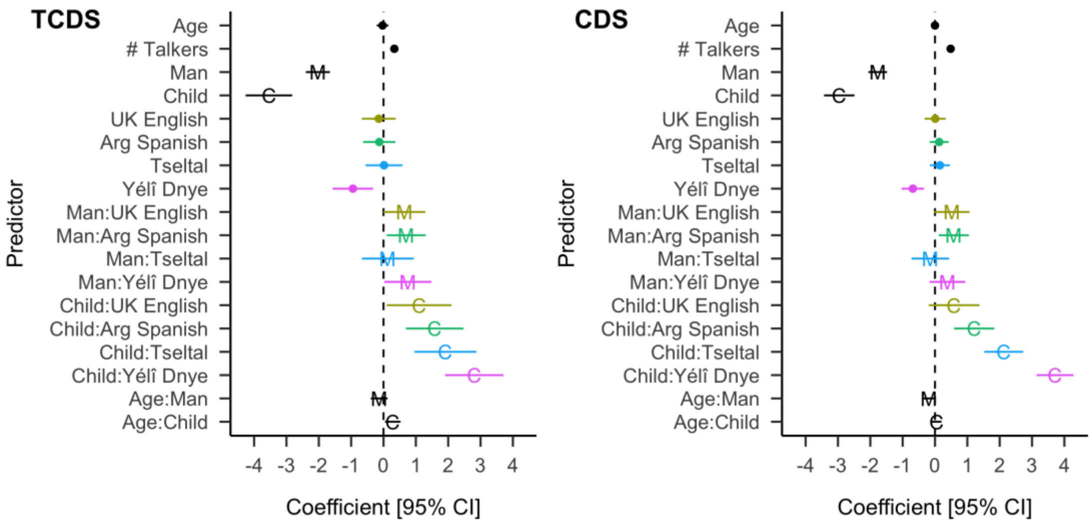

Figure SM1. Coefficients and 95% confidence intervals for the count models of TCDS (left) and CDS (i.e. TCDS + OCDS; right) for all included fixed effects. This figure differs from the similar one in the main text (Main Text Figure 3), which by contrast features TCDS and ADS. Color indicates population (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).

1 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 SM1, 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 $\geq .3$).

SM: CROSS-LINGUISTIC EVERYDAY LANGUAGE EXPERIENCES

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 ($p = 0.73$).

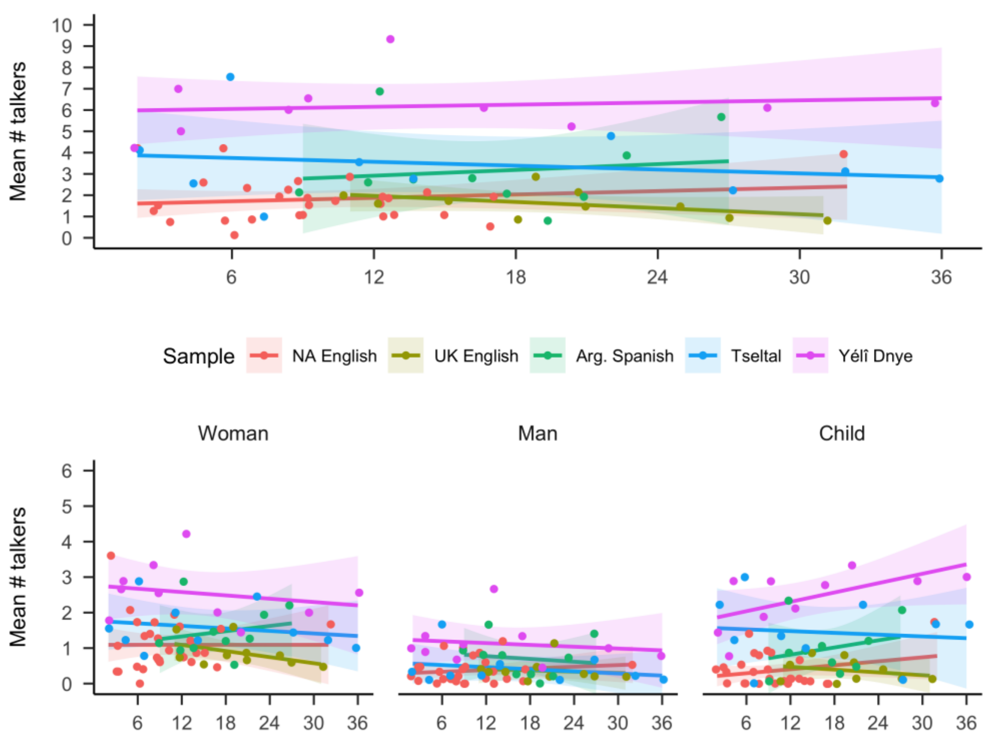

Figure SM2. Number of talkers present across language groups (colors), talker types (top: all talkers; bottom: individual talker types), and target child age (x-axis). Each datapoint represents the mean from one recording.

2 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 the number of talkers in the clip, and suggest that the presence of others has a consistent and substantial effect on how much input children encounter, particularly for ADS. Further, we anticipated that some of the apparent differences 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 the number of present talkers versus other differences in cultural and demographic context, language, and daily routines. Our main-text results thus reflect estimates of group difference controlling for the number of talkers as a separate and significant factor.

However, it is also the case that the number of talkers may systematically differ between language groups in a way that approximates important cross-group 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.

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 SM2. 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 this 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.

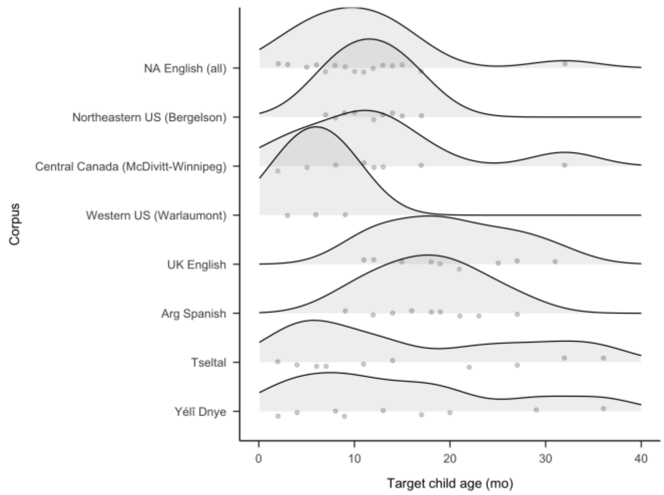

Figure SM3. Target child age for each corpus. Each datapoint represents a child's age from one recording.

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

The distribution of child age across corpora (Figure SM3) 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 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).

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., in the case that further 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 SM4) 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 statistical analysis.

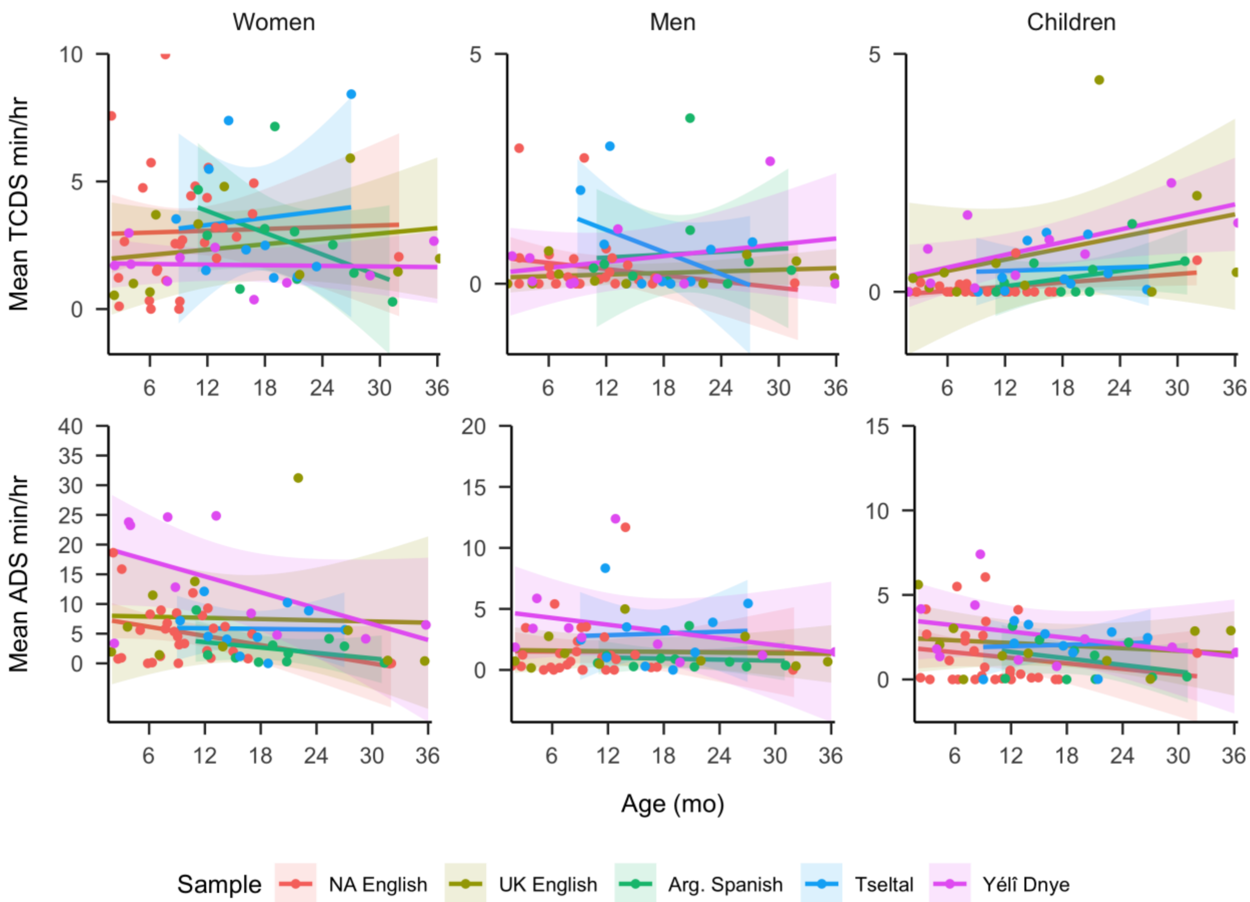

Figure SM4. TCDS (above) and ADS (below) min/hr rates across language groups (color) and talker types (panels), across target child age (x-axis). Each datapoint represents the mean from one recording. This figure is similar to Main Text Figure 2, but now additionally displays the data by child age. As 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.

4 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-group 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 versus 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 the 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.

The first interpretation provides greater nuance, but more importantly, it speaks to useful avenues forward in understanding consistent and observable levers of cross-group 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 group-related 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).

4.1 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 $\geq .15$).

4.2 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 $\geq .2$). 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 ($p = 0.31$).

4.3 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 the 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.

5 Individual North American English data

In the main-text analyses we pool together the North American English datasets. In Figure SM5 we present the primary descriptive figure from the main text, but now only including North American English data, and with datapoints broken out by individual corpus across target child age.

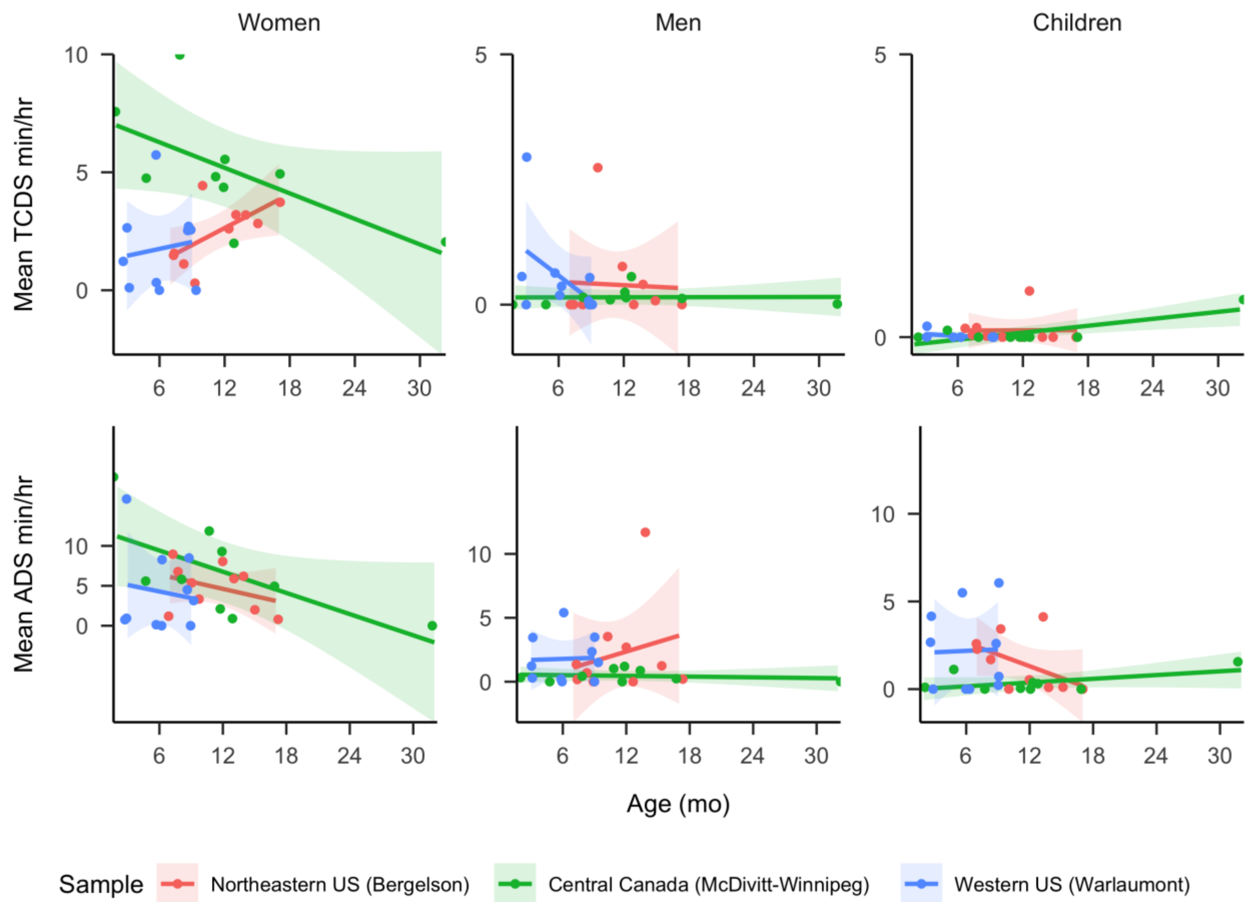

Figure SM5. TCDS (above) and ADS (below) min/hr rates across individual North American English corpora (color) and talker types (panels) across target child age (x-axis). Each datapoint represents the mean from one recording.

6 Full model output for TCDS, CDS, and ADS

The full zero-inflated negative binomial mixed-effects regression output tables for TCDS rate (Table SM2), CDS rate (Table SM3), and ADS rate (Table SM3) are presented below. Along with the output tables of TCDS and ADS from the main text (i.e., Language group reference level = "NA English") we show the alternative models with other reference levels for language group.

Table SM2: *Zero-inflated negative binomial mixed-effects regression output for TCDS min/hr.*

SM: CROSS-LINGUISTIC EVERYDAY LANGUAGE EXPERIENCES

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	Age	-0.03	0.092	-0.312	0.755	NA English
count	Man	-2.03	0.19	-10.693	< 0.001	NA English
count	Child	-3.54	0.367	-9.641	< 0.001	NA English
count	#Spkrs	0.33	0.044	7.619	< 0.001	NA English
count	UKEng	-0.15	0.267	-0.551	0.582	NA English
count	ArgSpa	-0.13	0.254	-0.517	0.605	NA English
count	Tseltal	0.02	0.291	0.06	0.952	NA English
count	Yeli	-0.95	0.319	-2.971	0.003	NA English
count	Man:UKEng	0.64	0.331	1.949	0.051	NA English
count	Child:UKEng	1.10	0.51	2.158	0.031	NA English
count	Man:ArgSpa	0.70	0.307	2.29	0.022	NA English
count	Child:ArgSpa	1.58	0.455	3.484	< 0.001	NA English
count	Man:Tseltal	0.13	0.409	0.326	0.745	NA English
count	Child:Tseltal	1.91	0.488	3.915	< 0.001	NA English
count	Man:Yeli	0.75	0.373	2.017	0.044	NA English
count	Child:Yeli	2.81	0.46	6.109	< 0.001	NA English
count	Age:Man	-0.13	0.132	-1.013	0.311	NA English
count	Age:Child	0.29	0.122	2.348	0.019	NA English
zero-inflation	(Intercept)	-2.67	1.938	-1.377	0.168	NA English
zero-inflation	Age	-1.54	1.397	-1.103	0.27	NA English
zero-inflation	UKEng	-14.31	1941.172	-0.007	0.994	NA English
zero-inflation	ArgSpa	-1.85	9.44	-0.196	0.845	NA English
zero-inflation	Tseltal	0.42	0.913	0.465	0.642	NA English
zero-inflation	Yeli	-1.45	3.607	-0.401	0.689	NA English
count	(Intercept)	1.04	0.191	5.473	< 0.001	UK English
count	Age	-0.03	0.093	-0.275	0.783	UK English
count	Man	-1.38	0.261	-5.312	< 0.001	UK English
count	Child	-2.44	0.349	-6.97	< 0.001	UK English
count	#Spkrs	0.34	0.043	7.791	< 0.001	UK English
count	NAEng	0.08	0.217	0.353	0.724	UK English
count	ArgSpa	0.04	0.242	0.166	0.868	UK English
count	Tseltal	0.03	0.265	0.096	0.923	UK English
count	Yeli	-0.74	0.288	-2.585	0.01	UK English
count	Man:NAEng	-0.64	0.332	-1.941	0.052	UK English
count	Child:NAEng	-1.11	0.51	-2.169	0.03	UK English
count	Man:ArgSpa	0.03	0.335	0.095	0.924	UK English
count	Child:ArgSpa	0.46	0.427	1.086	0.278	UK English
count	Man:Tseltal	-0.50	0.436	-1.145	0.252	UK English
count	Child:Tseltal	0.85	0.448	1.905	0.057	UK English
count	Man:Yeli	0.12	0.414	0.296	0.767	UK English
count	Child:Yeli	1.71	0.431	3.964	< 0.001	UK English
count	Age:Man	-0.14	0.131	-1.039	0.299	UK English
count	Age:Child	0.27	0.123	2.219	0.026	UK English
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.068	UK English
zero-inflation	Age	-1.62	1.281	-1.261	0.207	UK English
count	(Intercept)	1.08	0.193	5.608	< 0.001	Argentinian Spanish
count	Age	-0.03	0.093	-0.275	0.783	Argentinian Spanish
count	Man	-1.35	0.229	-5.907	< 0.001	Argentinian Spanish
count	Child	-1.97	0.266	-7.403	< 0.001	Argentinian Spanish

SM: CROSS-LINGUISTIC EVERYDAY LANGUAGE EXPERIENCES

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	#Spkrs	0.34	0.043	7.791	< 0.001	Argentinian Spanish
count	NAEng	0.04	0.204	0.178	0.859	Argentinian Spanish
count	UKEng	-0.04	0.242	-0.166	0.868	Argentinian Spanish
count	Tseltal	-0.02	0.262	-0.056	0.956	Argentinian Spanish
count	Yeli	-0.78	0.277	-2.827	0.005	Argentinian Spanish
count	Man:NAEng	-0.68	0.299	-2.26	0.024	Argentinian Spanish
count	Child:NAEng	-1.57	0.452	-3.468	< 0.001	Argentinian Spanish
count	Man:UKEng	-0.03	0.335	-0.095	0.925	Argentinian Spanish
count	Child:UKEng	-0.46	0.427	-1.086	0.278	Argentinian Spanish
count	Man:Tseltal	-0.53	0.417	-1.272	0.203	Argentinian Spanish
count	Child:Tseltal	0.39	0.392	0.997	0.319	Argentinian Spanish
count	Man:Yeli	0.09	0.395	0.23	0.818	Argentinian Spanish
count	Child:Yeli	1.24	0.372	3.343	< 0.001	Argentinian Spanish
count	Age:Man	-0.14	0.131	-1.039	0.299	Argentinian Spanish
count	Age:Child	0.27	0.123	2.219	0.026	Argentinian Spanish
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.069	Argentinian Spanish
zero-inflation	Age	-1.62	1.281	-1.261	0.207	Argentinian Spanish
count	(Intercept)	1.07	0.222	4.814	< 0.001	Tseltal
count	Age	-0.03	0.093	-0.275	0.783	Tseltal
count	Man	-1.88	0.356	-5.286	< 0.001	Tseltal
count	Child	-1.58	0.31	-5.108	< 0.001	Tseltal
count	#Spkrs	0.34	0.043	7.791	< 0.001	Tseltal
count	NAEng	0.05	0.235	0.217	0.828	Tseltal
count	UKEng	-0.03	0.265	-0.096	0.923	Tseltal
count	ArgSpa	0.02	0.262	0.056	0.956	Tseltal
count	Yeli	-0.77	0.29	-2.652	0.008	Tseltal
count	Man:NAEng	-0.14	0.404	-0.358	0.72	Tseltal
count	Child:NAEng	-1.96	0.483	-4.059	< 0.001	Tseltal
count	Man:UKEng	0.50	0.436	1.145	0.252	Tseltal
count	Child:UKEng	-0.85	0.448	-1.905	0.057	Tseltal
count	Man:ArgSpa	0.53	0.417	1.272	0.203	Tseltal
count	Child:ArgSpa	-0.39	0.392	-0.997	0.319	Tseltal
count	Man:Yeli	0.62	0.481	1.291	0.197	Tseltal
count	Child:Yeli	0.85	0.397	2.146	0.032	Tseltal
count	Age:Man	-0.14	0.131	-1.039	0.299	Tseltal
count	Age:Child	0.27	0.123	2.219	0.026	Tseltal
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.068	Tseltal
zero-inflation	Age	-1.62	1.281	-1.261	0.207	Tseltal
count	(Intercept)	0.30	0.237	1.263	0.207	Yélî Dnye
count	Age	-0.03	0.093	-0.275	0.783	Yélî Dnye
count	Man	-1.26	0.326	-3.872	< 0.001	Yélî Dnye
count	Child	-0.73	0.275	-2.649	0.008	Yélî Dnye
count	#Spkrs	0.34	0.043	7.791	< 0.001	Yélî Dnye
count	NAEng	0.82	0.253	3.239	0.001	Yélî Dnye
count	UKEng	0.74	0.288	2.585	0.01	Yélî Dnye
count	ArgSpa	0.78	0.277	2.827	0.005	Yélî Dnye
count	Tseltal	0.77	0.29	2.652	0.008	Yélî Dnye
count	Man:NAEng	-0.77	0.377	-2.032	0.042	Yélî Dnye
count	Child:NAEng	-2.81	0.463	-6.077	< 0.001	Yélî Dnye
count	Man:UKEng	-0.12	0.414	-0.296	0.767	Yélî Dnye

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	Child:UKEng	-1.71	0.431	-3.964	< 0.001	Yélî Dnye
count	Man:ArgSpa	-0.09	0.395	-0.23	0.818	Yélî Dnye
count	Child:ArgSpa	-1.24	0.372	-3.343	< 0.001	Yélî Dnye
count	Man:Tseltal	-0.62	0.481	-1.291	0.197	Yélî Dnye
count	Child:Tseltal	-0.85	0.397	-2.146	0.032	Yélî Dnye
count	Age:Man	-0.14	0.131	-1.039	0.299	Yélî Dnye
count	Age:Child	0.27	0.123	2.219	0.026	Yélî Dnye
zero-inflation	(Intercept)	-2.94	1.617	-1.822	0.069	Yélî Dnye
zero-inflation	Age	-1.62	1.281	-1.261	0.207	Yélî Dnye

Table SM3: *Zero-inflated negative binomial mixed-effects regression output for 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	Age	0.00	0.057	-0.015	0.988
count	Man	-1.77	0.147	-12.063	< 0.001
count	Child	-2.96	0.24	-12.309	< 0.001
count	#Spkrs	0.48	0.031	15.821	< 0.001
count	UKEng	0.01	0.165	0.039	0.969
count	ArgSpa	0.13	0.15	0.862	0.389
count	Tseltal	0.15	0.159	0.954	0.34
count	Yeli	-0.69	0.178	-3.855	< 0.001
count	Man:UKEng	0.52	0.278	1.872	0.061
count	Child:UKEng	0.59	0.399	1.48	0.139
count	Man:ArgSpa	0.58	0.238	2.454	0.014
count	Child:ArgSpa	1.21	0.318	3.809	< 0.001
count	Man:Tseltal	-0.15	0.296	-0.494	0.621
count	Child:Tseltal	2.12	0.309	6.885	< 0.001
count	Man:Yeli	0.38	0.281	1.368	0.171
count	Child:Yeli	3.71	0.29	12.782	< 0.001
count	Age:Man	-0.18	0.102	-1.792	0.073
count	Age:Child	0.05	0.077	0.677	0.498
zero-inflation	(Intercept)	-3.91	0.846	-4.629	< 0.001
zero-inflation	Age	0.17	0.505	0.342	0.733

SM: CROSS-LINGUISTIC EVERYDAY LANGUAGE EXPERIENCES

Table SM4: *Zero-inflated negative binomial mixed-effects regression output for ADS min/hr.*

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	(Intercept)	1.16	0.122	9.487	< 0.001	NA English
count	Age	-0.32	0.128	-2.45	0.014	NA English
count	Man	-1.00	0.128	-7.77	< 0.001	NA English
count	Child	-0.89	0.124	-7.163	< 0.001	NA English
count	#Spkrs	0.71	0.033	21.538	< 0.001	NA English
count	UKEng	-0.18	0.267	-0.693	0.488	NA English
count	ArgSpa	0.27	0.222	1.225	0.221	NA English
count	Tseltal	0.18	0.214	0.86	0.39	NA English
count	Yeli	-0.03	0.206	-0.132	0.895	NA English
count	Man:UKEng	0.38	0.281	1.355	0.176	NA English
count	Child:UKEng	0.60	0.261	2.312	0.021	NA English
count	Man:ArgSpa	0.34	0.229	1.481	0.139	NA English
count	Child:ArgSpa	0.18	0.217	0.82	0.412	NA English
count	Man:Tseltal	-0.27	0.278	-0.965	0.335	NA English
count	Child:Tseltal	0.39	0.234	1.669	0.095	NA English
count	Man:Yeli	-0.10	0.214	-0.446	0.655	NA English
count	Child:Yeli	-0.20	0.205	-0.985	0.324	NA English
count	Age:UKEng	0.02	0.256	0.059	0.953	NA English
count	Age:ArgSpa	0.10	0.251	0.39	0.696	NA English
count	Age:Tseltal	0.27	0.169	1.591	0.112	NA English
count	Age:Yeli	-0.11	0.167	-0.636	0.525	NA English
zero-inflation	(Intercept)	-16.40	1614.39	-0.01	0.992	NA English
zero-inflation	UKEng	-8.36	80051.62	0	1	NA English
zero-inflation	ArgSpa	14.35	1614.39	0.009	0.993	NA English
zero-inflation	Tseltal	-3.65	10750.76	0	1	NA English
zero-inflation	Yeli	-1.92	5218.114	0	1	NA English
count	(Intercept)	0.97	0.243	4.008	< 0.001	UK English
count	Age	-0.30	0.222	-1.35	0.177	UK English
count	Man	-0.62	0.25	-2.463	0.014	UK English
count	Child	-0.28	0.23	-1.239	0.215	UK English
count	#Spkrs	0.71	0.033	21.538	< 0.001	UK English
count	NAEng	0.18	0.267	0.693	0.488	UK English
count	ArgSpa	0.46	0.305	1.495	0.135	UK English
count	Tseltal	0.37	0.3	1.231	0.218	UK English
count	Yeli	0.16	0.295	0.534	0.593	UK English
count	Man:NAEng	-0.38	0.281	-1.355	0.175	UK English
count	Child:NAEng	-0.60	0.261	-2.312	0.021	UK English
count	Man:ArgSpa	-0.04	0.314	-0.13	0.897	UK English
count	Child:ArgSpa	-0.43	0.291	-1.464	0.143	UK English
count	Man:Tseltal	-0.65	0.351	-1.846	0.065	UK English
count	Child:Tseltal	-0.21	0.304	-0.702	0.482	UK English
count	Man:Yeli	-0.48	0.303	-1.569	0.117	UK English
count	Child:Yeli	-0.80	0.282	-2.855	0.004	UK English
count	Age:NAEng	-0.02	0.256	-0.059	0.953	UK English
count	Age:ArgSpa	0.08	0.31	0.267	0.789	UK English
count	Age:Tseltal	0.25	0.247	1.024	0.306	UK English
count	Age:Yeli	-0.12	0.247	-0.491	0.623	UK English

SM: CROSS-LINGUISTIC EVERYDAY LANGUAGE EXPERIENCES

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
zero-inflation	(Intercept)	-17.34	1960.801	-0.009	0.993	UK English
zero-inflation	NAEng	0.10	3149.282	0	1	UK English
zero-inflation	ArgSpa	15.30	1960.801	0.008	0.994	UK English
zero-inflation	Tseltal	-2.22	8571.378	0	1	UK English
zero-inflation	Yeli	-0.14	3812.044	0	1	UK English
count	(Intercept)	1.43	0.19	7.527	< 0.001	Argentinian Spanish
count	Age	-0.22	0.217	-1	0.317	Argentinian Spanish
count	Man	-0.66	0.19	-3.453	< 0.001	Argentinian Spanish
count	Child	-0.71	0.178	-3.985	< 0.001	Argentinian Spanish
count	#Spkrs	0.71	0.033	21.538	< 0.001	Argentinian Spanish
count	NAEng	-0.27	0.222	-1.225	0.221	Argentinian Spanish
count	UKEng	-0.46	0.305	-1.495	0.135	Argentinian Spanish
count	Tseltal	-0.09	0.259	-0.338	0.735	Argentinian Spanish
count	Yeli	-0.30	0.25	-1.197	0.231	Argentinian Spanish
count	Man:NAEng	-0.34	0.229	-1.481	0.139	Argentinian Spanish
count	Child:NAEng	-0.18	0.217	-0.82	0.412	Argentinian Spanish
count	Man:UKEng	0.04	0.314	0.13	0.897	Argentinian Spanish
count	Child:UKEng	0.43	0.291	1.464	0.143	Argentinian Spanish
count	Man:Tseltal	-0.61	0.312	-1.95	0.051	Argentinian Spanish
count	Child:Tseltal	0.21	0.267	0.797	0.426	Argentinian Spanish
count	Man:Yeli	-0.44	0.256	-1.699	0.089	Argentinian Spanish
count	Child:Yeli	-0.38	0.241	-1.572	0.116	Argentinian Spanish
count	Age:NAEng	-0.10	0.251	-0.39	0.696	Argentinian Spanish
count	Age:UKEng	-0.08	0.31	-0.267	0.789	Argentinian Spanish
count	Age:Tseltal	0.17	0.243	0.7	0.484	Argentinian Spanish
count	Age:Yeli	-0.20	0.241	-0.85	0.395	Argentinian Spanish
zero-inflation	(Intercept)	-2.05	0.483	-4.236	< 0.001	Argentinian Spanish
zero-inflation	NAEng	-15.54	2930.487	-0.005	0.996	Argentinian Spanish
zero-inflation	UKEng	-20.15	22258.09	-0.001	0.999	Argentinian Spanish
zero-inflation	Tseltal	-16.39	4747.69	-0.003	0.997	Argentinian Spanish
zero-inflation	Yeli	-14.80	2374.242	-0.006	0.995	Argentinian Spanish
count	(Intercept)	1.34	0.181	7.419	< 0.001	Tseltal
count	Age	-0.05	0.109	-0.427	0.669	Tseltal
count	Man	-1.26	0.247	-5.118	< 0.001	Tseltal
count	Child	-0.50	0.199	-2.505	0.012	Tseltal
count	#Spkrs	0.71	0.033	21.538	< 0.001	Tseltal
count	NAEng	-0.18	0.214	-0.86	0.39	Tseltal
count	UKEng	-0.37	0.3	-1.231	0.218	Tseltal
count	ArgSpa	0.09	0.259	0.338	0.735	Tseltal
count	Yeli	-0.21	0.24	-0.881	0.378	Tseltal
count	Man:NAEng	0.27	0.278	0.965	0.335	Tseltal
count	Child:NAEng	-0.39	0.234	-1.669	0.095	Tseltal
count	Man:UKEng	0.65	0.351	1.846	0.065	Tseltal
count	Child:UKEng	0.21	0.304	0.702	0.482	Tseltal
count	Man:ArgSpa	0.61	0.312	1.95	0.051	Tseltal
count	Child:ArgSpa	-0.21	0.267	-0.797	0.426	Tseltal
count	Man:Yeli	0.17	0.301	0.575	0.566	Tseltal
count	Child:Yeli	-0.59	0.257	-2.305	0.021	Tseltal
count	Age:NAEng	-0.27	0.169	-1.591	0.112	Tseltal
count	Age:UKEng	-0.25	0.247	-1.024	0.306	Tseltal

Model component	Term	Estimate (B)	Std. Error	t-value	p-value	Language group reference level
count	Age:ArgSpa	-0.17	0.243	-0.7	0.484	Tseltal
count	Age:Yeli	-0.38	0.154	-2.425	0.015	Tseltal
zero-inflation	(Intercept)	-16.73	2016.091	-0.008	0.993	Tseltal
zero-inflation	NAEng	-1.92	5351.675	0	1	Tseltal
zero-inflation	UKEng	-6.70	41081.03	0	1	Tseltal
zero-inflation	ArgSpa	14.68	2016.091	0.007	0.994	Tseltal
zero-inflation	Yeli	-0.84	3962.023	0	1	Tseltal
count	(Intercept)	1.13	0.173	6.525	< 0.001	Yélî Dnye
count	Age	-0.42	0.109	-3.874	< 0.001	Yélî Dnye
count	Man	-1.09	0.172	-6.35	< 0.001	Yélî Dnye
count	Child	-1.09	0.164	-6.658	< 0.001	Yélî Dnye
count	#Spkrs	0.71	0.033	21.538	< 0.001	Yélî Dnye
count	NAEng	0.03	0.206	0.132	0.895	Yélî Dnye
count	UKEng	-0.16	0.295	-0.534	0.593	Yélî Dnye
count	ArgSpa	0.30	0.25	1.197	0.231	Yélî Dnye
count	Tseltal	0.21	0.24	0.881	0.378	Yélî Dnye
count	Man:NAEng	0.10	0.214	0.446	0.655	Yélî Dnye
count	Child:NAEng	0.20	0.205	0.985	0.324	Yélî Dnye
count	Man:UKEng	0.48	0.303	1.569	0.117	Yélî Dnye
count	Child:UKEng	0.80	0.282	2.855	0.004	Yélî Dnye
count	Man:ArgSpa	0.44	0.256	1.699	0.089	Yélî Dnye
count	Child:ArgSpa	0.38	0.241	1.572	0.116	Yélî Dnye
count	Man:Tseltal	-0.17	0.301	-0.575	0.566	Yélî Dnye
count	Child:Tseltal	0.59	0.257	2.305	0.021	Yélî Dnye
count	Age:NAEng	0.11	0.167	0.636	0.525	Yélî Dnye
count	Age:UKEng	0.12	0.247	0.491	0.623	Yélî Dnye
count	Age:ArgSpa	0.20	0.241	0.85	0.395	Yélî Dnye
count	Age:Tseltal	0.38	0.154	2.425	0.015	Yélî Dnye
zero-inflation	(Intercept)	-18.10	4444.639	-0.004	0.997	Yélî Dnye
zero-inflation	NAEng	1.45	4805.431	0	1	Yélî Dnye
zero-inflation	UKEng	-4.61	29083.8	0	1	Yélî Dnye
zero-inflation	ArgSpa	16.05	4444.639	0.004	0.997	Yélî Dnye
zero-inflation	Tseltal	-0.92	7749.293	0	1	Yélî Dnye

7 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 raw data values. However, our statistical analyses take into account a number of factors that are not visible in that main-text plot. Here we show marginal-means plots from the models of TCDS (Figure SM6) and ADS (Figure SM7), displaying estimated values of each over child age, and talker type, for each language group

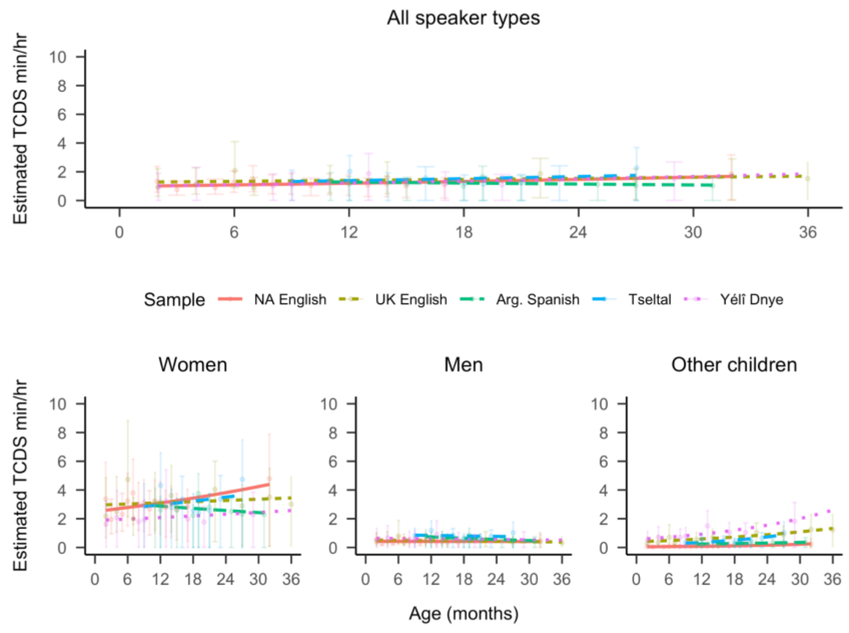

Figure SM6. Model-estimated rates of TCDS min/hr by target child age (x-axis) and corpus (color), shown all together in the top panel and split by talker type in the bottom panel. Each datapoint represents the estimated mean from one recording.

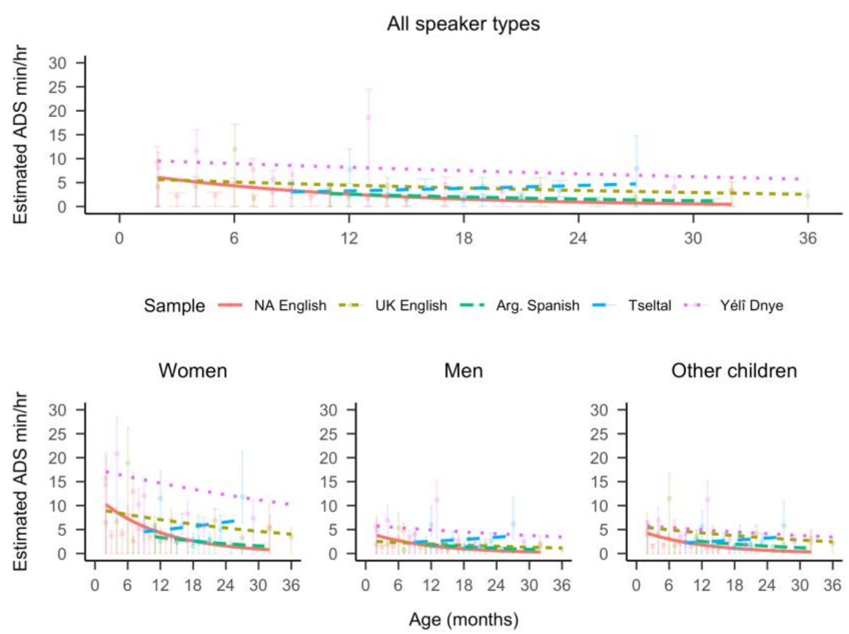

Figure SM7. Model-estimated rates of ADS min/hr by target child age (x-axis) and corpus (color), shown all together in the top panel and split by talker type in the bottom panel. Each datapoint represents the estimated mean from one recording.

8 Confusion among addressee annotation types

The agreement scores for addressee ranged from slight to substantial across corpora. In this final section of the Supplementary Materials, we give further information about how 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 (Figure SM8). 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.

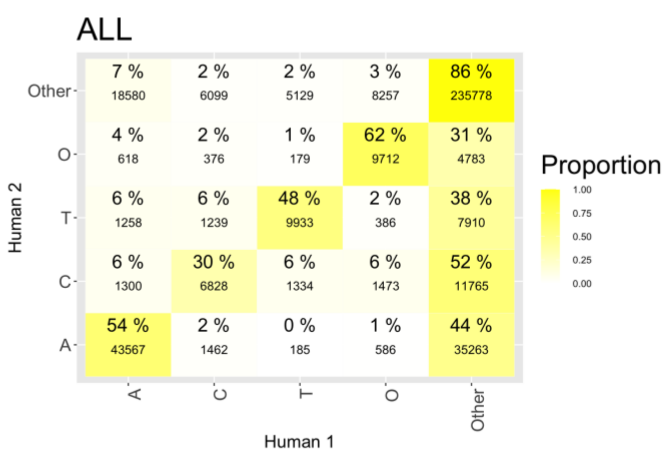

Figure SM8. Confusion matrix across all corpora for addressee annotations between the two human annotators.

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 (Figures SM9–SM15). 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).

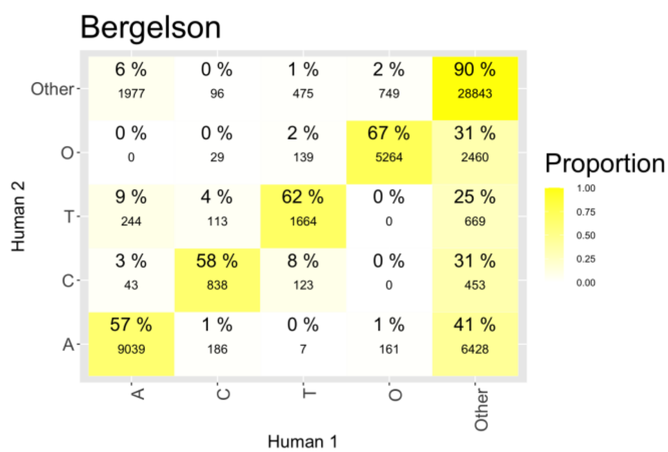

Figure SM9. Confusion matrix for the Bergelson (US English) corpus for addressee annotations between the two human annotators.

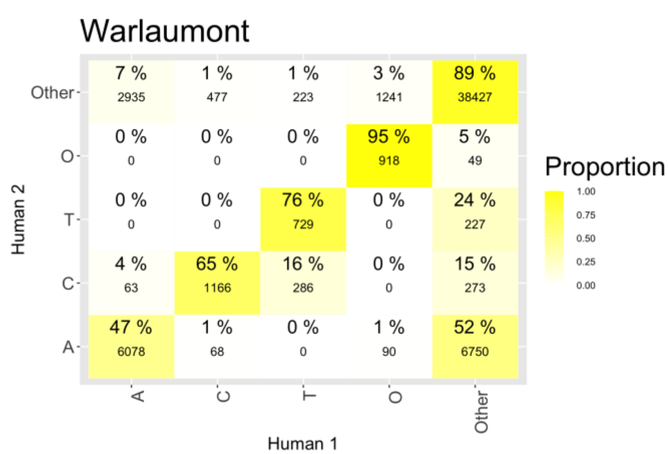

Figure SM10. Confusion matrix for the Warlaumont (US English) corpus for addressee annotations between the two human annotators.

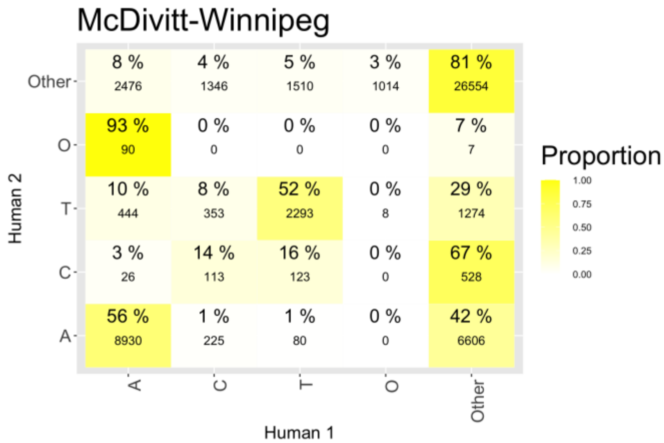

Figure SM11. Confusion matrix for the McDivitt/Winnipeg (Canadian English) corpus for addressee annotations between the two human annotators.

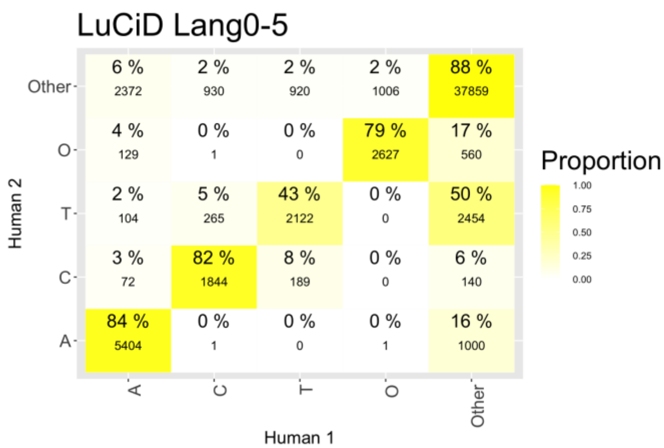

Figure SM12. Confusion matrix for the LuCiD Lang0–5 (UK English) corpus for addressee annotations between the two human annotators.

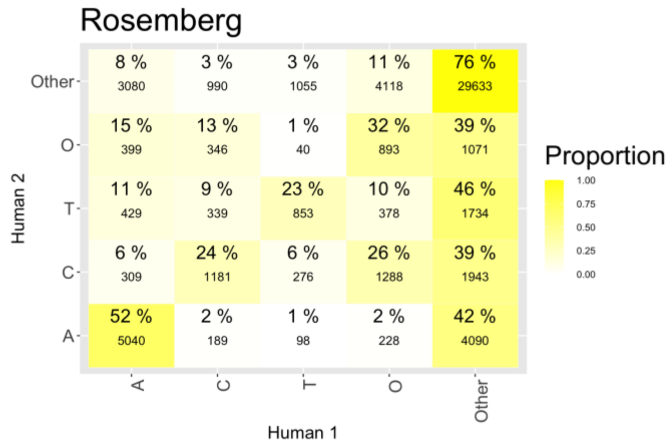

Figure SM13. Confusion matrix for the Rosemberg (Argentinian Spanish) corpus for addressee annotations between the two human annotators. Reliability annotators had non-native familiarity with a different variety of Spanish.

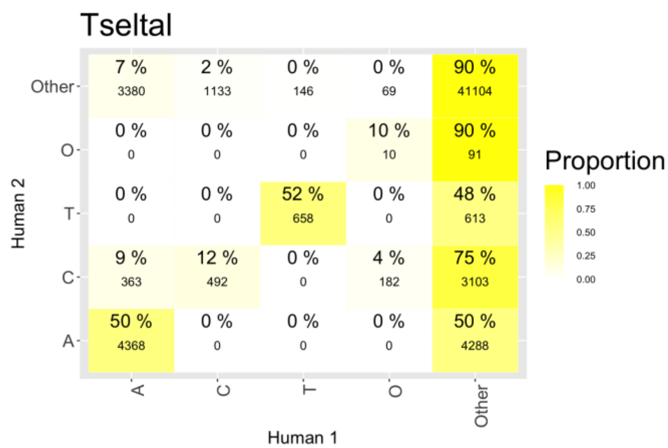

Figure SM14. Confusion matrix for the Tseltal corpus for addressee annotations between the two human annotators. Reliability annotators were completely unfamiliar with the language but there are occasional wordforms used in Spanish (e.g., borrowings) that may have been familiar to the annotators.

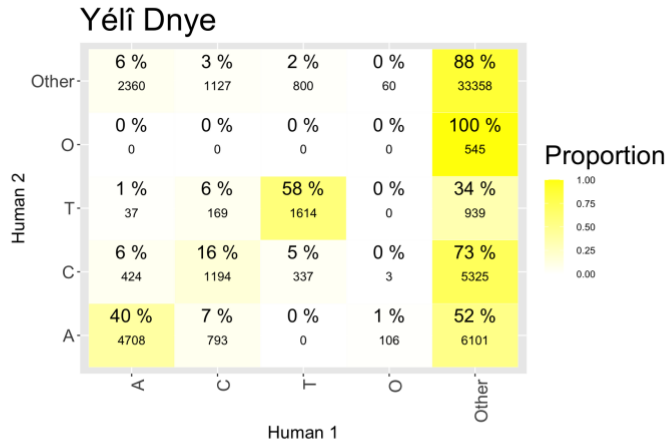

Figure SM15. Confusion matrix for the Yélî Dnye corpus for addressee annotations between the two human annotators. Reliability annotators were completely unfamiliar with the language but there are occasional wordforms used in English and the English-related creole Tok Pisin (e.g., borrowings) that may have been familiar to the annotators.

9 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