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

Table of Contents

Anonymous1

1 Institution

Author note

Redacted for anonymity

Correspondence concerning this article should be addressed to Anonymous, Address. E-mail: [name@institution.edu](mailto:name@institution.edu)

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

Before describing the contents of these Supplementary Materials, we remind the readers that our analyses are centered on the dependent variables of minutes-per-hour rate of target-child-directed language (TCDL), other-child-directed language (OCDL), all child-directed language (TCDL + OCDL), and adult-directed language (ADL). Please see the main manuscript for reasoning and details.

For those interested in knowing about the *total quantity* of child-directed language in the children’s environments (i.e., all input that is designed for children addressees), the first section includes analyses of all child-directed language (CDL) that parallel what is reported in the main text for target-child-directed language (TCDL).

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

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

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

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

The sixth section gives tabular regression outputs for the full binomial mixed-effects regression models of TCDL, CDL, and ADL from which the main-text and section 1 Supplementary Material results are reported.

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

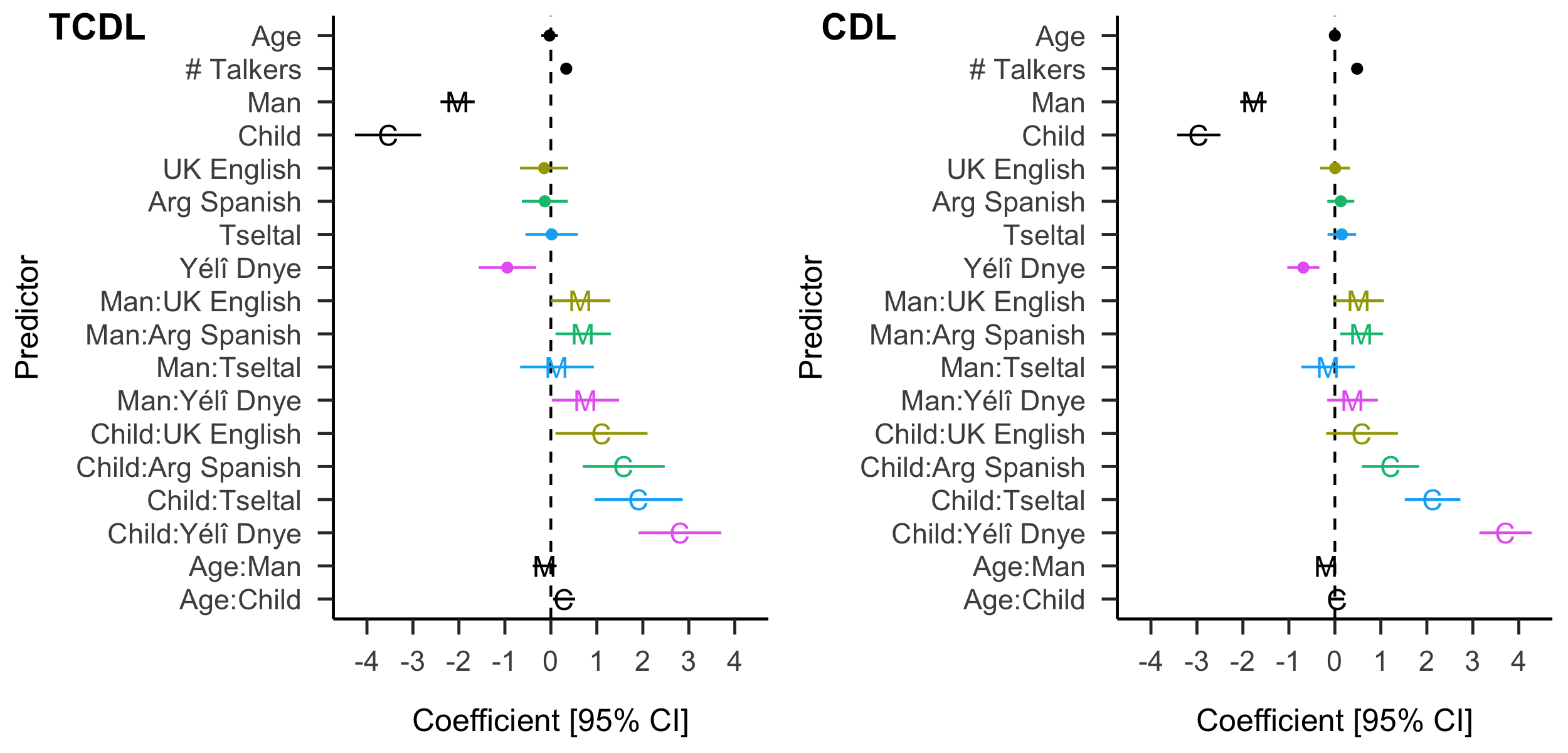
The eigth and final section shows confusion matrices for addressee-type annotations (e.g., target-child vs. other-child status of an utterance) overall and for each contributed corpus individually.

As in the main text for this study, all statistical analyses were conducted in R with the glmmTMB package (Brooks et al., 2017; R Core Team, 2019) and all figures were generated with ggplot2 (Wickham, 2016). Analysis scripts and anonymized data are available at URL\_MASKED\_FOR\_REVIEW.

Table 1:

*Average input rates per clip across participants for each corpus. Parentheses following the mean indicate the median and range across participants. OCDL indicates rate of input directed to non-target-child children; CDL sums rates of TCDL and OCDL.*

| Language | TCDL rate | OCDL rate | ADL rate | Mean proportion TCDL | Mean proportion any CDL |
| --- | --- | --- | --- | --- | --- |
| NA English | 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 |
| Arg. Spanish | 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 | 3.54 (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 |
| Tseltal | 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 |
| Yélî Dnye | 3.13 (2.95; 1.58-6.26) | 12.77 (13.57; 4.61-20.9) | 19.87 (17.1; 7.25-38.54) | 0.09 | 0.44 |



*Figure* *1.*  Coefficients and 95% confidence intervals for the count models of TCDL (left) and CDL (i.e. TCDL + OCDL; right) for all included fixed effects. This figure differs from the similar one in the main text, which by contrast features TCDL and ADL, not CDL. Color indicates population (North American English is the modeled reference level), C and M indicate effects related to child- and man-produced utterances, respectively (woman-produced utterances are set as the model reference level).

# 1 All child-directed language (CDL)

The analysis of directed linguistic input in the main text focuses on TCDL; 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 facilitated language learning. We here analyze *all* CDL in the recordings using the same factors as we did for TCDL in the main text. “CDL” here includes all utterances directed to the target child, plus all other observable child-directed language in the audio recording clips, including input directed at groups of children that may or may not include the target child. In other words “CDL” is here all hearable utterances that is directed to a child in the recording; comparable to what is measured, e.g., in Bergelson et al. (2019). Therefore this measure of CDL 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 CDL to that of TCDL reported in the main text.

On average, across all corpora children were exposed to 7.43 minutes of CDL per hour across audio clips (median = 6.18), with wide variation between children (range = 0.37–24.11). Our model of CDL rate was nearly identical to that used for TCDL rate in the main text: It included target child age, talker type, the number of talkers present in the clip, and linguistic community (i.e., corpus), with two additional two-way interactions (talker type by linguistic community 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 TCDL was that we did not include linguistic community as a predictor in the zero-inflation component because its inclusion caused model non-convergence. As a reminder, there was no significant effect of corpus in the zero-inflation model component of the main-text TCDL model. This fact, together with the qualitatively similar pattern of findings for CDL 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 = CDL\_min\_p\_hr ~ child.age + talker.type + num.spkrs + corpus + talker.type:corpus + child.age:talker.type + (1 | child\_id), ziformula = ~ child.age). The results are qualitatively highly similar to the TCDL model presented in the main text. The coefficients and 95% confidence intervals for all fixed effects in the CDL count model are shown in Figure 1, side by side with the same plot from the TCDL model, replicated from the main text.

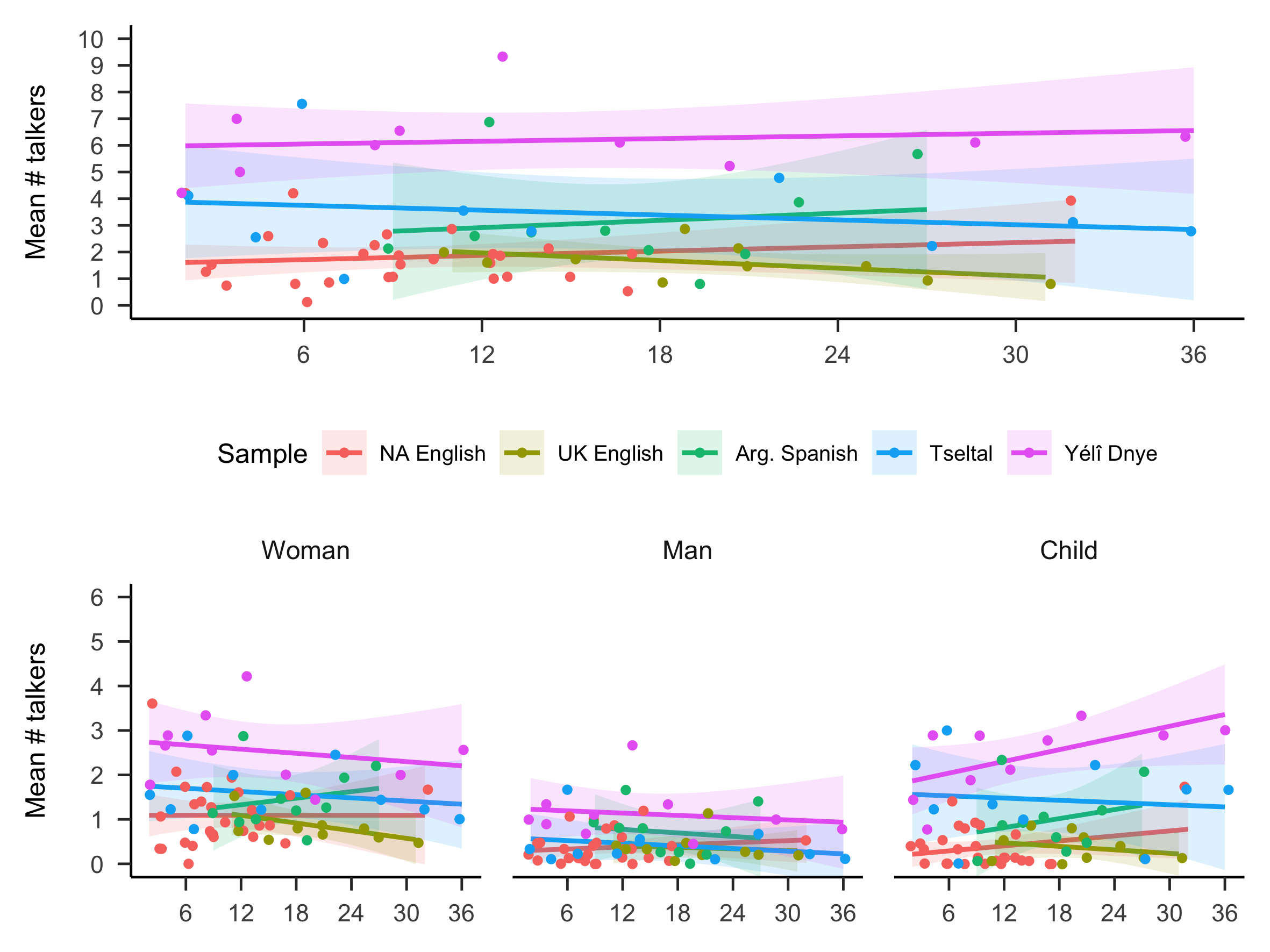
CDL input rate significantly differed by talker type, number of talker present, linguistic community, and the interaction between talker type and linguistic community. As with TCDL rate, CDL rate was significantly lower for men compared to women (*B* = -1.77, *SE* = 0.15, *z* = -12.06, *p* < 0.001) and for children compared to women (*B* = -2.96, *SE* = 0.24, *z* = -12.31, *p* < 0.001). CDL rate was, like TCDL rate, also significantly higher when there were more talkers present (*B* = 0.48, *SE* = 0.03, *z* = 15.82, *p* < 0.001).

As with TCDL, rates of CDL 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 corpora (all *p* > .3).

Interactions between talker type and linguistic community were overall similar, with some small differences. Men were previously found to produce significantly more TCDL in the Argentinian Spanish and Yélî Dnye samples (and marginally so for UK English) compared to North American English. When this measure is changed to CDL, 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) and the UK English difference remains marginal (*B* = 0.52, *SE* = 0.28, *z* = 1.87, *p* = 0.06). Children were previously found to produce significantly more TCDL in all four of the non-North American English samples compared to North American English. When this measure is changed to CDL, 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 TCDL and CDL. Whereas the previous analysis suggested that child-produced, but not man-produced, TCDL grows more with age compared to woman-produced TCDL, there are no significant differences across age by talker type with the input measure of CDL. The only evidence for an age-related change by talker type in CDL is marginally slower change in CDL over age for man-produced CDL compared to woman-produced TCDL (*B* = -0.18, *SE* = 0.10, *z* = -1.79, *p* = 0.07).

Like the model of TCDL rate, the zero-inflation component of the CDL rate regression suggested that clips with no CDL were not more likely for younger children (*p* = 0.73).



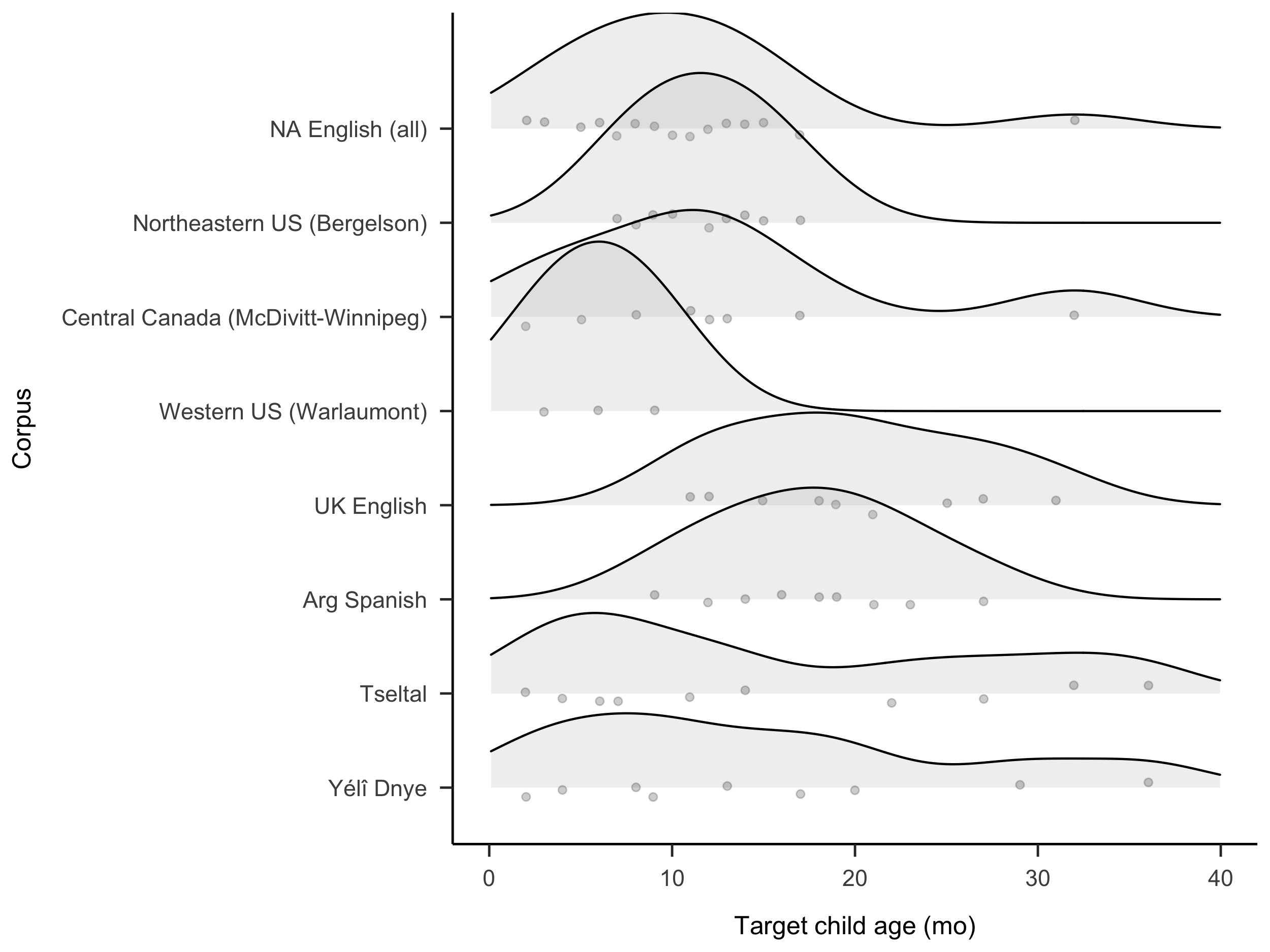
*Figure* *2.*  Number of talkers present across language groups, talker types, and target child age. Each datapoint represents the mean from one recording.

# 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 TCDL, CDL, and ADL, all show very strong effects of number of talkers in the clip, and suggest that the presence of others has a more-than-minimal effect on how much input children encounter, particularly for ADL. Further, we anticipated that some of the differences apparent between linguistic communities is 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-site differences in input rates is simply due to number of present talkers vs. other differences in culture, language, and daily life. Our main-text results thus reflect estimates of cultural difference controlling for number of talkers as a separate and significant factor. However, it is also the case that number of talkers may systematically differ between sites in a way that approximates important cultural differences. We here analyze the number of talkers detected in clips across sites and age, both overall and by talker type. Our aim is to illustrate the scale of cross-site 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 culture- and community-specific practices.

On average, and in addition to any vocalizations by the target child, a given child’s 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 crosscultural variation, which is apparent in Figure 2. In particular, the Yélî Dnye recordings show much higher rates of other talker presence compared to the other corpora, with averages of 2.54 women, 1.12 men, and 2.40 children. Compare to North American English with averages of 1.09 women, 0.36 men, and 0.36 children and UK English, with 0.86 women, 0.36 men, and 0.37 children. The Tseltal and Argentinian Spanish communities fall between, 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 presence of women contributing input and a slight uptick in children contributing input in some communities. 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. Thereby the current measure gives insight into effects of household and routine by site, but not adequately to make well substantiated claims at present.



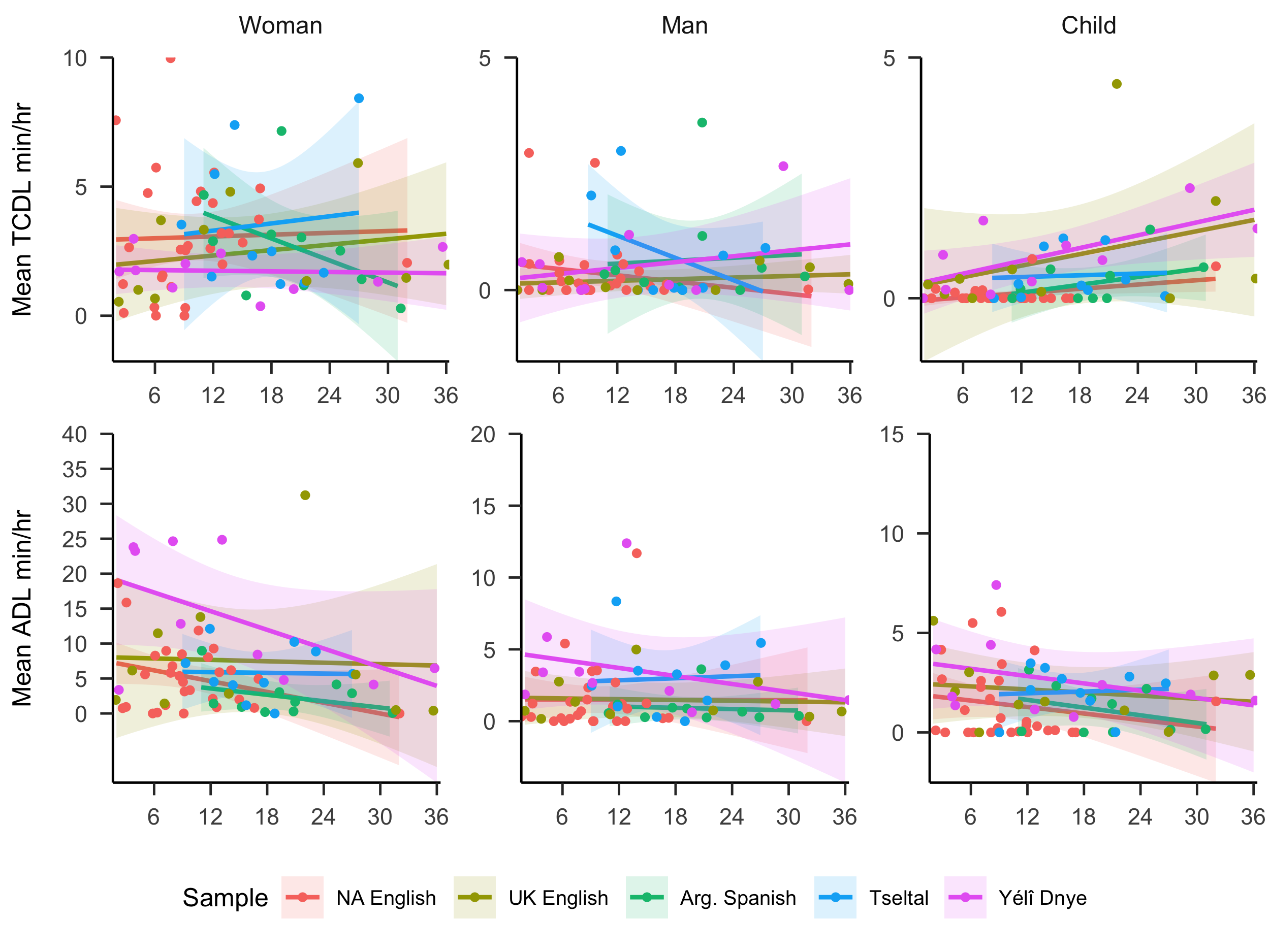
*Figure* *3.*  Target child age for each corpus. Each datapoint represents the mean from one recording.

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

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

In principle, age differences between corpora, even if present, are not necessarily a problem for the present study—in layperson’s 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 corpora could mean that there are true age effects present in the data that we cannot detect under present circumstances (i.e., supposing increased data would substantially change the linear fits of age in the model). As a reminder, we found no evidence for an overall effect of target child age (neither a decrease nor an increase) in the primary models of child-directed language (i.e., TCDL in the main text and CDL here in the supplementary materials). We do find a significant decrease in ADL associated with child age. In our view, the primary concern is then whether we are missing overall effects of child age on TCDL and CDL. However, considering the much better age coverage of our four other corpora—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 CDL 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 3) so that the interested reader can get some informal and qualitative impression of potential differences that might be detected if more data were to be added in future work. Please note that any apparent visual differences here, as in the main text figures and tables, do not have the benefit of random-effects controls that are applied in our analysis via the use of statistical models.



*Figure* *4.*  TCDL (above) and ADL (below) min/hr rates across language groups and talker types across target child age. Each datapoint represents the mean from one recording. This figure is similar to Figure 2 in the main text, but now additionall displays the data by child age. As is apparent, age effects are minimal for TCDL across corpora and talker type whereas there is a general decrease in ADL across corpora and talker type.

# 4 Simple models of age and cross-site difference in TCDL, CDL, and ADL rates

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

By including these factors in the model, as we have in the main-text models of TCDL and ADL and in the model of CDL above, we can gain a more detailed perspective on the shared features that drive variation in input rate between and within corpora. 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, Tenejapa, 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 crosscuts the corpora, 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 ADL 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 vs. child vs. 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 we wish to obtain a more general picture of the differences in children’s language experiences across these communities. Consider number of talkers present in the clip.

As shown above, there is systematic variation across our corpora in the number of present talkers: for example, Yélî Dnye-acquiring children are surrounded by substantially more talkers than children in the other corpora. There may be two ways of looking at Yélî children’s experience of TCDL: (1) All else being equal, Yélî Dnye-acquiring children hear significantly less TCDL 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 TCDL 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 TCDL as North American English-acquiring children, if not more.

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

In this analysis, therefore, we replicate our models of TCDL, CDL, and ADL, 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 corpus (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 + corpus + (1 | child\_id), ziformula = ~ child.age + corpus).

## 4.1 Target-child-directed language (TCDL)

The count model of the simpler regression of TCDL (N = 915, log-likelihood = -2,031.92, overdispersion estimate = 6.67) showed a marginal effect of child age (*B* = 0.16, *SE* = 0.10, *z* = 1.69, *p* = 0.09) and no significant effects of corpus in the count model (all *p* > .34). The zero-inflation component similarly showed no significant effects of age or corpus (all *p* > .15).

## 4.2 All child-directed language (CDL)

The count model of the simpler regression of CDL (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 corpus: CDL 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 corpus showed significant difference in the rate of CDL compared to North American English (all *p* > 0.2). However, in this case the zero-inflation component showed that both Tseltal and Yélî Dnye were significantly less likely than North American English to have clips with zero CDL (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 show that zero-CDL 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 CDL, the rate of CDL is significantly higher for Yélî Dnye-acquiring children than North American English-acquiring children. The zero-inflation component demonstrated no effect of child age.

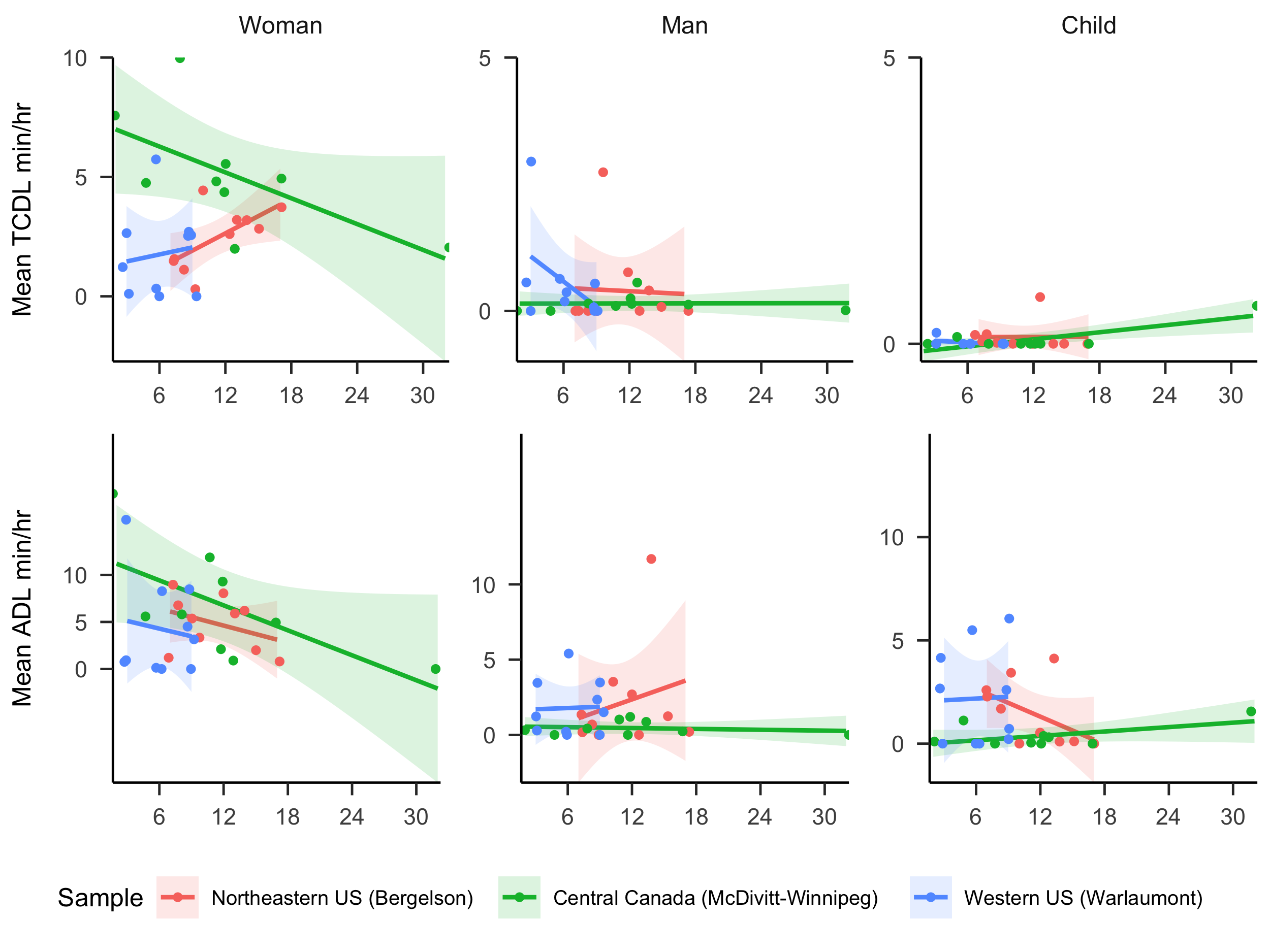
## 4.3 Adult-directed language (ADL)

The count model of the simpler regression of ADL (N = 915, log-likelihood = -2,517.12, overdispersion estimate = 14.51) showed significant effects of both child age and corpus. ADL decreased with child age (*B* = -0.23, *SE* = 0.11, *z* = -2.16, *p* = 0.03). ADL 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 corpus showed significant difference in the rate of ADL compared to North American English (all *p* > 0.19). As with CDL, the zero-inflation model component revealed further structure in the data: zero-ADL clips were significantly less likely in Yélî Dnye data and marginally less likely in Tseltal 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-ADL clips were significantly more likely for North American English-acquiring children than Yélî Dnye-acquiring children (and marginally for Tseltal-acquiring children) and that, for clips with non-zero amounts of ADL, the rate of ADL is significantly higher for Yélî Dnye-acquiring children than North American English-acquiring children. Consistent with the other models, there is no evidence of an age effect in the zero-inflation model component.

Pulling these results together with those reported in the main text (TCDL, ADL) and above (CDL), two primary points are worth highlighting. First, Yélî Dnye looks very different when number of talkers is removed from the model—it has equivalent overall rates of TCDL, higher rates of CDL and ADL, and is less likely to have a zero-CDL or zero-ADL 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 corpora included here. Second, we lose sight of the critical and cross-corpus 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. We here present the primary descriptive figure from the main text, but with data broken out by individual corpus and not by language variety.



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

# 6 Full model output for TCDL, CDL, and ADL

The full zero-inflated negative binomial mixed-effects regression output tables for TCDL, CDL, and ADL rate are presented below.

Table 2:

*Full output of the zero-inflated negative binomial mixed-effects regression of TCDL min/hr*

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

Table 3:

*Full output of the zero-inflated negative binomial mixed-effects regression of CDL min/hr*

| Model component | Term | Estimate (B) | Std. Error | t-value | p-value |
| --- | --- | --- | --- | --- | --- |
| count | (Intercept) | 1.37 | 0.086 | 16.008 | < 0.001 |
| count | tchiyr.std | 0.00 | 0.057 | -0.015 | 0.988 |
| count | SpkrTypeMan | -1.77 | 0.147 | -12.063 | < 0.001 |
| count | SpkrTypeChild | -2.96 | 0.24 | -12.309 | < 0.001 |
| count | nsk.std | 0.48 | 0.031 | 15.821 | < 0.001 |
| count | group\_corpNEUK\_English | 0.01 | 0.165 | 0.039 | 0.969 |
| count | group\_corpNEArg\_Spanish | 0.13 | 0.15 | 0.862 | 0.389 |
| count | group\_corpNETseltal | 0.15 | 0.159 | 0.954 | 0.34 |
| count | group\_corpNEYeli\_Dnye | -0.69 | 0.178 | -3.855 | < 0.001 |
| count | SpkrTypeMan:group\_corpNEUK\_English | 0.52 | 0.278 | 1.872 | 0.061 |
| count | SpkrTypeChild:group\_corpNEUK\_English | 0.59 | 0.399 | 1.48 | 0.139 |
| count | SpkrTypeMan:group\_corpNEArg\_Spanish | 0.58 | 0.238 | 2.454 | 0.014 |
| count | SpkrTypeChild:group\_corpNEArg\_Spanish | 1.21 | 0.318 | 3.809 | < 0.001 |
| count | SpkrTypeMan:group\_corpNETseltal | -0.15 | 0.296 | -0.494 | 0.621 |
| count | SpkrTypeChild:group\_corpNETseltal | 2.12 | 0.309 | 6.885 | < 0.001 |
| count | SpkrTypeMan:group\_corpNEYeli\_Dnye | 0.38 | 0.281 | 1.368 | 0.171 |
| count | SpkrTypeChild:group\_corpNEYeli\_Dnye | 3.71 | 0.29 | 12.782 | < 0.001 |
| count | tchiyr.std:SpkrTypeMan | -0.18 | 0.102 | -1.792 | 0.073 |
| count | tchiyr.std:SpkrTypeChild | 0.05 | 0.077 | 0.677 | 0.498 |
| count | sd\_\_(Intercept) | 0.23 |  |  |  |
| zero-inflation | (Intercept) | -3.91 | 0.846 | -4.629 | < 0.001 |
| zero-inflation | tchiyr.std | 0.17 | 0.505 | 0.342 | 0.733 |

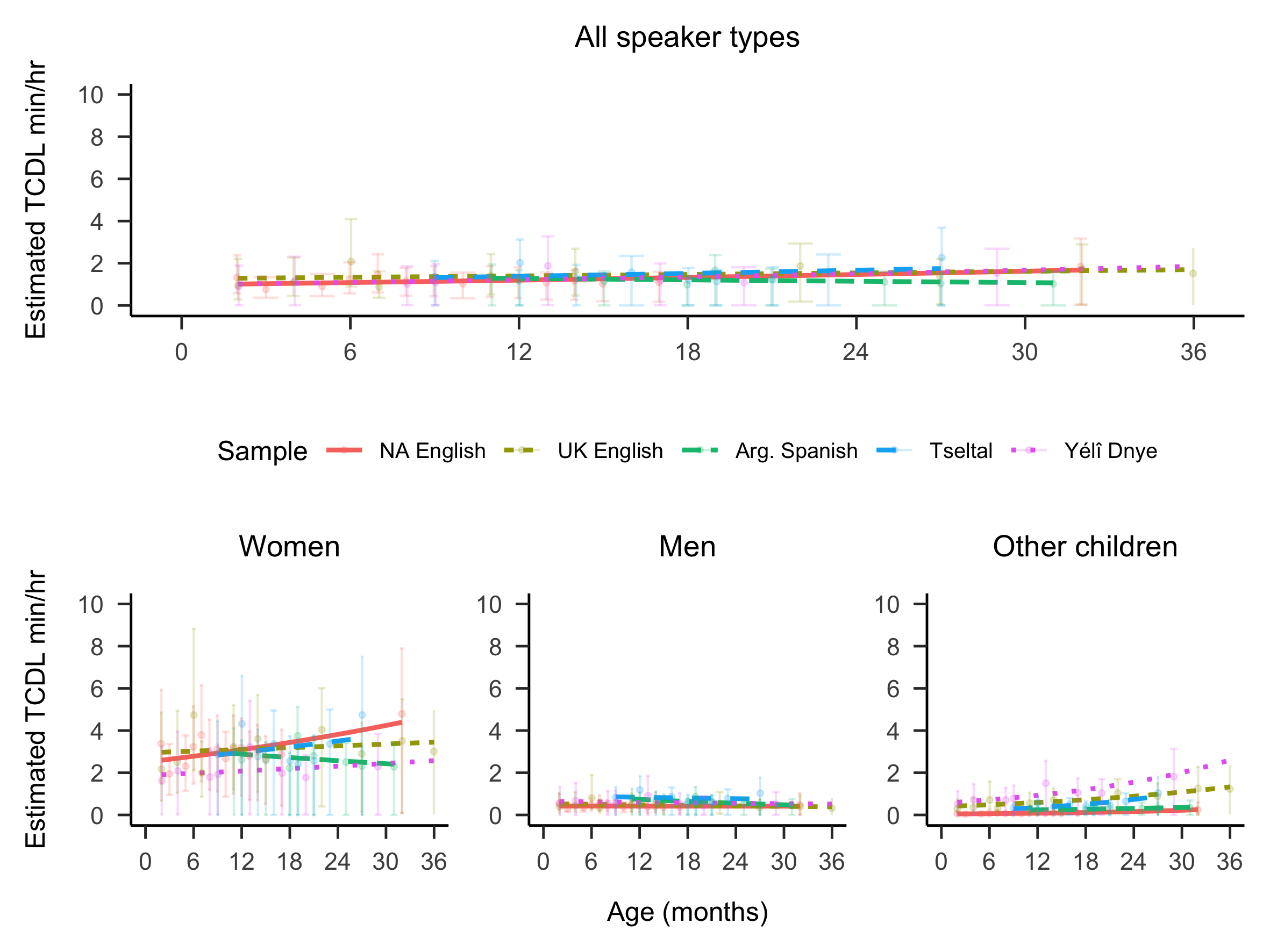
Table 4:

*Full output of the zero-inflated negative binomial mixed-effects regression of ADL min/hr*

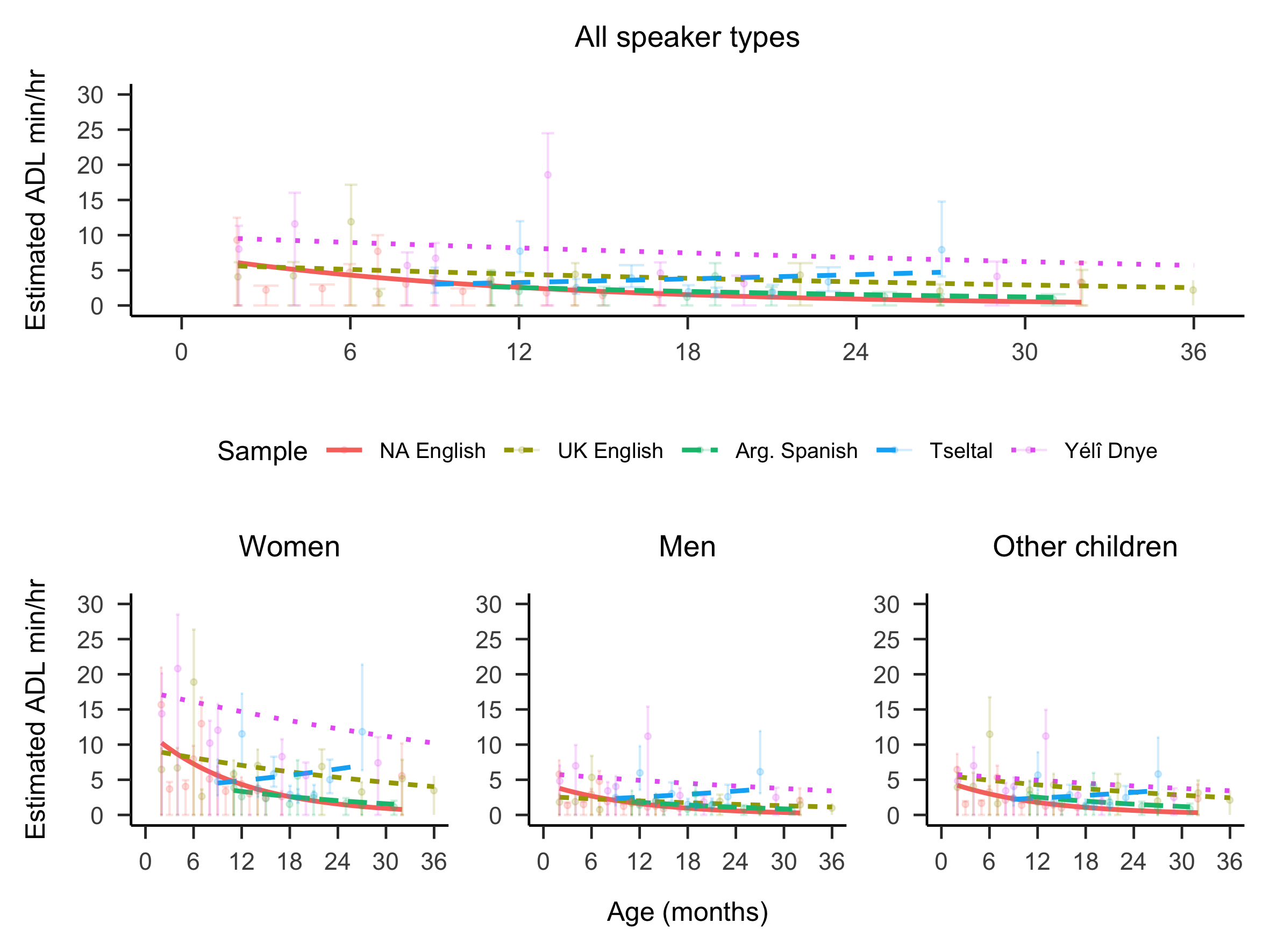
| Model component | Term | Estimate (B) | Std. Error | t-value | p-value |
| --- | --- | --- | --- | --- | --- |
| count | (Intercept) | 1.16 | 0.122 | 9.487 | < 0.001 |
| count | tchiyr.std | -0.32 | 0.128 | -2.45 | 0.014 |
| count | SpkrTypeMan | -1.00 | 0.128 | -7.77 | < 0.001 |
| count | SpkrTypeChild | -0.89 | 0.124 | -7.163 | < 0.001 |
| count | nsk.std | 0.71 | 0.033 | 21.538 | < 0.001 |
| count | group\_corpNEUK\_English | -0.18 | 0.267 | -0.693 | 0.488 |
| count | group\_corpNEArg\_Spanish | 0.27 | 0.222 | 1.225 | 0.221 |
| count | group\_corpNETseltal | 0.18 | 0.214 | 0.86 | 0.39 |
| count | group\_corpNEYeli\_Dnye | -0.03 | 0.206 | -0.132 | 0.895 |
| count | SpkrTypeMan:group\_corpNEUK\_English | 0.38 | 0.281 | 1.355 | 0.176 |
| count | SpkrTypeChild:group\_corpNEUK\_English | 0.60 | 0.261 | 2.312 | 0.021 |
| count | SpkrTypeMan:group\_corpNEArg\_Spanish | 0.34 | 0.229 | 1.481 | 0.139 |
| count | SpkrTypeChild:group\_corpNEArg\_Spanish | 0.18 | 0.217 | 0.82 | 0.412 |
| count | SpkrTypeMan:group\_corpNETseltal | -0.27 | 0.278 | -0.965 | 0.335 |
| count | SpkrTypeChild:group\_corpNETseltal | 0.39 | 0.234 | 1.669 | 0.095 |
| count | SpkrTypeMan:group\_corpNEYeli\_Dnye | -0.10 | 0.214 | -0.446 | 0.655 |
| count | SpkrTypeChild:group\_corpNEYeli\_Dnye | -0.20 | 0.205 | -0.985 | 0.324 |
| count | tchiyr.std:group\_corpNEUK\_English | 0.02 | 0.256 | 0.059 | 0.953 |
| count | tchiyr.std:group\_corpNEArg\_Spanish | 0.10 | 0.251 | 0.39 | 0.696 |
| count | tchiyr.std:group\_corpNETseltal | 0.27 | 0.169 | 1.591 | 0.112 |
| count | tchiyr.std:group\_corpNEYeli\_Dnye | -0.11 | 0.167 | -0.636 | 0.525 |
| count | sd\_\_(Intercept) | 0.35 |  |  |  |
| zero-inflation | (Intercept) | -16.40 | 1613.376 | -0.01 | 0.992 |
| zero-inflation | group\_corpNEUK\_English | -8.36 | 79972.675 | 0 | 1 |
| zero-inflation | group\_corpNEArg\_Spanish | 14.35 | 1613.377 | 0.009 | 0.993 |
| zero-inflation | group\_corpNETseltal | -3.65 | 10742.64 | 0 | 1 |
| zero-inflation | group\_corpNEYeli\_Dnye | -1.92 | 5214.752 | 0 | 1 |

# 7 Marginal means estimates of TCDL and ADL

In the main manuscript the plotted data of TCDL and ADL across corpora and linguistic community are shown based on the actual data. However, our analyses take into account a number of factors that are not visible in that main-text plot. Here we show marginal-means plots from the models of ADL and TCDL, displaying estimated values of each over child age, and talker type, for each corpus.



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



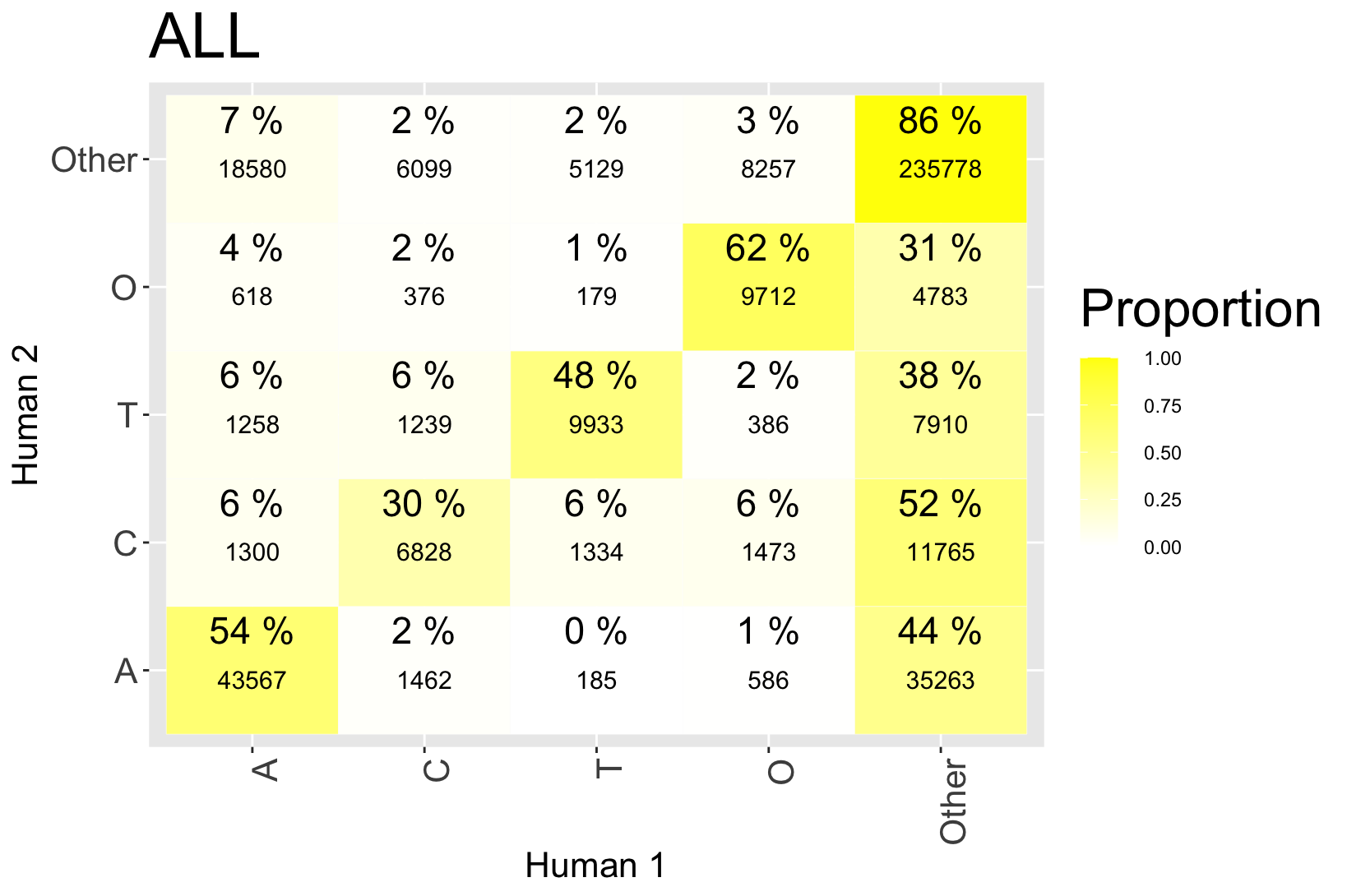
*Figure* *7.*  Model estimated rates of ADL min/hr by target child age and corpus, shown all together in the top panel and split by talker type in the bottom panel. Each datapoint represents the mean from one recording.

# 8 Confusion among addressee annotation types

The agreement scores for addressee ranged from slight to substantial across corpora. We lay out here further information about how these agreement scores are derived and distributed in this continuous, naturalistic dataset.

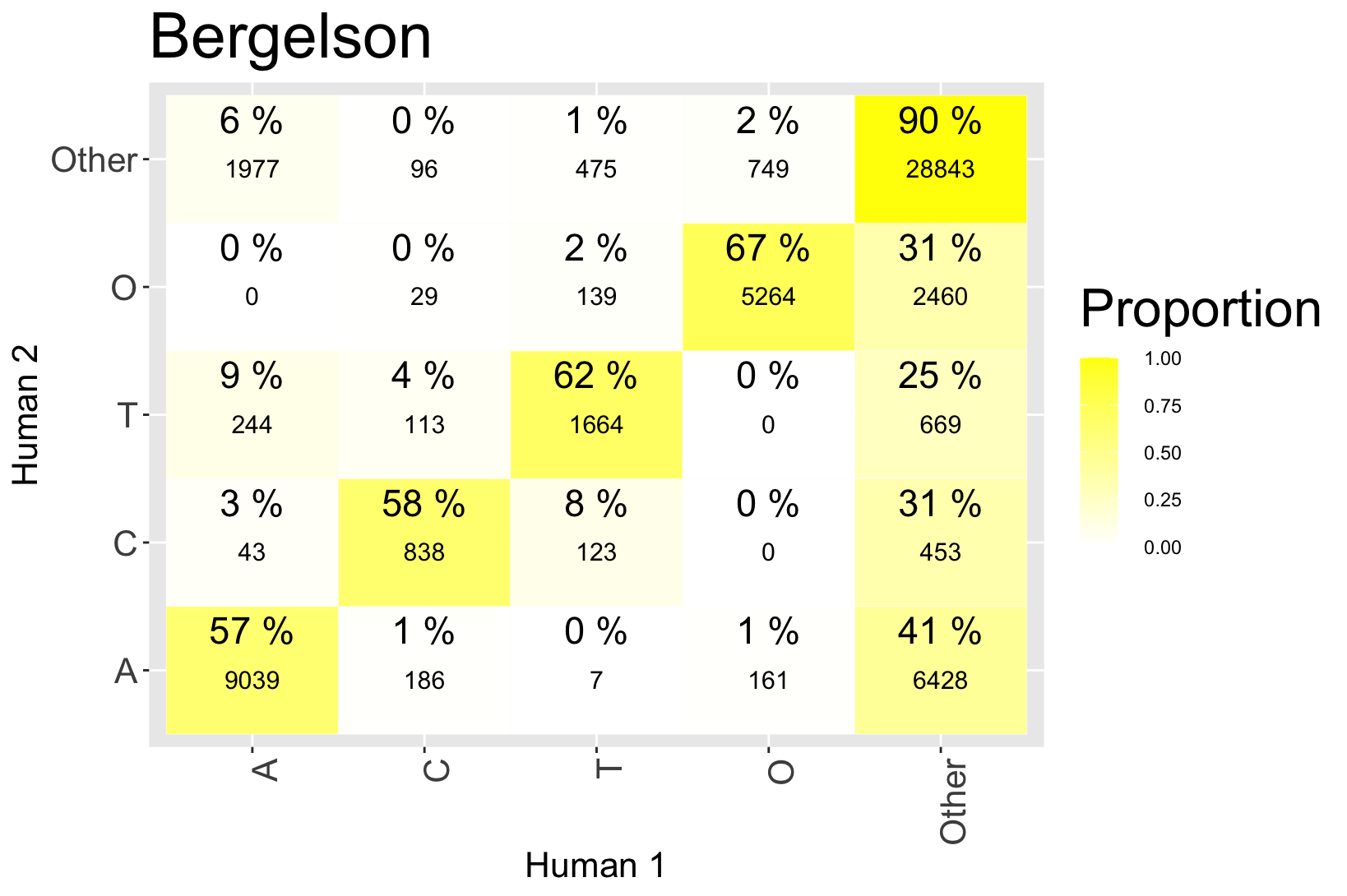
We show below a confusion matrix displaying the first vs. second annotator’s impressions of addressee across all corpora. To deal with the fact that we have continuous data (not, e.g., individual clips that are multiply coded, as in Bergelson et al. (2019) or Weisleder and Fernald (2013)) we compute 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 much so. Most cases of disagreement are really only with the cases of silence or overlapping speech—a kind of error inherited from the process of speech diarization.

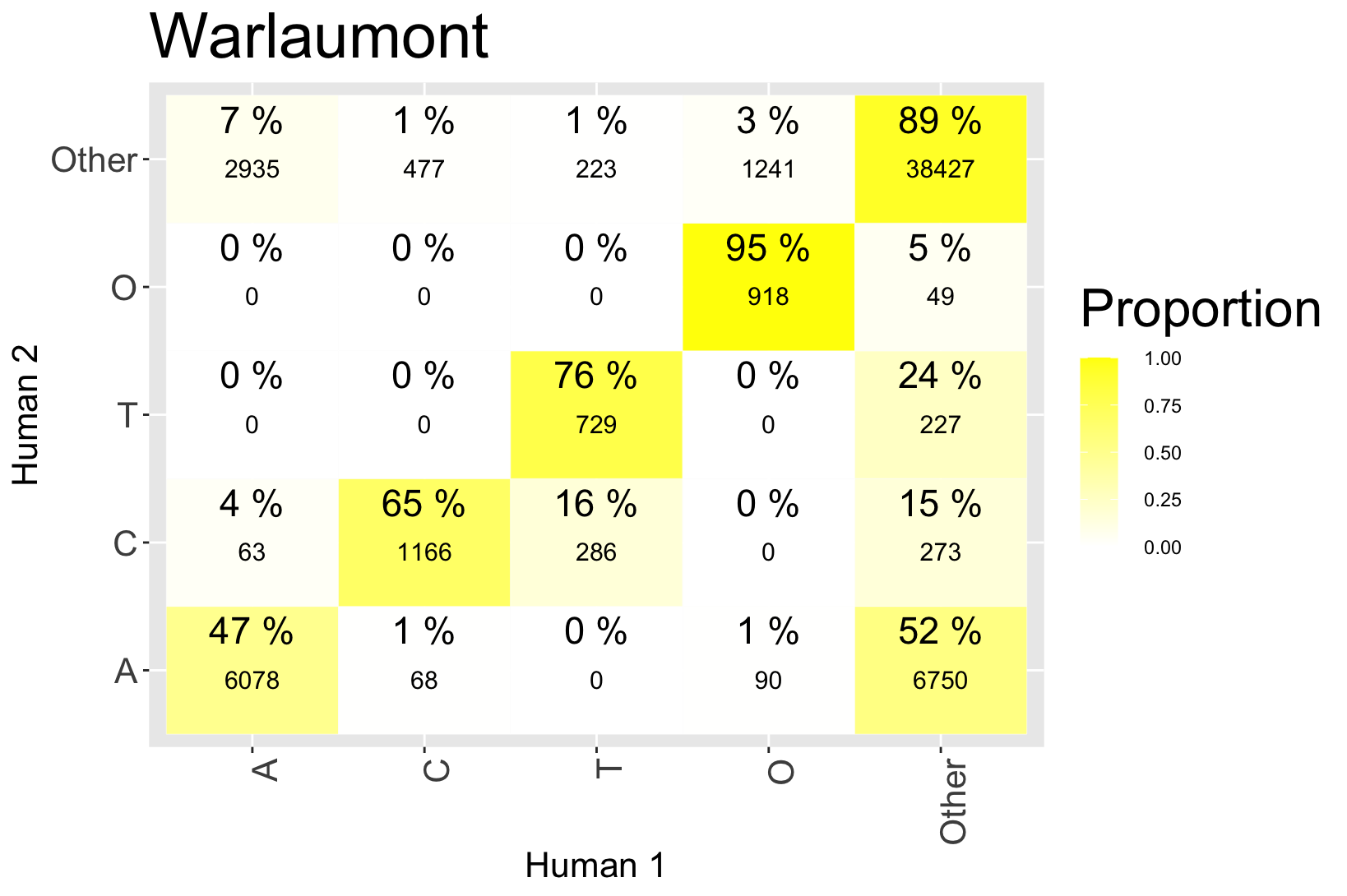


*Figure* *8.*  Confusion matrix across all corpora for addressee annotations between the two human annotators.

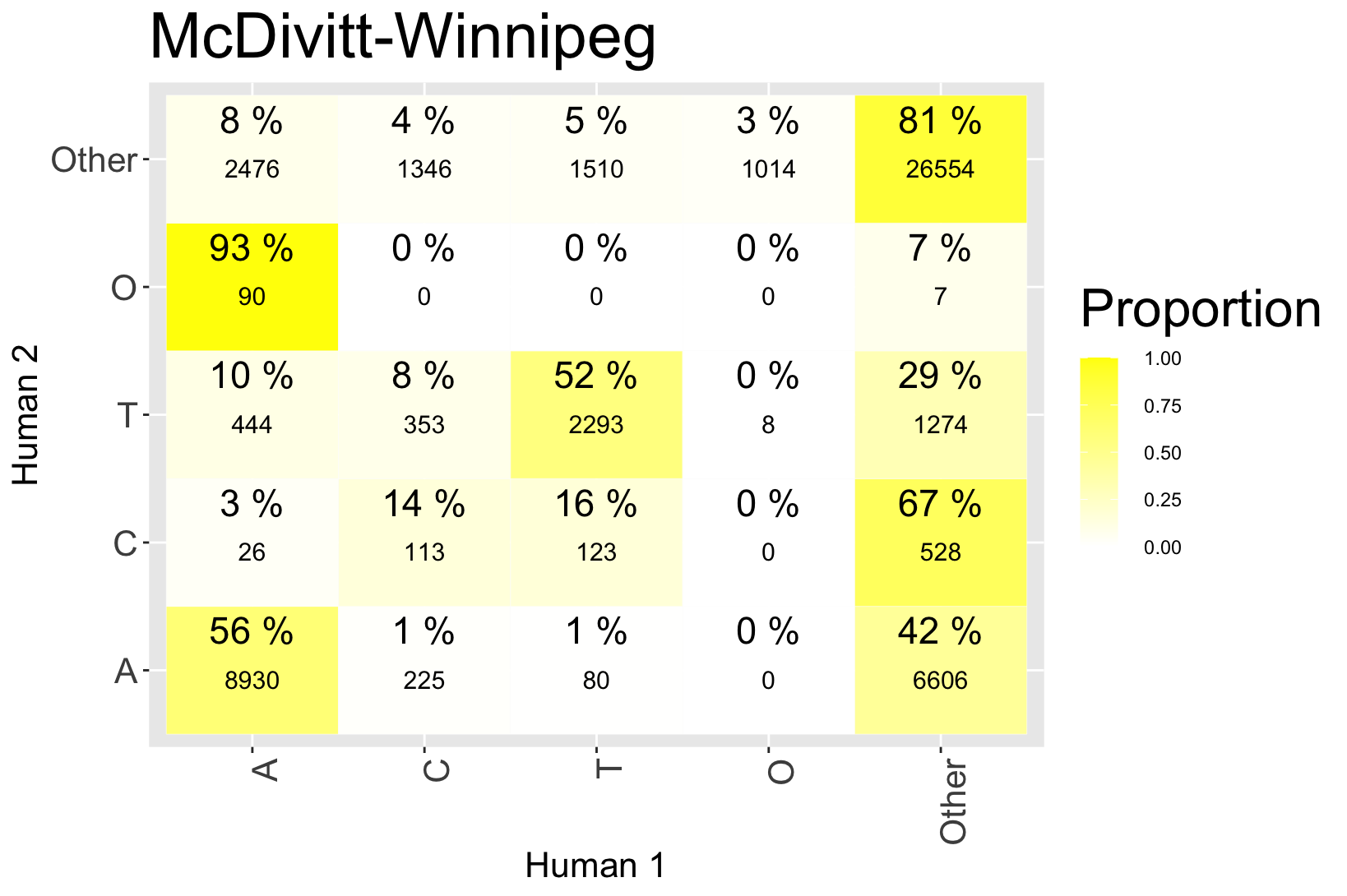
However, this is the sum over *all* of the data that we have. We had anticipated that reliability would be substantially worse in the Argentinian Spanish, Tseltal, and Yélî Dnye data because the reliability annotators did not actually fluently speak (if they had any knowledge at all) of the data they were annotating. 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 decision. This pattern is visible in the same matrix, but now created for each individual corpus. We take this as indirect evidence in favor of the only *partial* recoverability of register and addressee-specific features across diverse languages (see also Soderstrom et al., 2021).



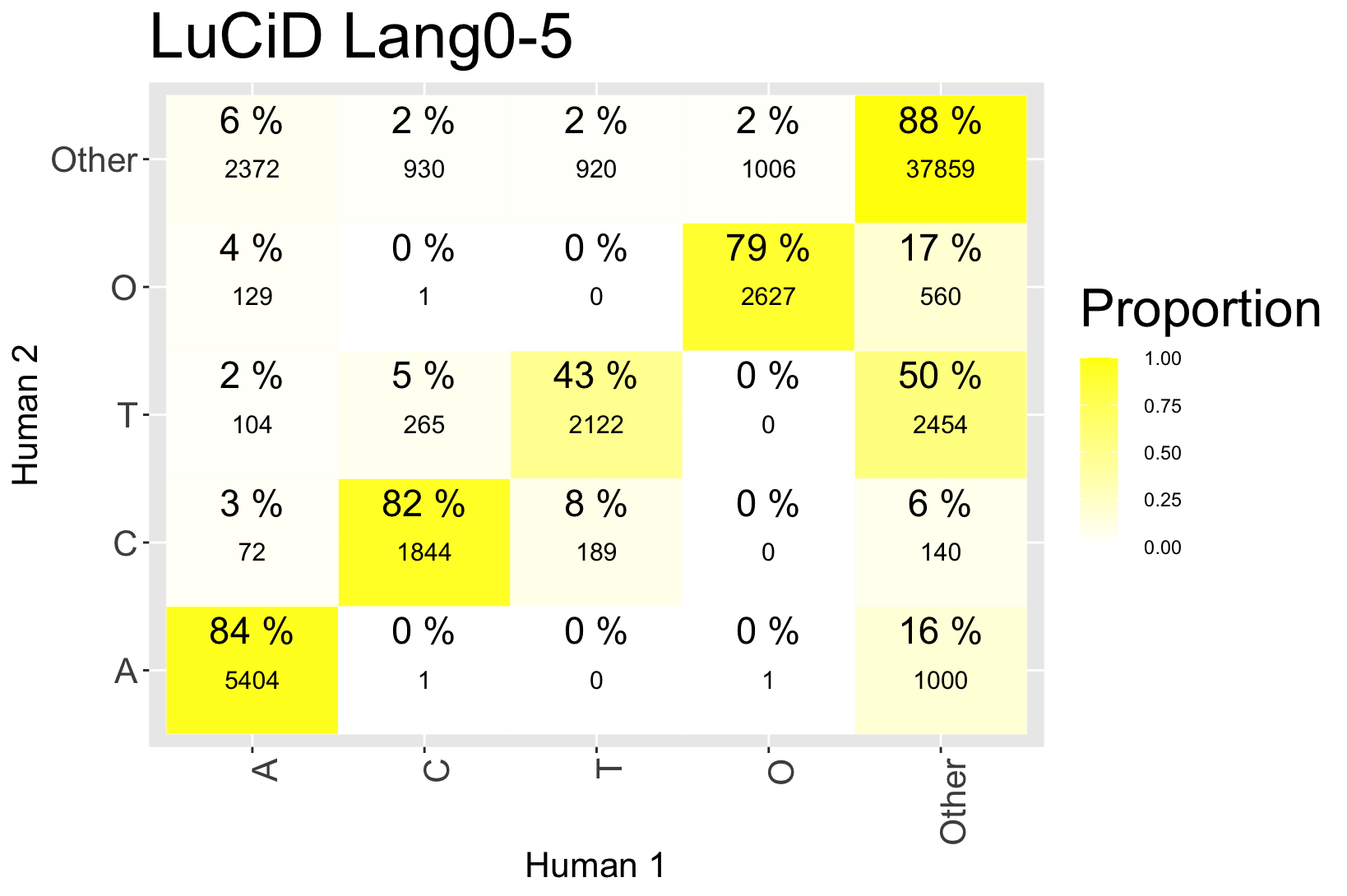
*Figure* *9.*  Confusion matrix for the Bergelson (US English) corpus for addressee annotations between the two human annotators.



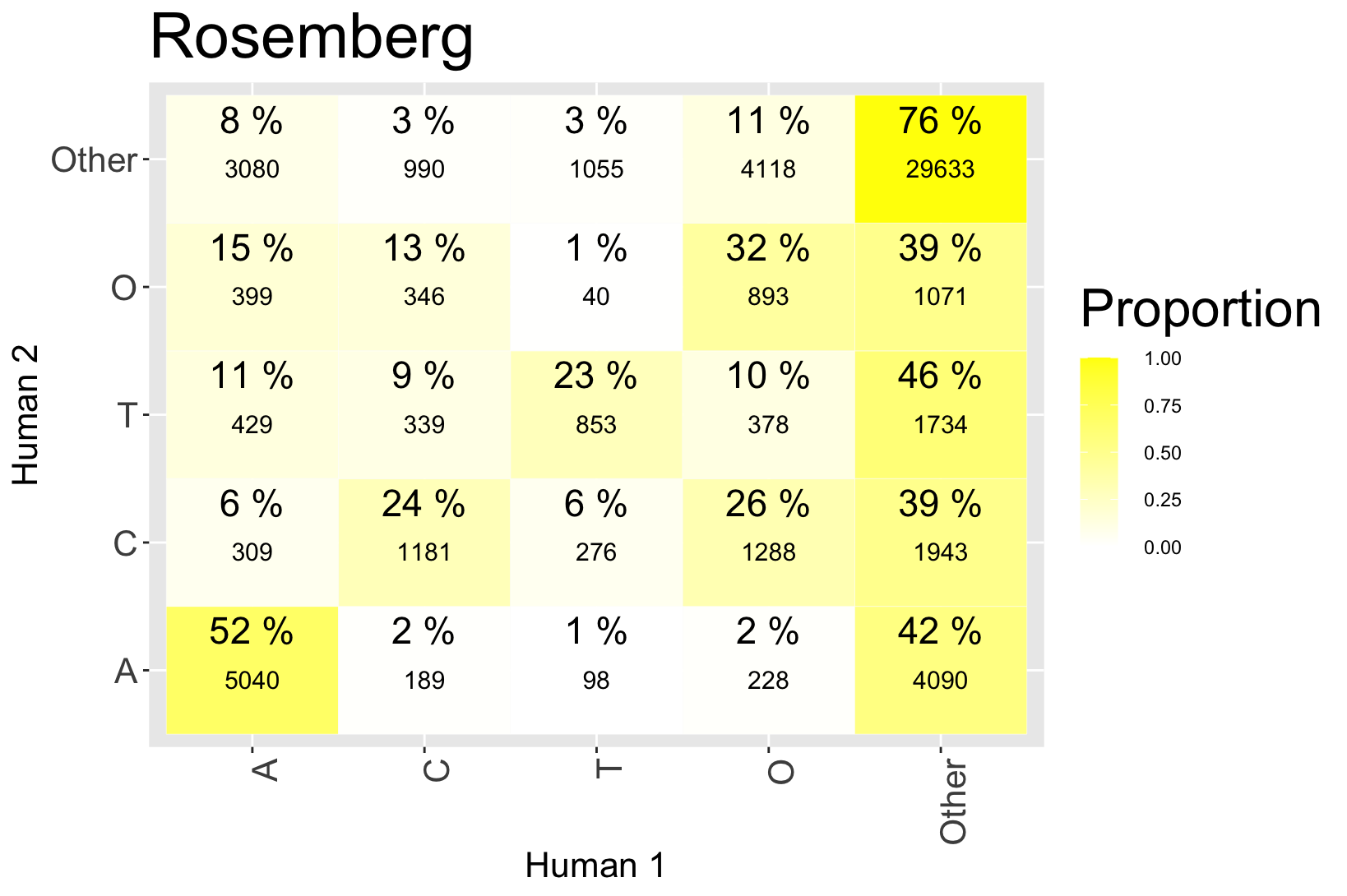
*Figure* *10.*  Confusion matrix for the Warlaumont (US English) corpus for addressee annotations between the two human annotators.



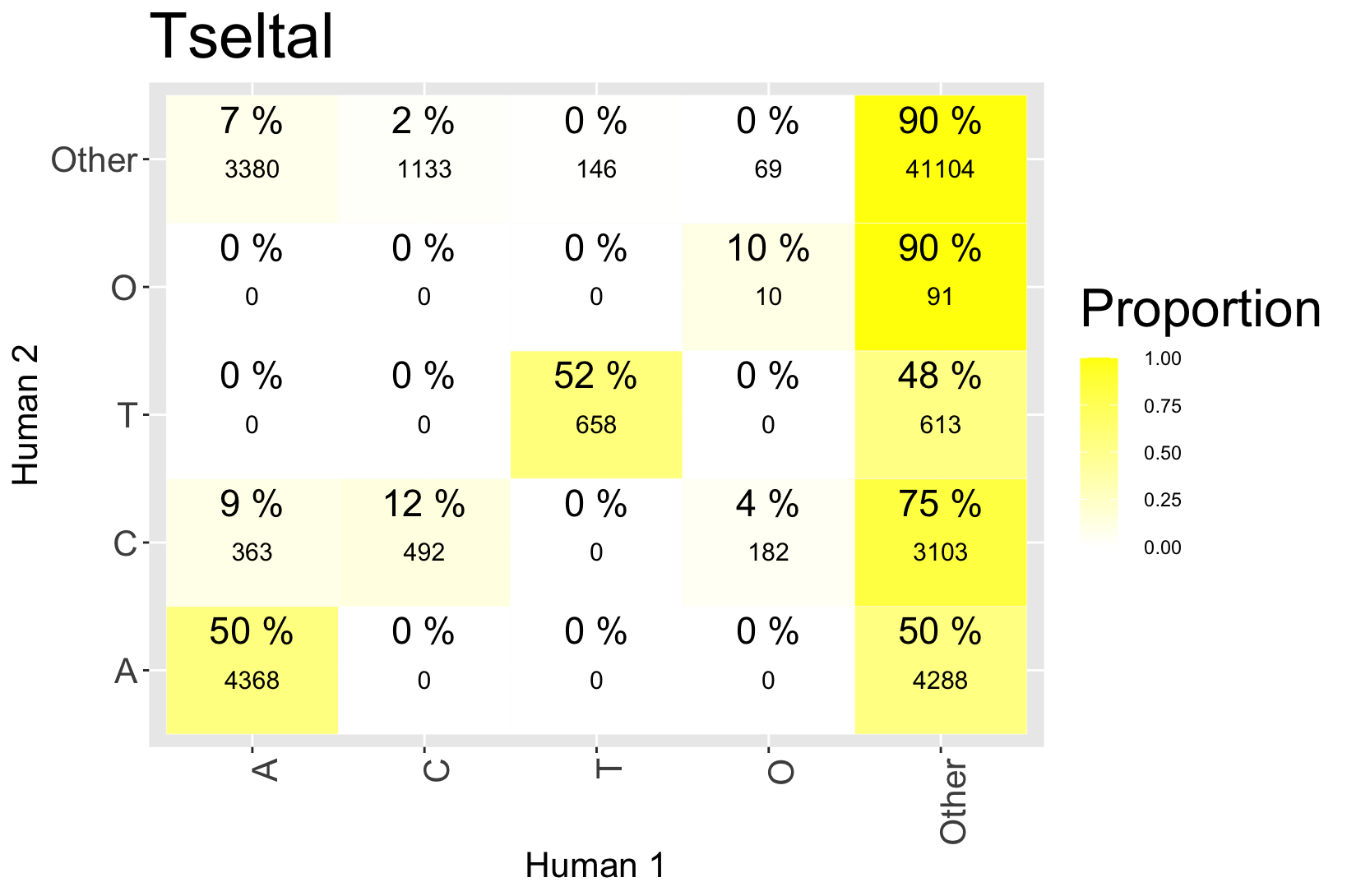
*Figure* *11.*  Confusion matrix for the McDivitt/Winnipeg (Canadian English) corpus for addressee annotations between the two human annotators.



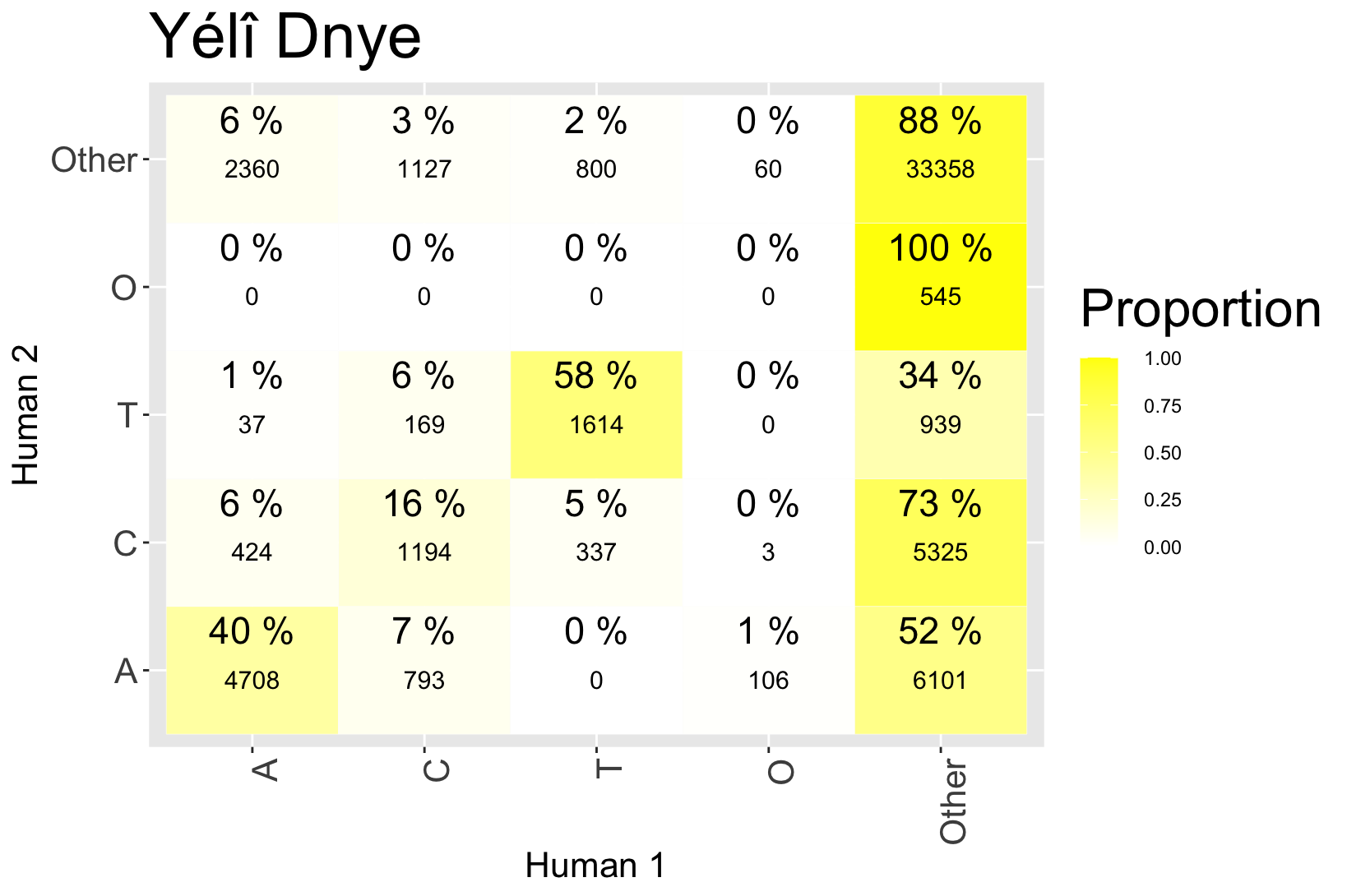
*Figure* *12.*  Confusion matrix for the LuCiD Lang0–5 (UK English) corpus for addressee annotations between the two human annotators.



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



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



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

# 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>

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*.

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>