Supplementary Materials: Early language experience in a Papuan village

Marisa Casillas[1], Penelope Brown[1], & Stephen C. Levinson[1]

[1] Max Planck Institute for Psycholinguistics

Author Note

Correspondence concerning this article should be addressed to Marisa Casillas, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: Marisa.Casillas@mpi.nl

7 Abstract

8

Supplementary Materials: Early language experience in a Papuan village

Full model outputs

In these Supplementary Materials we give the full model output tables for each analysis in the main text, including re-leveled versions of each model to show all three of the two-way contrasts between the three-level time-of-day factor (i.e., morning vs. midday, morning vs. afternoon, and midday vs. afternoon) as well as, for each of the measures, a histogram showing how each variable is distributed (i.e., because they are non-normal and/or zero-inflated) and a figure showing the distribution of model residuals. For every negative binomial model, we also include the full model output table and residual plots for matching gaussian mixed-effects regressions which uses a log-transformed dependent measure. Such gaussian models with log-transformed measures are an alternative solution to analyzing non-normal distributions sometimes used in psycholinguistics, but are not suitable for the current data given how our speech environment measures are distributed, particularly in the randomly sampled clips (see, e.g., Figures 1, 7, 10, 13, 19). Overall, the gaussian models show a qualitatively similar pattern of results. These analyses are structured as identically as possible to those in Casillas and colleagues' (forthcoming) study on Tseltal Mayan child language environments.

How to interpret the model output

All models were run with the glmm-TMB library in R (Brooks et al., 2017a, 2017b). Note that, in the negative binomial regressions, the dependent variables have been rounded to the nearest integer (e.g., 3.2 minutes of TCDS per hour becomes 3 minutes per hour in the model).

The predictors in the models are abbreviated as follows: tchiyr.std = centered,

standardized target child age in months; stthr.tri = the start time of the clip as either morning, midday, or afternoon; hsz.std = centered, standardized household size of the target child; nsk.std = centered, standardized number of speakers present in the clip, aclew_child_id = the unique identifier for each child. The predictors are sometimes combined in two-way interactions, as shown below with a ":" separator between predictor names (e.g., tchiyr.std:nsk.std = a two-way interaction of target child age and number of speakers present).

In each model output table, the "component" shows what kind of model the estimate derives from (e.g., the zero-inflated models include both a conditional "cond" set of predictors, random effects, and zero-inflation "zi" predictors). The "term" is the estimated predictor. The "statistic" is the estimated z-statistic for each predictor's effect. The other labels are self-explanatory.

As more data are added to this corpus, the analyses will also be updated, as will this supplementary model information, all of which will be available online at: https://middycasillas.shinyapps.io/Rossel_Child_Language_Environment/.

Target-child-directed speech (TCDS)

Random clips. TCDS rate in the random clips demonstrated a skewed distribution with extra cases of zero Figure 1. We therefore modeled it using a zero-inflated negative binomial mixed-effects regression in the main text: results for the two models demonstrating all pairwise effects of time of day are shown in Table 1 and Table 2. The residuals for the default model (Table 1) are shown in Figure 2.

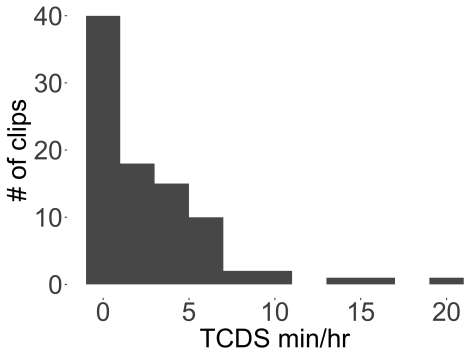

Figure 1. The distribution of TCDS rates found across the 90 random clips.

Table 1

Full output of the zero-inflated negative binomial mixed-effects regression of TCDS min/hr for the random sample, with midday as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	0.69	0.32	2.16	0.03
cond	tchiyr.std	0.73	0.23	3.20	0.00
cond	stthr.trimorning	0.80	0.36	2.23	0.03
cond	stthr.triafternoon	0.26	0.35	0.73	0.46
cond	hsz.std	-0.21	0.12	-1.69	0.09
cond	nsk.std	-0.04	0.16	-0.27	0.79
cond	tchiyr.std:stthr.trimorning	-0.59	0.30	-1.94	0.05
cond	tchiyr.std:stthr.triafternoon	-0.60	0.29	-2.04	0.04
cond	tchiyr.std:nsk.std	-0.03	0.11	-0.26	0.80
zi	(Intercept)	-9.28	11.51	-0.81	0.42
zi	nsk.std	-5.66	7.44	-0.76	0.45
random_effect	aclew_child_id	0.00	NA	NA	NA

Table 2

Model output of the zero-inflated negative binomial mixed-effects regression of TCDS min/hr for the random sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	0.95	0.19	4.99	0.00
cond	tchiyr.std	0.14	0.19	0.72	0.47
cond	stthr.tri.amidday	-0.26	0.35	-0.73	0.46
cond	stthr.tri.amorning	0.54	0.26	2.10	0.04
cond	hsz.std	-0.21	0.12	-1.69	0.09
cond	nsk.std	-0.04	0.16	-0.27	0.79
cond	tchiyr.std:stthr.tri.amidday	0.60	0.29	2.04	0.04
cond	tchiyr.std:stthr.tri.amorning	0.01	0.27	0.03	0.98
cond	tchiyr.std:nsk.std	-0.03	0.11	-0.26	0.80
zi	(Intercept)	-9.28	11.51	-0.81	0.42
zi	nsk.std	-5.66	7.44	-0.76	0.45
random_effect	aclew_child_id	0.00	NA	NA	NA

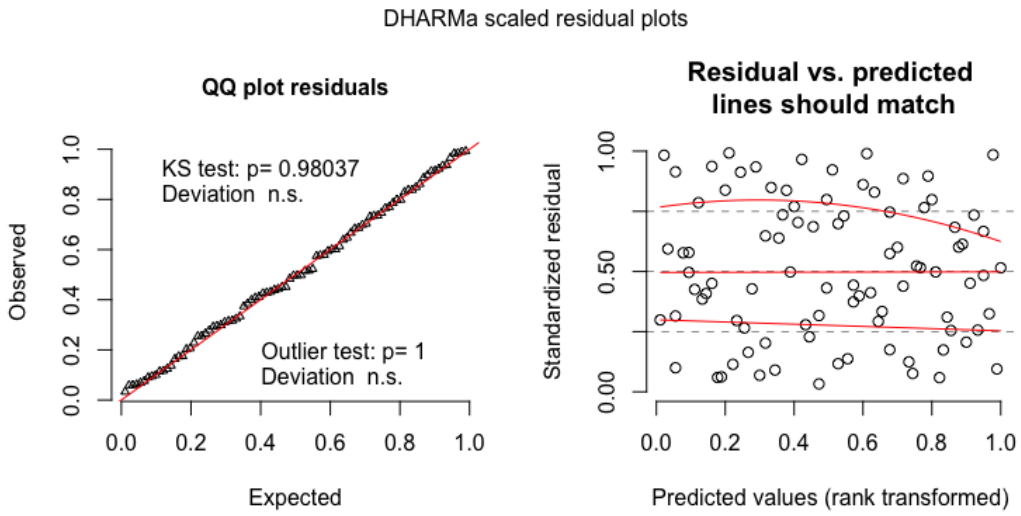

Figure 2. The model residuals from the zero-inflated negative binomial mixed-effects regression of TCDS min/hr for the random sample.

As an alternative analysis we generated parallel models of TCDS rate in the random clips using gaussian mixed-effects regression with logged values of TCDS: results for the two models demonstrating all pairwise effects of time of day are shown in Table 3 and Table 4. The residuals for the default gaussian model (Table 3) are shown in Figure 3.

Table 3

Full output of the gaussian mixed-effects regression of TCDS min/hr for the random sample, with midday as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	0.89	0.18	5.04	0.00
cond	tchiyr.std	0.48	0.17	2.80	0.00
cond	stthr.trimorning	0.40	0.24	1.68	0.09
cond	stthr.triafternoon	0.09	0.21	0.42	0.67
cond	hsz.std	-0.11	0.09	-1.26	0.21
cond	nsk.std	0.03	0.09	0.35	0.73
cond	tchiyr.std:stthr.trimorning	-0.39	0.25	-1.56	0.12
cond	tchiyr.std:stthr.triafternoon	-0.41	0.22	-1.88	0.06
cond	tchiyr.std:nsk.std	-0.03	0.08	-0.33	0.74
random_effect	aclew_child_id	0.00	NA	NA	NA
random_effect	Residual	0.79	NA	NA	NA

Table 4

Model output of the gaussian mixed-effects regression of TCDS min/hr for the random sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	0.98	0.12	8.11	0.00
cond	tchiyr.std	0.08	0.13	0.58	0.56
cond	stthr.tri.amidday	-0.09	0.21	-0.42	0.67
cond	stthr.tri.amorning	0.31	0.20	1.56	0.12
cond	hsz.std	-0.11	0.09	-1.26	0.21
cond	nsk.std	0.03	0.09	0.35	0.73
cond	tchiyr.std:stthr.tri.amidday	0.41	0.22	1.88	0.06
cond	tchiyr.std:stthr.tri.amorning	0.02	0.22	0.10	0.92
cond	tchiyr.std:nsk.std	-0.03	0.08	-0.33	0.74
random_effect	aclew_child_id	0.00	NA	NA	NA
random_effect	Residual	0.79	NA	NA	NA

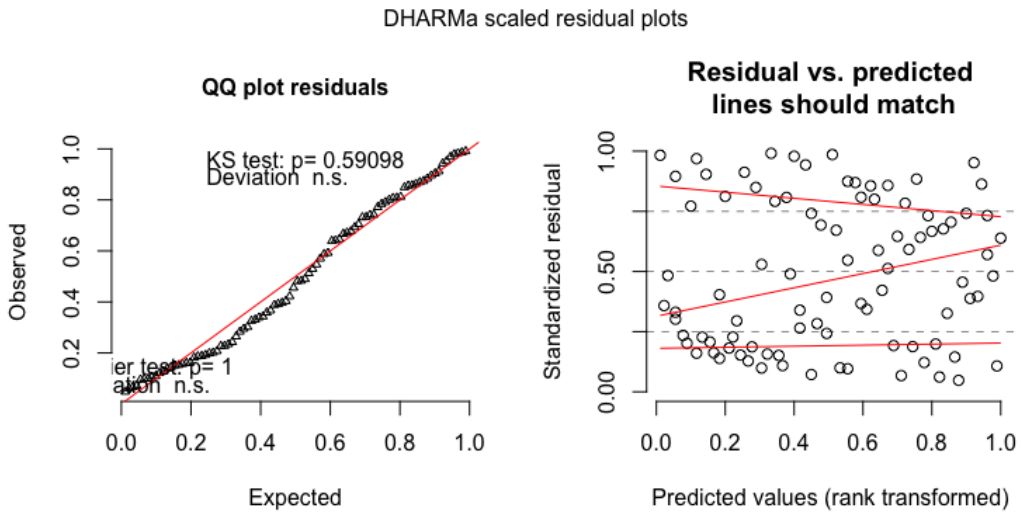

Figure 3. The model residuals from the gaussian mixed-effects regression of TCDS min/hr for the random sample.

Turn-taking clips. TCDS rate in the turn-taking clips demonstrated a slightly skewed, but unimodal distribution Figure 4. We therefore modeled it using a plain (i.e., non-zero-inflated) negative binomial mixed-effects regression in the main text: results for the two models demonstrating all pairwise effects of time of day are shown in Table 5 and Table 6. The residuals for the default model (Table 5) are shown in Figure 5.

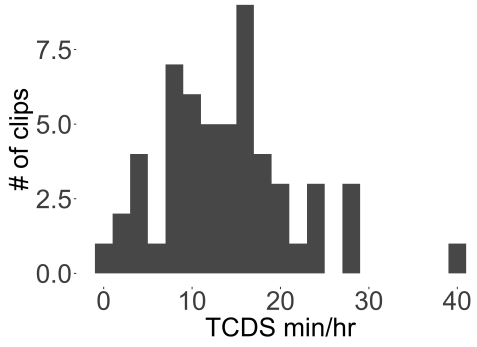

Figure 4. The distribution of TCDS rates found across the 59 turn-taking clips.

Table 5

Full output of the negative binomial mixed-effects regression of TCDS min/hr for the turn-taking sample, with midday as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.39	0.25	9.45	0.00
cond	tchiyr.std	-0.63	0.27	-2.33	0.02
cond	stthr.trimorning	0.22	0.28	0.77	0.44
cond	stthr.triafternoon	0.34	0.27	1.24	0.22
cond	hsz.std	-0.02	0.08	-0.26	0.79
cond	nsk.std	-0.04	0.09	-0.52	0.60
cond	tchiyr.std:stthr.trimorning	0.53	0.28	1.89	0.06
cond	tchiyr.std:stthr.triafternoon	0.60	0.28	2.17	0.03
cond	tchiyr.std:nsk.std	-0.15	0.11	-1.35	0.18
random_effect	aclew_child_id	0.00	NA	NA	NA

Table 6

Model output of the negative binomial mixed-effects regression of TCDS min/hr for the turn-taking sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.73	0.11	23.84	0.00
cond	tchiyr.std	-0.02	0.13	-0.18	0.86
cond	stthr.tri.amidday	-0.34	0.27	-1.24	0.22
cond	stthr.tri.amorning	-0.12	0.17	-0.69	0.49
cond	hsz.std	-0.02	0.08	-0.26	0.79
cond	nsk.std	-0.04	0.09	-0.52	0.60
cond	tchiyr.std:stthr.tri.amidday	-0.60	0.28	-2.17	0.03
cond	tchiyr.std:stthr.tri.amorning	-0.07	0.17	-0.42	0.68
cond	tchiyr.std:nsk.std	-0.15	0.11	-1.35	0.18
random_effect	aclew_child_id	0.00	NA	NA	NA

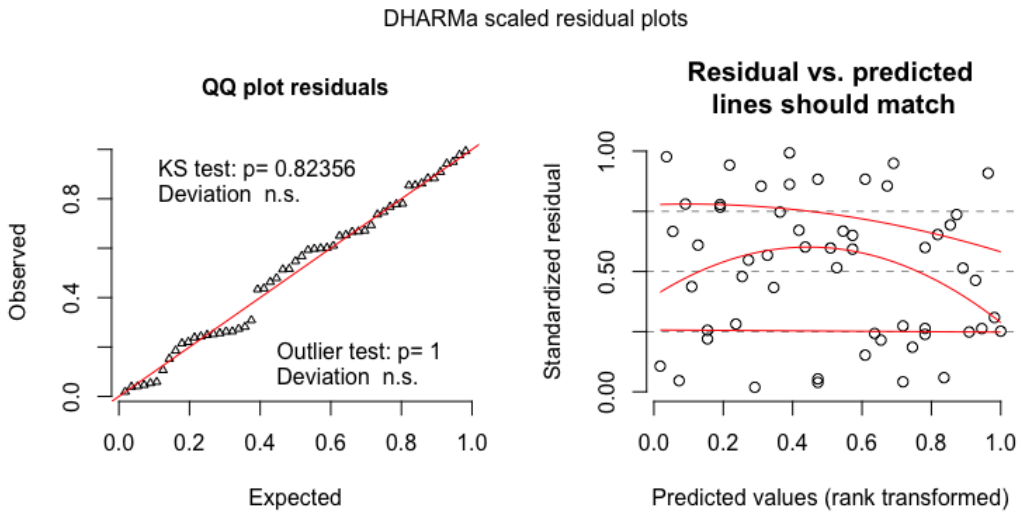

Figure 5. The model residuals from the negative binomial mixed-effects regression of TCDS min/hr for the turn-taking sample.

As an alternative analysis we generated parallel models of TCDS rate in the turn-taking clips using gaussian mixed-effects regression with logged values of TCDS: results for the two models demonstrating all pairwise effects of time of day are shown in Table 7 and Table 8. The residuals for the default gaussian model (Table 7) are shown in Figure 6.

Table 7

Full output of the gaussian mixed-effects regression of TCDS min/hr for the turn-taking sample, with midday as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.32	0.24	9.69	0.00
cond	tchiyr.std	-0.84	0.24	-3.44	0.00
cond	stthr.trimorning	0.21	0.29	0.72	0.47
cond	stthr.triafternoon	0.36	0.26	1.36	0.18
cond	hsz.std	-0.05	0.08	-0.57	0.57
cond	nsk.std	-0.05	0.09	-0.58	0.56
cond	tchiyr.std:stthr.trimorning	0.75	0.26	2.88	0.00
cond	tchiyr.std:stthr.triafternoon	0.81	0.26	3.14	0.00
cond	tchiyr.std:nsk.std	-0.18	0.12	-1.48	0.14
random_effect	aclew_child_id	0.09	NA	NA	NA
random_effect	Residual	0.53	NA	NA	NA

Table 8

Model output of the gaussian mixed-effects regression of TCDS min/hr for the turn-taking sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.68	0.13	20.54	0.00
cond	tchiyr.std	-0.03	0.16	-0.19	0.85
cond	stthr.tri.amidday	-0.36	0.26	-1.36	0.18
cond	stthr.tri.amorning	-0.15	0.21	-0.73	0.47
cond	hsz.std	-0.05	0.08	-0.57	0.57
cond	nsk.std	-0.05	0.09	-0.58	0.56
cond	tchiyr.std:stthr.tri.amidday	-0.81	0.26	-3.14	0.00
cond	tchiyr.std:stthr.tri.amorning	-0.07	0.20	-0.33	0.74
cond	tchiyr.std:nsk.std	-0.18	0.12	-1.48	0.14
random_effect	aclew_child_id	0.09	NA	NA	NA
random_effect	Residual	0.53	NA	NA	NA

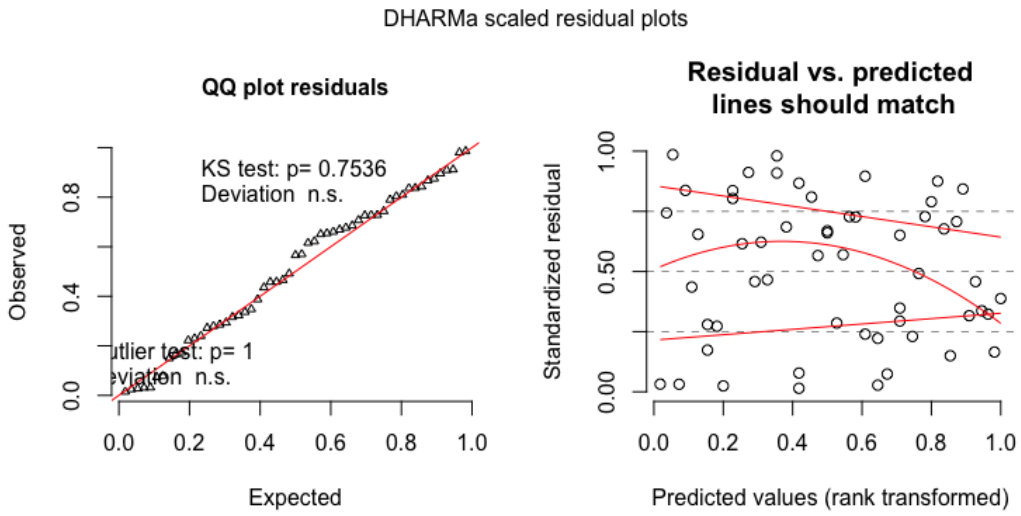

Figure 6. The model residuals from the gaussian mixed-effects regression of TCDS min/hr for the turn-taking sample.

Other-directed speech (ODS)

Random clips. ODS rate in the random clips demonstrated a skewed distribution, but without extra cases of zero Figure 7. We therefore modeled it using a negative binomial mixed-effects regression without zero inflation in the main text: results for the two models demonstrating all pairwise effects of time of day are shown in Table 9 and Table 10. The residuals for the default model (Table 9) are shown in Figure 8.

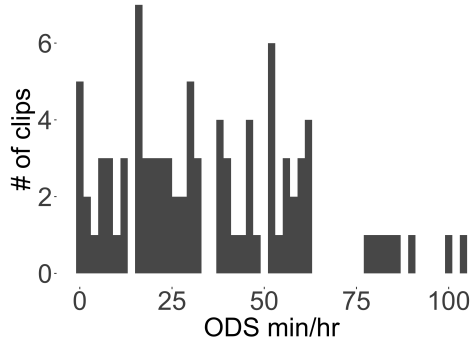

Figure 7. The distribution of ODS rates found across the 90 random clips.

Table 9

Full output of the negative binomial mixed-effects regression of ODS min/hr for the random sample, with midday as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	3.26	0.14	23.99	0.00
cond	tchiyr.std	-0.57	0.17	-3.28	0.00
cond	stthr.trimorning	0.20	0.16	1.19	0.23
cond	stthr.triafternoon	0.26	0.15	1.68	0.09
cond	hsz.std	-0.02	0.06	-0.32	0.75
cond	nsk.std	0.50	0.05	10.07	0.00
cond	tchiyr.std:stthr.trimorning	0.65	0.20	3.23	0.00
cond	tchiyr.std:stthr.triafternoon	0.28	0.20	1.43	0.15
cond	tchiyr.std:nsk.std	0.04	0.05	0.87	0.38
random_effect	aclew_child_id	0.00	NA	NA	NA

Table 10

Model output of the negative binomial mixed-effects regression of ODS min/hr for the random sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	3.51	0.08	42.78	0.00
cond	tchiyr.std	-0.29	0.09	-3.12	0.00
cond	stthr.tri.amidday	-0.26	0.15	-1.68	0.09
cond	stthr.tri.amorning	-0.06	0.13	-0.48	0.63
cond	hsz.std	-0.02	0.06	-0.32	0.75
cond	nsk.std	0.50	0.05	10.07	0.00
cond	tchiyr.std:stthr.tri.amidday	-0.28	0.20	-1.43	0.15
cond	tchiyr.std:stthr.tri.amorning	0.37	0.15	2.50	0.01
cond	tchiyr.std:nsk.std	0.04	0.05	0.87	0.38
random_effect	aclew_child_id	0.00	NA	NA	NA

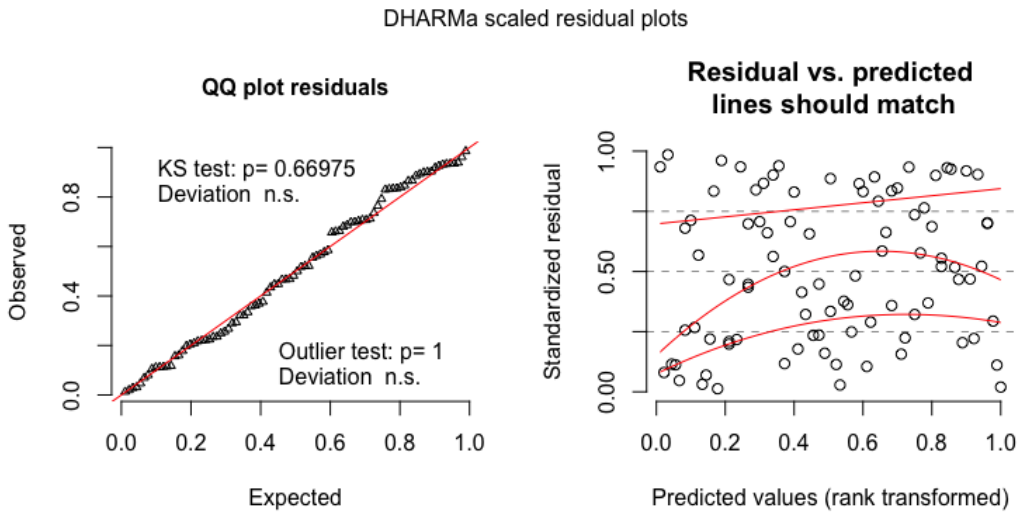

Figure 8. The model residuals from the zero-inflated negative binomial mixed-effects regression of ODS min/hr for the random sample.

As an alternative analysis we generated parallel models of ODS rate in the random clips using gaussian mixed-effects regression with logged values of ODS: results for the two models demonstrating all pairwise effects of time of day are shown in Table 11 and Table 12. The residuals for the default gaussian model (Table 11) are shown in Figure 9.

Table 11

Full output of the gaussian mixed-effects regression of ODS min/hr for the random sample, with midday as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	3.06	0.16	18.79	0.00
cond	tchiyr.std	-0.48	0.16	-2.98	0.00
cond	stthr.trimorning	0.26	0.20	1.25	0.21
cond	stthr.triafternoon	0.28	0.18	1.55	0.12
cond	hsz.std	0.00	0.10	0.03	0.98
cond	nsk.std	0.68	0.08	8.82	0.00
cond	tchiyr.std:stthr.trimorning	0.57	0.21	2.70	0.01
cond	tchiyr.std:stthr.triafternoon	0.09	0.18	0.51	0.61
cond	tchiyr.std:nsk.std	0.04	0.07	0.63	0.53
random_effect	aclew_child_id	0.20	NA	NA	NA
random_effect	Residual	0.66	NA	NA	NA

Table 12

Model output of the gaussian mixed-effects regression of ODS min/hr for the random sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	3.34	0.12	28.26	0.00
cond	tchiyr.std	-0.38	0.13	-3.04	0.00
cond	stthr.tri.amidday	-0.28	0.18	-1.55	0.12
cond	stthr.tri.amorning	-0.03	0.16	-0.16	0.87
cond	hsz.std	0.00	0.10	0.03	0.98
cond	nsk.std	0.68	0.08	8.82	0.00
cond	tchiyr.std:stthr.tri.amidday	-0.09	0.18	-0.51	0.61
cond	tchiyr.std:stthr.tri.amorning	0.48	0.18	2.64	0.01
cond	tchiyr.std:nsk.std	0.04	0.07	0.63	0.53
random_effect	aclew_child_id	0.20	NA	NA	NA
random_effect	Residual	0.66	NA	NA	NA

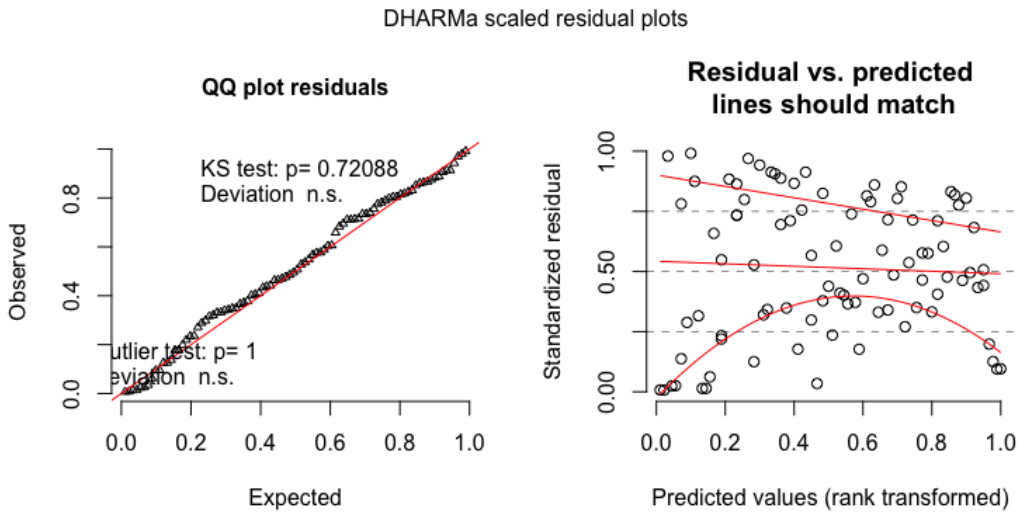

Figure 9. The model residuals from the gaussian mixed-effects regression of ODS min/hr for the random sample.

Turn-taking clips. ODS rate in the turn-taking clips demonstrated a skewed distribution Figure 10. We therefore modeled it using a negative binomial mixed-effects regression without zero inflation in the main text: results for the two models demonstrating all pairwise effects of time of day are shown in Table 13 and Table 14. The residuals for the default model (Table 13) are shown in Figure 11.

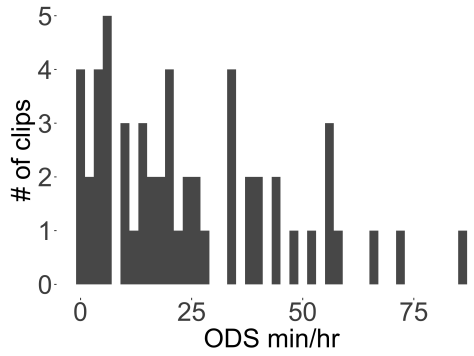

Figure 10. The distribution of ODS rates found across the 59 turn-taking clips.

Table 13

Full output of the negative binomial mixed-effects regression of ODS min/hr for the turn-taking sample, with morning as the reference level for time of day (note that most default models have midday as the reference level for time of day; the default model is changed here due to convergence issues).

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.62	0.33	7.89	0.00
cond	tchiyr.std	-0.04	0.33	-0.14	0.89
cond	stthr.trimorning	0.43	0.34	1.25	0.21
cond	stthr.triafternoon	0.35	0.35	1.00	0.32
cond	hsz.std	0.03	0.12	0.27	0.78
cond	nsk.std	0.56	0.08	6.76	0.00
cond	tchiyr.std:stthr.trimorning	-0.15	0.33	-0.44	0.66
cond	tchiyr.std:stthr.triafternoon	0.03	0.35	0.08	0.93
cond	tchiyr.std:nsk.std	-0.16	0.11	-1.51	0.13
random_effect	aclew_child_id	0.28	NA	NA	NA

Table 14

Model output of the negative binomial mixed-effects regression of ODS min/hr for the turn-taking sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.96	0.16	18.58	0.00
cond	tchiyr.std	-0.02	0.18	-0.08	0.93
cond	stthr.tri.amidday	-0.35	0.35	-1.00	0.32
cond	stthr.tri.amorning	0.08	0.17	0.47	0.64
cond	hsz.std	0.03	0.12	0.27	0.78
cond	nsk.std	0.56	0.08	6.76	0.00
cond	tchiyr.std:stthr.tri.amidday	-0.03	0.35	-0.08	0.93
cond	tchiyr.std:stthr.tri.amorning	-0.18	0.20	-0.86	0.39
cond	tchiyr.std:nsk.std	-0.16	0.11	-1.51	0.13
random_effect	aclew_child_id	0.28	NA	NA	NA

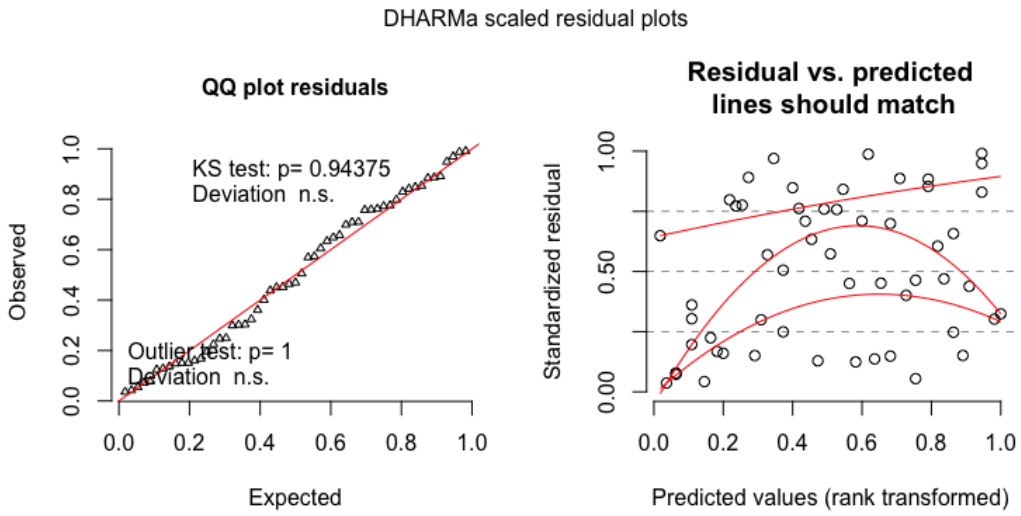

Figure 11. The model residuals from the zero-inflated negative binomial mixed-effects regression of ODS min/hr for the turn-taking sample.

As an alternative analysis we generated parallel models of ODS rate in the turn-taking clips using gaussian mixed-effects regression with logged values of ODS: results for the two models demonstrating all pairwise effects of time of day are shown in Table 15 and Table 16. The residuals for the default gaussian model (Table 15) are shown in Figure 12.

Table 15

Full output of the gaussian mixed-effects regression of ODS min/hr for the turn-taking sample, with midday as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.55	0.29	8.92	0.00
cond	tchiyr.std	-0.12	0.30	-0.40	0.69
cond	stthr.trimorning	0.37	0.32	1.16	0.25
cond	stthr.triafternoon	0.31	0.30	1.02	0.31
cond	hsz.std	0.04	0.13	0.35	0.72
cond	nsk.std	0.75	0.11	6.73	0.00
cond	tchiyr.std:stthr.trimorning	-0.07	0.30	-0.24	0.81
cond	tchiyr.std:stthr.triafternoon	0.21	0.30	0.70	0.48
cond	tchiyr.std:nsk.std	-0.20	0.14	-1.37	0.17
random_effect	aclew_child_id	0.26	NA	NA	NA
random_effect	Residual	0.61	NA	NA	NA

Table 16

Model output of the gaussian mixed-effects regression of ODS min/hr for the turn-taking sample, with afternoon as the reference level for time of day.

component	term	estimate	std.error	statistic	p.value
cond	(Intercept)	2.87	0.17	17.12	0.00
cond	tchiyr.std	0.09	0.20	0.45	0.65
cond	stthr.tri.amidday	-0.31	0.30	-1.02	0.31
cond	stthr.tri.amorning	0.06	0.22	0.28	0.78
cond	hsz.std	0.04	0.13	0.35	0.72
cond	nsk.std	0.75	0.11	6.73	0.00
cond	tchiyr.std:stthr.tri.amidday	-0.21	0.30	-0.70	0.48
cond	tchiyr.std:stthr.tri.amorning	-0.28	0.22	-1.25	0.21
cond	tchiyr.std:nsk.std	-0.20	0.14	-1.37	0.17
random_effect	aclew_child_id	0.26	NA	NA	NA
random_effect	Residual	0.61	NA	NA	NA

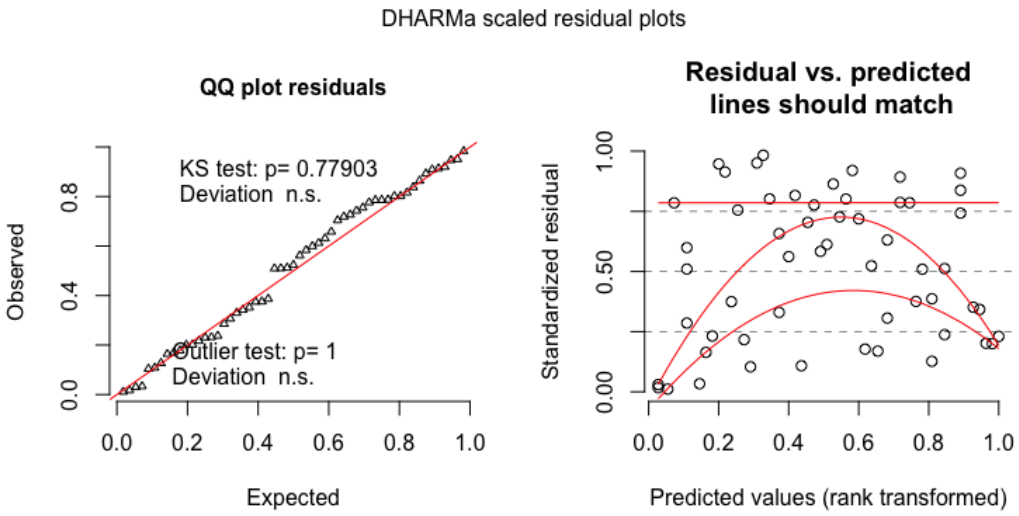

Figure 12. The model residuals from the gaussian mixed-effects regression of ODS min/hr for the turn-taking sample.

References

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., ... Bolker, B. M. (2017a). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*, 378–400.

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

Casillas, M., Brown, P., & Levinson, S. C. (forthcoming). Early language experience in a tseltal mayan village. *Child Development*, *XX*, XX–XX.