Mealtime conversations between parents and their 2-year-old children in five cultural
contexts

Manuel Bohn[1], Wilson Filipe da Silva Vieira[1], Marta Giner Torréns[2], Joscha Kärtner[2],

Shoji Itakura[3], Lilia Cavalcante[4], Daniel Haun[1], Moritz Köster[5,*], & Patricia Kanngiesser[6,*]

[1] Department of Comparative Cultural Psychology, Max Planck Institute for Evolutionary

Anthropology, Leipzig, Germany

[2] Department of Psychology, University of Münster, Münster, Germany

[3] Doshisha University, Center for Baby Science, Kyoto, Japan

[4] Department of Behavior Theory and Research, Federal University of Pará, Belém, Brazil

[5] University of Regensburg, Institute of Psychology, Regensburg, Germany

[6] University of Plymouth, School of Psychology, Plymouth, UK

[*] joint senior author

Author Note

Correspondence concerning this article should be addressed to Manuel Bohn, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. E-mail: manuel_bohn@eva.mpg.de

32   Abstract

33   Children all over the world learn language, yet, the contexts in which they do so varies

34   substantially. This variation needs to be systematically quantified to build robust and

35   generalizable theories of language acquisition. We compared communicative interactions

36   between parents and their two-year-old children (N = 99 families) during mealtime across

37   five cultural settings (Brazil, Ecuador, Argentina, Germany, Japan) and coded the amount

38   of talk and gestures as well as their conversational embedding (interlocutors, speech acts,

39   themes). We found a comparable pattern of communicative interactions across cultural

40   settings, which were attenuated in ways that likely reflect local norms and values. These

41   results suggest that children encounter similarly structured communicative environments

42   across diverse cultural contexts and will inform theories of language learning.

43   *Keywords:* Language acquisition, Communication, Gesture, Cross-cultural

44   psychology, Parent-child interaction

45   Word count: X

Mealtime conversations between parents and their 2-year-old children in five cultural contexts

Children learn language in interactions with language-competent others (Bohn & Frank, 2019; Bruner, 1983; Clark, 2009; Levinson & Holler, 2014; Tomasello, 2009). Social interactions between children and their social partners are structured by norms, values, and beliefs that vary substantially across cultural and historical contexts (Rogoff et al., 2003). As a consequence, children may encounter dramatically different language learning environments. Yet, the fact that children usually achieve fluency in their local language(s) suggests that they use a suite of compensatory learning strategies to adapt flexibly to their respective learning environment (Cristia, 2022; Kidd & Garcia, 2022; Rowe & Weisleder, 2020). Explaining how children accomplish this feat poses a serious theoretical and empirical challenge. Detailed documentation of learning environments across cultural contexts is needed to inform theorizing about children's learning processes. In this paper, we contribute to this effort by reporting on cross-cultural variation in parent-child communicative interactions in a semi-structured setting: meals involving parents and their 2-year-old child.

In recent decades, research on language acquisition has focused, to a large extent, on variation in language input and, in particular, the number of words children hear in naturalistic settings. This line of work was sparked by the finding that children who receive more input – especially speech directly addressing them – have better language skills (Bang, Bohn, Ramirez, Marchman, & Fernald, 2022; Hart & Risley, 1995; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Shneidman & Goldin-Meadow, 2012; Walker, Greenwood, Hart, & Carta, 1994; Weisleder & Fernald, 2013). From a theoretical perspective, more language input increases children's opportunities for learning word-meaning mappings and allows them to build a larger vocabulary (Jones & Rowland, 2017; Kachergis, Marchman, & Frank, 2022; McMurray, Horst, & Samuelson, 2012). The introduction of daylong audio

recording devices and automated coding algorithms has provided further momentum to this endeavor (Cristia et al., 2021; Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011; Lavechin, Bousbib, Bredin, Dupoux, & Cristia, 2020). As a consequence, the quantity of direct language input plays a central role in theories and formal models of language learning (Braginsky, Yurovsky, Marchman, & Frank, 2019; Goodman, Dale, & Li, 2008; Kachergis et al., 2022; Swingley & Humphrey, 2018).

However, like most of developmental psychology (Amir & McAuliffe, 2020; Nielsen, Haun, Kärtner, & Legare, 2017), research on language acquisition has largely focused on affluent societies of the global north and, as a consequence, the resulting theoretical proposals may fail to generalize to other cultural contexts. As studies in a greater variety of cultural settings have begun to accumulate (Altınkamış, Kern, & Sofu, 2014; Bergelson et al., 2019; Bunce et al., 2020; Casillas, Brown, & Levinson, 2021; Choi, 2000; Cristia, Dupoux, Gurven, & Stieglitz, 2019; Loukatou, Scaff, Demuth, Cristia, & Havron, 2021; Tardif, Shatz, & Naigles, 1997), they have revealed substantial cultural variation in how much direct input children receive (Cristia, 2022; see also Sperry, Sperry, & Miller, 2019 for variation within an English-speaking sample). Yet, children still reach major milestones in language development at similar ages (Brown & Gaskins, 2014; Casillas, Brown, & Levinson, 2020). These findings highlight that theories and models of language learning need to extend beyond quantity of input and also include learning processes that compensate for variation in input (Bang, Mora, Munévar, Fernald, & Marchman, 2022; Casillas, 2022; Jones & Rowland, 2017; Kachergis et al., 2022; Meylan & Bergelson, 2022).

It has been suggested that these compensatory learning processes leverage structural features of social interactions in which language is used (Casillas et al., 2020; Rogoff, Paradise, Arauz, Correa-Chávez, & Angelillo, 2003; Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016). Pragmatic accounts of language learning offer an explanation for how children use contextual information (e.g., Bohn & Frank, 2019; Tomasello, 2009): Social interactions, especially routines, follow predictable patterns that

99   make it easier for children to infer what speakers are communicating about (Barbaro &

100 Fausey, 2022; Lieven, 1994; Masek, Ramirez, McMillan, Hirsh-Pasek, & Golinkoff, 2021).

101 For instance, Roy, Frank, DeCamp, Miller, and Roy (2015) found that words were more

102 easily learned when they were primarily used in a distinct spatial and temporal context.

103 Similarly, establishing common ground over the course of an interaction provides

104 information about the speaker's intention independent of the words that are being used

105 (Bohn & Köymen, 2018; Bohn, Tessler, Merrick, & Frank, 2021). For example, Bohn, Le,

106 Peloquin, Köymen, and Frank (2021) showed that children identify the referent of an

107 ambiguous word by inferring the topic of an ongoing conversation (see also Akhtar, 2002).

108 These findings help to explain why the amount of conversational turn-taking in

109 parent-child interactions predicts child language outcomes (Donnelly & Kidd, 2021; Romeo

110 et al., 2018). Turn-taking results in continuous, structured conversations that provide

111 information-rich learning opportunities.

112   In order to assess whether children can use structural features to complement direct

113 verbal input, it is crucial to compare communicative interactions between adults and

114 children across cultural settings. However, to our knowledge, there are very few

115 quantitative comparisons. While ethnographic descriptions offer important and rich

116 insights into individual cultural settings (see e.g., De León, 2011; Gaskins, 2006),

117 quantitative comparisons are essential for understanding gradual cultural differences

118 (Broesch et al., 2021; Hewlett, Lamb, Shannon, Leyendecker, & Schölmerich, 1998; Köster

119 et al., 2022) and offer core input for theory building (see language input studies discussed

120 earlier).

121   One of the challenges of cross-cultural work lies in selecting an appropriate context

122 for comparing the structure of communicative interactions (Broesch, Lew-Levy, Kärtner,

123 Kanngiesser, & Kline, 2022). Prior work has shown that the amount of language input

124 children receive varies substantially across routine activities. For example, Soderstrom and

125 Wittebolle (2013) found that Canadian adults spoke most during book reading and

structured playtime (see also Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2019). Such activities, however, are very specific to industrialized societies and less frequent or absent in other cultural contexts. A cross-culturally recurrent, and hence particularly promising, context for cross-cultural research is mealtime: across societies, meals are social events that are structured by – and used to transmit – cultural norms, values and beliefs (Blum-Kulka, 2012; Fjellström, 2004; Köster et al., 2022; Ochs & Shohet, 2006). Furthermore, mealtimes have been a fruitful context for studying caregiver-child communication, for example, in the U.S. (e.g., Beals, 1993, 1997; Snow & Beals, 2006).

The current study

The goal of this study was to compare communicative interactions between parents and their children during mealtime across diverse cultural settings. We aimed for a naturalistic but comparable setup by a) asking families to record in their homes, b) recruiting families with a single – usually the first – child between 2 and 3 years of age and c) focusing on 10-minute-long episodes during which three family members (mother, father, one child) were present. Even though the constellation of two parents and one child might be less representative of the overall family demographics in some settings, it allowed us to directly quantify and compare communicative interactions. We obtained recordings from five different cultural settings, including families living in the Global South and North, as well as in urban and rural settings: the city of Buenos Aires, Argentina, small villages in the Apeú region, Brazil, small villages close to Cotacachi, Ecuador, the city of Münster, Germany, and the city of Kyoto, Japan.

We coded and analyzed the data along nine dimensions that focused on the quantity of talk and gestures as well as their conversational embedding (interlocutors, speech acts, themes). In a first step, we analyzed if and how these dimensions differed in the five cultural settings. In a second step, we asked whether some cultural settings are more similar to one another. The five cultural settings offer an interesting perspective on the

factors influencing mealtime conversations. For example, communicative interaction patterns could cluster by country (five clusters; one cluster per country), or by language family and geographical region (three clusters; Argentina, Brazil, Ecuador vs. Germany vs. Japan) or by degree of urbanization (two clusters; urban: Argentina, Germany, Japan vs. rural: Brazil, Ecuador). Based on previous work, we expected less direct input to children in the rural contexts (Cristia, 2022) but – given a lack of comparable previous work – we had no specific predictions for variation in the structure of communicative interactions.

Methods

Participants

The final sample consisted of 99 families from five cultural contexts. This included 20 families from the city of Buenos Aires, Argentina (urban setting), 18 families from villages in the Amazon region near Apeú, Brazil (rural setting), 13 from villages near Cotacachi, Ecuador (rural setting), 24 families from the city of Münster, Germany (urban setting) and 24 families from the city of Kyoto, Japan (urban setting). For the recording sessions, all families comprised a father, a mother and a child aged between 2 years and 3 years, 2 months. Almost all children were the first child in the family. Some videos partly included additional children (n = 1 for Argentina, Brazil and Ecuador).

Additional families were recorded but they did not meet the inclusion criteria of at least one recording of a meal that lasted for at least ten minutes, initially included all three family members and had all family members visible in the recording. This resulted in the exclusion of 11 families from Münster, Germany, 34 from Apeú, Brazil, five from Buenos Aires, Argentina, 39 from Cotacachi, Ecuador and five from Kyoto, Japan.

The recordings were collected as part of a larger cross-cultural investigation into parent-child interactions and findings on parental teaching behaviors have been published

177 by Köster et al. (2022). We refer to this earlier work for a detailed description of each

178 cultural setting. In the following we only provide a short overview.

179   **Argentina.**   Families lived in the metropolitan area of Buenos Aires, Argentina,

180 which comprises around 15.2 million people. They were recruited via personal contacts of

181 the local experimenter. The family language was Rioplatense Spanish. Compensation

182 included small toys for children and USD 10 for parents. Most parents had completed a

183 university degree (mothers: 74%; fathers: 52%) and engaged in paid professional labor

184 (mothers: 87%; fathers: 78%). The majority of children (91%) either attended

185 kindergarten or were looked after by a nanny or a family member other than the parents.

186   **Brazil.**   Families lived in villages of around 50 - 300 families in the Amazon region

187 near Apeú, approximately 1.5 hours east of Belém, the capital of the state of Pará. They

188 were recruited with the help of a local public health office. The family language was

189 Brazilian Portuguese. Compensation included small toys for children and a certificate of

190 participation for parents. Most parents had completed secondary school (~12 years of

191 schooling, mothers: 50%; fathers: 56%). Mothers worked mainly as housewives (83%) while

192 fathers engaged in paid labor (100%). Some families engaged in traditional subsistence

193 activities such as tapioca farming, livestock breeding, or açaí and fruit harvesting. In line

194 with employment status, the majority of children were looked after by their mothers.

195   **Ecuador.**   Families identified as belonging to the Kichwa community and lived in

196 villages with 800-5,000 inhabitants located within 1 hour (by car) of the city of Cotacachi

197 in the Imbabura province. They were recruited via personal contacts mediated by the

198 community president. The family language was Ecuadorian Spanish with elements of

199 Kichwa. Compensation included food (e.g., rice or oat) and USD 4. Most parents had

200 completed primary school (~10 years of schooling, mothers: 50%; fathers: 56%). Mothers

201 worked mainly as housewives (59%) while fathers engaged in paid labor (77%). Around

202 40% of children were looked after by a person other than the mother during the day.

Germany.   Families lived in Münster in the state of North-Rhine-Westphalia, a city with ~310,000 inhabitants. They were recruited via a participant database of the Developmental Psychology lab at the University of Münster. Compensation included a voucher of EUR 15 for a local toy store. Most parents had completed a university degree (mothers: 71%; fathers: 71%) and engaged in paid professional labor (mothers: 92%; fathers: 92%). All children either attended kindergarten or were looked after by a nanny during the day.

Japan.   Families lived in the city of Kyoto, in the Kansai metropolitan region, with around 1.5 million inhabitants. They were recruited via a participant database of the Center for Baby Science at Doshisha University. Compensation was JPY 3000. Most parents had completed a university degree (mothers: 92%; fathers: 83%) and engaged in paid professional labor (mothers: 71%; fathers: 100%). Most children (80%) attended kindergarten.

The study was approved by the ethics committee of the Free University of Berlin. Recordings took place between September 2017 and March 2019. Informed verbal consent was obtained from both parents and written consent from one of the parents.

Procedure

We visited families twice. On the first visit, an experimenter (familiar with the local language) instructed parents on how to use the video camera and what to record. We encouraged families to record two instances of the meal they commonly shared together, which happened in the evening for most families. The cameras were equipped with a wide-angle lens and set up to capture all family members during the meal. In addition to video, the cameras also recorded sound. On the second visit, the experimenter asked about the recordings and encouraged families to record additional meals if they had not already recorded two sessions. In the end, we collected socio-demographic information and interviewed the mothers (unrelated to the present study).

Coding

We scanned all recordings for sections that captured a meal event, lasted at least 10 minutes, and included all three family members. For each family, we selected one such section for in-depth coding and excluded all families for which we did not find such a section (see above for the number of excluded families).

We coded videos using ELAN (Wittenburg, Brugman, Russel, Klassman, & Sloetjes, 2006) version 6.4. The primary coder was either a native (Germany, Japan, Brazil) or a highly fluent (Argentina, Ecuador) speaker of the local language. For Ecuador, a native speaker translated sections containing Kichwa into Spanish before the primary coder coded them.

In a first pass, the primary coder created a tier for each speaker and marked segments in which this person was speaking or using a gesture. In a second pass, the coder transcribed all utterances into the local language and coded their conversational embedding. We defined utterances as sections of continuous talk by one person. If speakers paused for more than 2 seconds, we coded two utterances with 2 (or more) seconds of silence in between. We used the following codes to capture the conversational embedding of each utterance:

Speaker.   Here we coded who produced the utterance. The speaker could either be `child`, `mother`, or `father`. All sections containing no speech were coded as `silence`.

Recipient.   Here we coded who the utterance was addressed to. Codes could either be `child`, `mother`, `father`, `both` or `other`, where `other` was used either when a fourth person (e.g., over the phone) was addressed or the speaker was talking to themselves (e.g., child babbling or singing). If an utterance addressed two people in sequence, the second addressee was coded as the recipient.

Themes and rounds.   Here we coded the conversational coherence of the different utterances. For that we defined `themes` as sequences of utterances that related to one

255 another. This applies for example to sequences of questions and answers but also to

256 sequences in which the content of an utterance is directly related to the content of the

257 previous utterance. Please note that such themes were coded locally and were not the same

258 as topics. For example, if father and child exchanged four utterances about the child's day

259 in the kindergarten this was coded as one theme. If the same topic (day at the

260 kindergarten) came up later again, this was coded as a separate theme. Each utterance

261 within a theme was counted as a `round` to capture the sequence and length of a theme.

262 Thus, each utterance was assigned a number for the theme and a number for the round

263 within theme. Themes could have interjections of one or two utterances. After more than

264 two interjections we coded a new theme. For example, if father and child talked about food

265 and the mother made an unrelated comment in between, the mother's comment would be

266 coded as a separate theme while the other theme continued around it:

267   Child: "I want more" (theme (t) 1, round (r) 1)

268   Father: "Do you want more soup?" (t1, r2)

269   Mother: "Phew, I'm hot (t2, r1)

270   Child: "No, bread (t1, r3)

271   Father: "I'll get some" (t1, r4)

272   **Speech acts.**   Each utterance was coded as either being a `question`, `assertion` or

273 `imperative`. Imperatives were only coded if the the utterance was grammatically

274 structured as an imperative. For example "Pass me the salt!" was coded as an imperative

275 while "You should give me the salt." was not.

276   **Referential gestures.**   We also coded the frequency of two types of referential

277 gestures for each individual. `Points` were coded when someone indicated an object,

278 location or person in the environment, either using a finger (often index finger), the head

279 or an object (e.g., cutlery). Reaches and hold-outs were not coded as points. `Iconic`

280 gestures were coded when someone depicted an object or action using their hands and/or

body (e.g., pretending to hold a knife and cut to instruct the child how to cut a cucumber).
Conventional gestures such as head shaking, nodding or shrugging were not coded.

Reliability coding. For each cultural setting, we selected 15% of videos and had them re-coded by a second coder (native speaker of the respective language). The second coder relied on the sequencing of the primary coder. Inter-rater reliability was generally very good. For recipient, the agreement between coders was 88% ($\kappa = 0.83$), for speech acts it was 91% ($\kappa = 0.78$) and for gestures it was 96% ($\kappa = 0.81$). To get inter-rater reliability for the coding of themes, we asked whether the two coders agreed on whether a given utterance belonged to the same theme as the previous utterance or belonged to a new theme. Once again, agreement between coders was high (agreement $= 87\%$, $\kappa = 0.74$).

Analysis and Results

For each of the research questions (see below), we defined a response variable and then used Bayesian multilevel regression models fit via the function `brm` from the package `brms` (Bürkner, 2017) to model the effect of cultural setting and – whenever applicable – that of the different individuals involved in the conversation. To make inferences about the importance of predictors, we compared a set of nested models including cultural setting and individual as predictors to each other and to a null model that did not include them to test if these predictors improved model fit. Following McElreath (2018), we compared models using Widely Applicable Information Criteria (WAIC) and the corresponding weights. This approach favors models that have high out-of-sample predictive accuracy in that they achieve a good fit to the data with the minimal set of parameters.

We modeled the effect of cultural settings as random effects and interactions between additional variables (e.g., speaker identity) and setting as random slopes within cultural setting (`brms` notation: `(variable|setting)`). This approach partially pools model estimates and is thought to yield more generalizable results because it avoids overfitting the model to the observed data (Gelman & Hill, 2006; McElreath, 2018). For each model

307 comparison, we visualized the predictions of the winning model and interpreted them based

308 on their posterior means and 95% Credible Intervals (CrI). We used default priors built

309 into `brms` for all parameters.

How much time did families spend talking?

311   First, we ask how much time families spent talking as opposed to being silent and

312 how this varied across cultural settings. The dependent variable in this case was the total

313 lengths of all sections coded as silence for each family (modeled as a normal distribution).

314 We compared a null model including only an overall intercept (`silence ~ 1`) to a model

315 including cultural setting (`silence ~ 1 + (1|setting)`).

316   The model comparison clearly favored the model including cultural setting (WAIC =

317 338.84, se = 14.93, weight = 1.00) over the null model (WAIC = 362.36, se = 14.97, weight

318 = 0.00). The model predicted an average of 4.95 [95%CrI = 3.80 - 6.07] minutes of silence

319 across cultural settings. Ecuador and Brazil had longer sections of silence compared to

320 Argentina and Germany, with Japan falling in the middle (see Figure 1A).

How much talk is directed at each family member?

322   Next, we asked whom talk was directed to, that is, how much "input" each family

323 member received. The dependent variable was the total lengths of utterances directed at

324 each individual in a family. This variable was right-skewed and we therefore modeled it as

325 a skewed normal distribution. Given that the analysis above showed that the amount of

326 overall talk differed across cultural settings, the null model already included a random

327 effect for setting (`input ~ 1 + (1|setting) + (1|family)`). We compared it to two

328 alternative models, one assuming that input additionally differed across recipients (`input`

329 `~ recipient + (1|setting) + (1|family)`) and one assuming that this effect in turn

330 varies across settings (`input ~ recipient + (recipient|setting) + (1|family)`).

The model comparison favored the two alternative models, with a slight preference for the simpler model that did not assume the effect of recipients to vary across cultural setting (WAIC = 705.72, se = 30.16, weight = 0.74; model assuming variation across settings: WAIC = 707.82, se = 30.15, weight = 0.26). We observed that, across settings, more talk was directed at children compared to the two parents with fathers being talked to the least (see 1B).

Which family member talks the most?

In the next analysis, we asked how talking time was distributed across the different family members. The dependent variable was the total lengths of utterances of each individual in a family, which was also right-skewed and modeled as a skewed normal distribution. Given previous results, the null model included a random effect for setting (`talk ~ 1 + (1|setting) + (1|family)`). The first alternative model assumed that talk differed across speakers (`talk ~ recipient + (1|setting) + (1|family)`), the second assumed that this effect interacted with setting (`talk ~ recipient + (recipient|setting) + (1|family)`).

The model comparison clearly favored the interaction model assuming that the the difference between speakers varied across settings (WAIC = 755.92, se = 25.20, weight = 1.00; model assuming no interaction: WAIC = 772.14, se = 24.65, weight = 0.00). Figure 1C shows that even though mothers talked the most in all settings, this effect was much more pronounced in Japan, Germany and Argentina compared to Ecuador and Brazil.

How many gestures are being used?

To conclude the first set of analysis, we looked at variation in gesture production. Iconic gestures were produced at a much lower rate (only ~15% of the 1484 gestures were iconic gestures), resulting in many empty cells for combinations of individual and cultural

setting. This made it difficult to analyze points and iconic gestures separately and we instead decided to combine them. Thus, the dependent variable was the number of gestures produced by each individual. We modeled this distribution as a zero-inflated poisson distribution to account for the fact that some individuals did not produce any gestures.

The null model only included an intercept and a random effect of family (`gestures ~ 1 + (1|family)`). There were three alternative models: the first included producer (child, mother, father) as a fixed effect (`gestures ~ producer + (1|family)`), the second model added to this a random effect for setting (`gestures ~ producer + (1|setting) + (1|family)`) and the third model included an additional random slope for interlocutors within setting to model the interaction (`gestures ~ producer + (producer|setting) + (1|family)`).

The model comparison clearly favored the model assuming that the number of gestures produced varied between individuals within cultural settings (interaction model; WAIC = 1602.79, se = 49.79, weight = 1.00; second best model (without interaction): WAIC = 1670.90, se = 53.44, weight = 0.00). Overall, there were slightly fewer gestures in Ecuador and Brazil. Looking at the different individuals, we saw that – across settings – children produced the most gestures, followed by mothers and then fathers. This pattern was less pronounced in Brazil and Argentina and notably reversed in Ecuador, where children produced hardly any gestures (see Figure 1D).

Who talks to whom?

To address the question of who talks to whom we categorized the conversational partners of each utterance as either being mother and father, child and mother or child and father. We then used a categorical model to predict the proportion with which each of these categories occurred. The null model only included an intercept and a random effect of family (`partners ~ 1 + (1|family)`) while the alternative model assumed that these
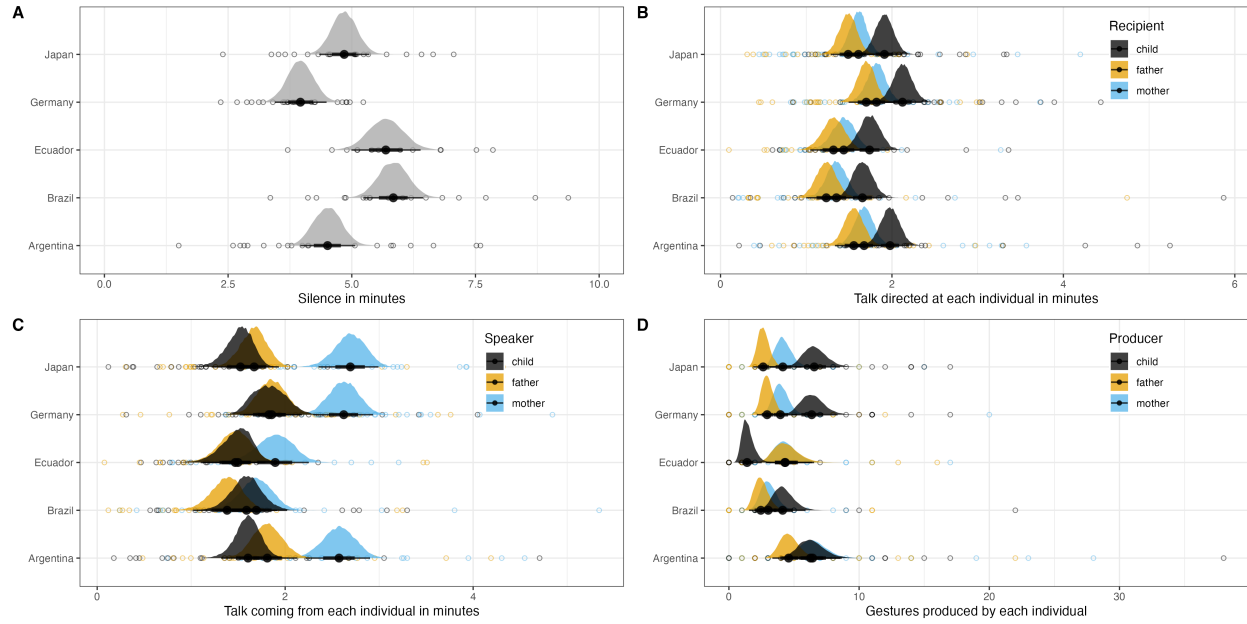
Figure 1. A: Silence across cultural settings. B: Talk directed at the different individuals. C: Time spent talking by the different individuals. D: Number of gestures (points and iconic gestures combined) produced by each individual. In B-D: color denotes the individual. Distributions show the predicted values based on the respective model with solid points and error bars showing the mean with 66% and 95% CrI. Light points show the aggregated data for each family and – whenever applicable – individual.

380 proportions differ across settings (`partners ~ 1 + (1|setting) + (1|family)`).

381 The model comparison yielded no clear difference between models, suggesting no
382 substantial differences in the proportion of conversational partners across settings (null
383 model: WAIC = 28107.31, se = 116.83, weight = 0.38; alternative model: WAIC =
384 28106.33, se = 116.90, weight = 0.62). Compared to an equal split (proportion of 0.33 for
385 each category), conversations between mother and child were slightly more frequent and
386 conversations between child and father less frequent except for Brazil where conversations
387 between mother and father were less likely (see Figure 2A).

Who uses which speech acts?

As the next step, we analyzed how the different speakers used speech acts – assertions, imperatives, and questions. That is, we predicted the proportion with which each speech act occurred using a categorical model. We investigated whether the types of speech acts used varied with speakers as well as cultural setting. The null model only included an intercept and a random effect of family (`speech_act ~ 1 + (1|family)`). There were three alternative models: the first included speaker as an additional fixed effect (`speech_act ~ speaker + (1|family)`), the second model added to this a random effect for setting (`speech_act ~ speaker + (1|setting) + (1|family)`) and the third model included and additional random slope for speaker within setting to model the interaction between speaker and setting (`speech_act ~ speaker + (speaker|setting) + (1|family)`).

The model comparison clearly favored the interaction model assuming that the type of speech act varied across speakers within cultural setting (WAIC = 23591.46, se = 180.20, weight = 1.00; second best model (without interaction): WAIC = 23689.02, se = 181.03, weight = 0.00). The general pattern was that assertions were the most frequent type of speech act, followed by questions and imperatives. This ordering was much more pronounced in children in that they hardly used questions or imperatives. Variation across settings was most notable in that both mothers and fathers from Brazil and Ecuador were substantially more likely to use imperatives compared to the other three settings (see Figure 2B).

How many people are involved in a theme?

Next, we turned to themes as the focus of analysis. As a first step, we asked how many different speakers were involved in a theme. To be involved in a theme, an individual had to produce at least one utterance. Please note that it was possible for themes to have
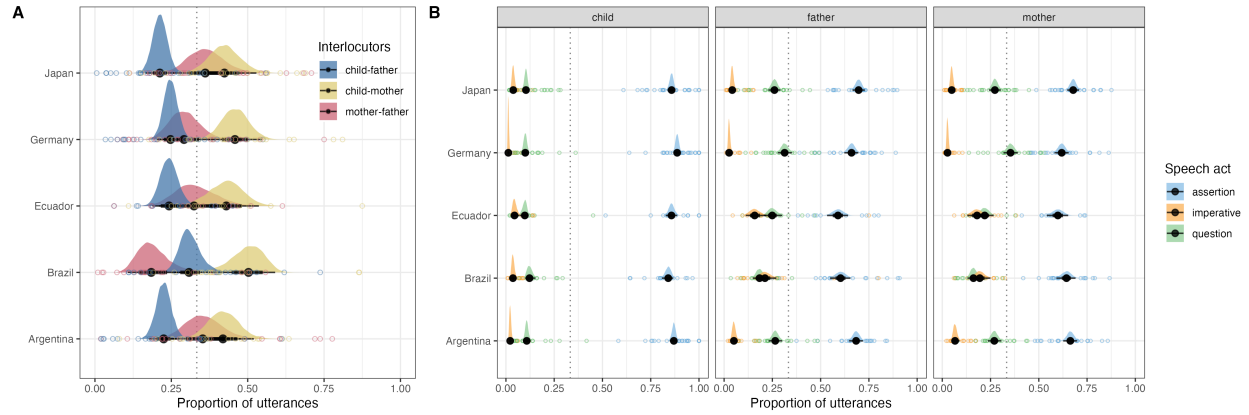
Figure 2. A: Proportion of utterances that were exchanged by a pair of interlocutors. Color shows the interlocutors involved in the utterance regardless of direction (i.e., identity of speaker and listener). B:Proportion of utterances that belonged to a certain class of speech acts. Facets show different speakers, color denotes the type of speech act. Distributions show the predicted values based on the respective model with solid points and error bars showing the mean with 66% and 95% CrI. Light points show the aggregated data for each family.

only one speaker. In fact, this was the case for 34%% of all utterances. These themes were mostly single utterances that occurred when someone made an unrelated comment or asked a question but did not receive an answer. We counted the number of speakers involved in each theme (1, 2, or 3) and modeled the resulting distribution using a binomial model. Note that this approach does not take into account the length of each theme. We compared a null model including only an overall intercept (`no_speakers ~ 1`) to a model including cultural setting (`no_speakers ~ 1 + (1|setting)`).

The model comparison favored the model including cultural setting (WAIC = 6544.58, se = 41.13, weight = 0.95) over the null model (WAIC = 6550.36, se = 40.88, weight = 0.05). Figure 3A shows that the number of speakers involved in a theme was relatively similar across cultural settings, with Brazil being the notable exception in having, on average, more speakers per theme.

Who initiates themes?

In the following analysis, we asked whether there are differences among speakers and cultural settings in who initiated a theme. For each theme, we only selected the first utterance and used a categorical model to predict the probability with which each individual was the speaker of that utterance and thus the initiator of the theme. Once again, we compared a null model including only an overall intercept (`initiator ~ 1`) to a model including cultural setting (`initiator ~ 1 + (1|setting)`).

The model comparison favored the model including cultural setting (WAIC = 6566.90, se = 26.07, weight = 0.73) over the null model (WAIC = 6568.84, se = 25.58, weight = 0.27). However, the difference between models was rather small, suggesting that there were no pronounced differences between cultural settings. Overall, there were no huge differences between the three individuals in terms of the probability of being the initiator of a theme (range: 0.26 to 0.41). Compared to an equal split, mothers were slightly more likely to initiate themes and fathers less likely. This relative pattern held for all cultural settings, except Brazil, where the child was the most likely initiator of a theme (see Figure 3B).

How long do themes last?

We finished the analysis of themes by asking about variation in how long themes lasted (i.e., how many rounds there were in a theme). For each theme, we noted its length (i.e., the maximum round) and the main interlocutors. For that, we counted how many utterances were exchanged between all possible pairs in each theme and classified each theme as being mainly a conversation between those interlocutors who exchanged the most utterances. As a consequence, we excluded all themes that only had a single round and only involved a single speaker. The dependent variable (length of the theme) was heavily right-skewed and close to zero and we, therefore, used a log-normal distribution to model it.

450   The null model only included an intercept and a random effect of family

451 (`theme_length ~ 1 + (1|family)`). There were three alternative models: the first

452 included interlocutors as a fixed effect (`theme_length ~ interlocutors + (1|family)`),

453 the second model added to this a random effect for setting (`theme_length ~`

454 `interlocutors + (1|setting) + (1|family)`) and the third model included and

455 additional random slope for interlocutors within setting to model the interaction between

456 interlocutors and setting (`theme_length ~ interlocutors + (interlocutors|setting)`

457 `+ (1|family)`).

458   The model comparison clearly favored the interaction model assuming that the

459 difference in length of themes for each pair of interlocutors varied across cultural settings

460 (WAIC = 11657.48, se = 106.30, weight = 1.00; second best model (without interaction):

461 WAIC = 11671.33, se = 106.59, weight = 0.00). The average predicted length of a theme

462 across interlocutors and settings was 5.71 rounds [95%CrI = 3.95 - 8.35]. Figure 3C

463 indicates a variable pattern across cultural settings. In Japan, themes were approximately

464 equally long for all pairs of interlocutors. In the other settings, conversations between

465 mother and father were shorter compared to conversations between one of the parents and

466 the child. This pattern was less pronounced in Ecuador compared to Germany, Brazil and

467 Argentina. Overall, themes lasted slightly longer in Brazil compared to the other settings.

Family level clustering

469   In this final analysis, we took a more holistic look at the data and tried to identify

470 patterns across the communicative dimensions analyzed above. That is, we asked if there

471 were clusters within our sample that represent different communicative profiles. This

472 allowed us to see a) if families clustered based on cultural settings and b) how the different

473 cultural settings clustered with each other. To construct the data set for this analysis, we

474 computed the following dimensions for each family: the amount of Silence, the proportion

475 of utterances coming from each individual (`Father speaker`, `Mother speaker`, and `Child`
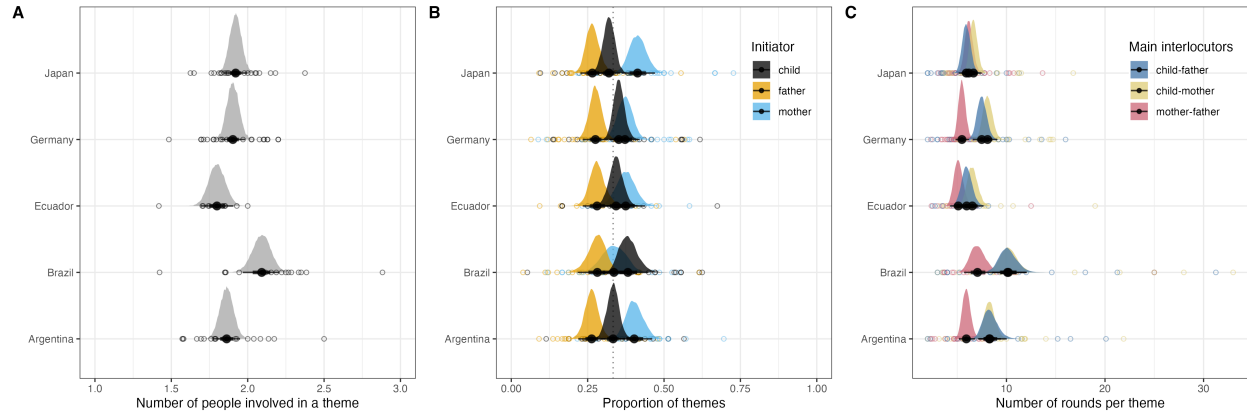
Figure 3. A: Average number of people involved in a theme. B: Proportion of themes as a function of who initiated them. Color shows the initiator. C: Number of rounds per theme depending on the interlocutors involved. Color shows the interlocutors who exchanged the most utterances within a given theme. Distributions show the predicted values based on the respective model with solid points and error bars showing the mean with 66% and 95% CrI. Light points show the aggregated data for each family.

speaker), the proportion of `Questions`, `Assertions`, and `Imperatives`, the number of `Gestures`, the number of `Themes`, the average number of `Rounds` per theme, and the average number of `Speakers per theme`. Please note that more granular dimensions (e.g., gestures or speech act types separate for each individual) would have been possible. However, because this would have meant that each dimension would have had to be estimated based on less data (resulting in a more noisy estimate), we decided to use a more coarse approach.

We performed *k*-means clustering on the data using the function `kmeans` from the `stats` package which is a native component of `R`. This analysis partitions the data into *k* clusters so that the sum of squares from points to the assigned cluster centers – in the multidimensional space that is defined by the different dimensions – is minimized. We used the default *Hartigan-Wong* algorithm to find these cluster centers (Hartigan & Wong, 1979). To determine the number of clusters, we used the *silhouette* and *elbow* methods via

the function `fviz_nbclust` from the `factoextra` package (Kassambara & Mundt, 2020). Both suggested two clusters as the optimal solution.

Figure 4A visualizes the clustering of families based on this analysis. The first cluster (blue), included mainly families from Argentina, Germany and Japan. Within the cluster, there was no further clustering of families by cultural setting. The second cluster (gold), mainly comprised families from Ecuador and Brazil. Within that cluster, families further tended to cluster by cultural setting, with families from Brazil being more similar to each other compared to families from Ecuador.

In comparison to the first cluster, the second cluster (mainly Ecuador and Brazil) was characterized by overall less talk (more silence), a higher proportion of child- compared to parental-talk, and fewer gestures. Furthermore, there were fewer themes, but themes had more speakers and lasted longer. Finally, there was a higher proportion of imperatives and thus fewer assertions and questions (see Figure 4B).

Figure 4C shows the correlations between the different dimensions across clusters. Besides some expected patterns (e.g., negative correlation between proportion of talk from the different individuals) there were some notable associations: more silence was associated with a higher proportion of imperatives, themes had more rounds the more speakers were involved, and a larger number of questions was associated with more themes.

Discussion

We investigated parent-child communicative interactions during mealtimes in five cultural settings. Each family comprised a father, mother and one child and we analyzed 10 minutes of video recordings. We found that families from Ecuador and Brazil communicated less overall compared to families from Argentina and Germany, with Japan falling in the middle. Across settings, there was a common pattern in how talk was distributed across family members: mothers talked the most and children were addressed
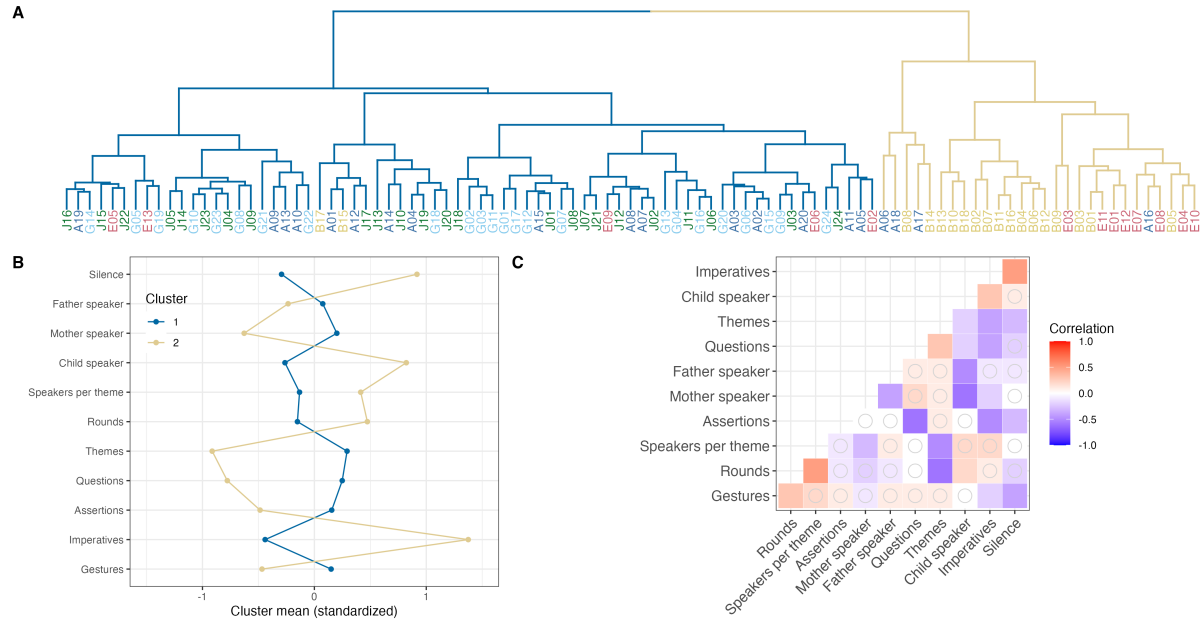
Figure 4. A: Dendrogram visualizing the similarity between families based on a cluster analysis assuming two clusters. Line colors show the two clusters, color of letters for family corresponds to the different cultural settings The first letter of the family name denotes the cultural setting (e.g., J = Japan). B: Mean values for the two clusters for each (standardized) dimension on which the cluster analysis was based. C: Pearson correlations between the different dimensions entering the cluster analysis. Color of cells shows the size and direction of the correlation coefficient. Cells without circles show correlations with p-values < 0.05.

most frequently. Assertions were the most common type of speech act for all speakers in all settings, followed by questions and imperatives. However, mothers and fathers form Brazil and Ecuador were more likely to use imperatives, mirroring the findings from Köster et al. (2022) that parents used a higher number of prompts towards their children in these contexts. The number of themes – parts of coherent utterances – tended to be longer and involved more people in Brazil compared to the other settings. When investigating how families clustered based on their communicative interaction patterns, we found what can be described as an urban-rural split, with families from urban settings (Argentina, Germany, Japan) being more similar to each other compared to families from rural settings (Brazil,

523 Ecuador). These systematic, quantitative comparisons provide an important step towards

524 understanding the similarities and differences in communicative contexts in which children

525 learn language.

526   Our findings echo how Barrett (2020; see also Kärtner, Schuhmacher, & Torréns,

527 2020) summarized much of cross-cultural research in the last two decades: *variation on a*

528 *theme.* For every aspect of communicative interaction we investigated, there was a

529 dominant pattern which described behavior in most of the cultural settings, but which was

530 often attenuated in one or two settings. Attenuation meant that the predicted means for

531 some of the settings were shifted while the distributions of families were largely

532 overlapping. For example, on average, the number of people involved in a theme was

533 around 1.8, with the highest predicted average for Brazil ($\sim$ 2.1) and the lowest for

534 Ecuador ($\sim$1.6), yet, the minimum family average in Brazil was 1.40 and the maximum for

535 Ecuador was 2. Or, mothers talked the most in all settings but the difference compared to

536 father and child was less pronounced in Ecuador and Brazil. Thus, we may tentatively

537 conclude that these overlaps in communicative patterns allow children to use similar

538 learning strategies across settings – in particular those strategies that leverage the

539 structure of the communicative context (Casillas et al., 2020; Rogoff, Paradise, et al., 2003;

540 Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016).

541   The overall pattern – or *theme* – can be summarized as being child-centered. Across

542 cultural settings, most talk was directed towards the child and themes had more

543 conversational turns (i.e., number of rounds) when the child was involved. The latter

544 finding corresponds well with the idea that children's language learning benefits from

545 coherent and structured interactions (Casillas et al., 2020; Rogoff, Paradise, et al., 2003;

546 Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016). Mothers seemed to

547 be the driving force behind this pattern: they spoke the most, initiated most themes and

548 most of the themes they were involved in also included the child. This aligns with the

549 former analyses of these videos showing that mothers teach more compared to fathers

(Köster et al., 2022) and a recent study by Broesch et al. (2021) who described mothers as the primary interaction partners for young children across five cultural settings. Fathers spoke less and were less likely to be involved in a conversation with the child. As mentioned above, this overall pattern was attenuated in some of the cultural settings and in the following we will take a closer look at this variation.

The cluster analyses showed that families' communicative interaction patterns co-varied with the degree of urbanization. Families from Brazil and Ecuador were more similar to each other than they were to families from Argentina, Germany and Japan. Interestingly, within the rural cluster, there seemed to be a further grouping by setting. This was not the case within the urban cluster: even though they lived in very different geographical regions and spoke very different languages. That is, families from Argentina, Germany and Japan were not more similar to families from the same setting than they were to families from the other settings. However, the urban/rural split was by no means complete in that some of the families from Brazil and Ecuador were assigned to the urban cluster and some families from Argentina were grouped in the rural cluster. A similar difference between urban and rural settings was found when analyzing parental teaching behavior for these samples but with a stronger sub-clustering of families in the urban cluster (Köster et al., 2022). Taken together, these results show that variation in communicative interactions did not – at least not primarily – originate from the languages that were spoken, but was likely due to norms, values and beliefs prevalent in the respective cultural settings. Below we discuss in more detail how such norms, values and beliefs can explain the variation we found.

Families from Brazil and Ecuador had longer periods of silences and produced fewer gestures. This mirrors results by Cristia (2022) who synthesized 29 studies on naturalistic language input and found that infants growing up in rural settings heard less child-directed speech compared to children growing up in urban settings. For the Kichwa community in Ecuador, Sánchez-Parga (2010) reports a norm that meals are supposed to be taken in

577 silence. In our sample, such norms seemed to have influenced mothers' communication the

578 most: there was less talk by mothers in Ecuador compared to the other settings (except

579 Brazil), while the amount of talk by fathers and children was relatively similar. However,

580 given that all family members talked, it is worth pointing out that this norm – at least in

581 the present study – mainly had an attenuating effect.

582   Children communicated in very similar ways across settings: they mostly made

583 assertions and rarely asked questions or used imperatives. Parents' communication in the

584 different settings were also very similar in that they mostly made assertions, asked

585 relatively few questions and hardly used any imperatives. Notably, the rate of imperatives

586 was substantially higher in the rural settings in Brazil and Ecuador [see also]. For rural

587 Brazil, Köster, Cavalcante, Vera Cruz de Carvalho, Dôgo Resende, and Kärtner (2016)

588 reported that mothers assigned tasks to their children in a more assertive way compared to

589 mothers from urban Germany (see also Keller et al., 2004 for similar findings from rural

590 Costa Rica). Furthermore, when Köster et al. (2022) coded teaching behavior in the same

591 samples, they found that a higher rate of parents in Brazil and Ecuador prompted their

592 children to do something. Finally, in a study on norm enforcement, children living in rural

593 settings themselves used more imperatives than norm-protest when reacting to a peer's

594 perceived norm violation (Kanngiesser et al., 2022). Thus, the higher rate of imperatives

595 likely reflects cultural norms and beliefs about how children should behave and how they

596 learn (Keller, 2007).

Limitations

598   We see the mealtime setting in which we investigated communicative interactions

599 among family members as a strength of the current study, but acknowledge that it comes

600 with important limitations. The constellation of mother, father and one child is probably

601 more representative for the urban contexts of Argentina, Germany and Japan than the

602 rural settings. Thus, it would be interesting to see if and how our observed patterns are

attenuated when more people (especially more children and extended family members) take part in the meal. Based on our current findings, we would anticipate similar rates of change across cultural settings. For example, we would expect that the presence of a second child would lower the rate of talk addressed to the other child in a similar way in all cultural settings. Of course, this prediction – as well as all our results – can only generalize to cultural settings in which the interaction format of joint mealtimes exists.

Furthermore, our sample was a convenience sample in that we relied on established contacts and collaborations to recruit families in different settings. As such, the grouping into rural and urban contexts is confounded with the normative belief systems of particular regions. Thus, we do not think that living in a rural setting per se affects communicative interactions in a systematic way but the specific cultural norms and practices associated with rural subsistence in these settings produced the patterns we observed. Future work should combine our quantitative approach with a qualitative assessment of the local norms surrounding communication and mealtime to better understand the link between norms, values and beliefs and communicative behavior.

Finally, we did not obtain a measure of children's language abilities. As such, we can only speculate to what extent the different interaction patterns directly affected children's language learning. Obtaining such measures would be a valuable extension of our work.

Conclusions

Our findings offer important insights into the variable and constant aspects of children's language learning environments across diverse cultural settings. For all aspects of communication we investigated in the current study, a common pattern emerged across cultural settings suggesting that children can rely on similar information sources and learning processes. This common pattern was attenuated in some of the settings in a way that likely reflected particular local norms, values, beliefs and ecologies. This exemplifies

the importance of quantitative cross-cultural research for theory building in language

acquisition.

References

Akhtar, N. (2002). Relevance and early word learning. *Journal of Child Language*, *29*(3), 677–686.

Altınkamış, N. F., Kern, S., & Sofu, H. (2014). When context matters more than language: Verb or noun in french and turkish caregiver speech. *First Language*, *34*(6), 537–550.

Amir, D., & McAuliffe, K. (2020). Cross-cultural, developmental psychology: Integrating approaches and key insights. *Evolution and Human Behavior*, *41*(5), 430–444.

Bang, J. Y., Bohn, M., Ramirez, J., Marchman, V. A., & Fernald, A. (2022). Spanish-speaking caregivers' use of referential labels with toddlers is a better predictor of later vocabulary than their use of referential gestures. *PsyArXiv*.

Bang, J. Y., Mora, A., Munévar, M., Fernald, A., & Marchman, V. A. (2022). *Time to talk: Multiple sources of variability in caregiver verbal engagement during everyday activities in english-and spanish-speaking families in the US*.

Barbaro, K. de, & Fausey, C. M. (2022). Ten lessons about infants' everyday experiences. *Current Directions in Psychological Science*, *31*(1), 28–33.

Barrett, H. C. (2020). Towards a cognitive science of the human: Cross-cultural approaches and their urgency. *Trends in Cognitive Sciences*, *24*(8), 620–638.

Beals, D. E. (1993). Explanatory talk in low-income families' mealtime conversations. *Applied Psycholinguistics*, *14*(4), 489–513.

Beals, D. E. (1997). Sources of support for learning words in conversation: Evidence from mealtimes. *Journal of Child Language*, *24*(3), 673–694.

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*(1), e12724.

Blum-Kulka, S. (2012). *Dinner talk: Cultural patterns of sociability and socialization in family discourse*. Routledge.

Bohn, M., & Frank, M. C. (2019). The pervasive role of pragmatics in early language.

657   *Annual Review of Developmental Psychology*, *1*(1), 223–249.

658   Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development*

659   *Perspectives*, *12*(2), 104–108.

660   Bohn, M., Le, K. N., Peloquin, B., Köymen, B., & Frank, M. C. (2021). Children's

661   interpretation of ambiguous pronouns based on prior discourse. *Developmental Science*,

662   *24*(3), e13049.

663   Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children

664   integrate information sources to infer the meaning of words. *Nature Human Behaviour*,

665   *5*(8), 1046–1054.

666   Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and

667   variability in children's word learning across languages. *Open Mind*, *3*, 52–67.

668   Broesch, T., Carolan, P. L., Cebioğlu, S., Rueden, C. von, Boyette, A., Moya, C., . . .

669   Kline, M. A. (2021). Opportunities for interaction: Natural observations of children's

670   social behavior in five societies. *Human Nature.*

671   Broesch, T., Lew-Levy, S., Kärtner, J., Kanngiesser, P., & Kline, M. (2022). A roadmap to

672   doing culturally grounded developmental science. *Review of Philosophy and Psychology*,

673   1–23.

674   Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. In N. J.

675   Enfield, P. Kockelman, & J. Sidnell (Eds.), *Cambridge handbook of linguistic*

676   *anthropology* (pp. 187–226). Cambridge University Press.

677   Bruner, J. (1983). *Child's talk: Learning to use language.* New York: Norton.

678   Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Migdalek, M., et

679   al.others. (2020). *A cross-cultural examination of young children's everyday language*

680   *experiences.*

681   Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan.

682   *Journal of Statistical Software*, *80*(1), 1–28.

683   Casillas, M. (2022). Learning language in vivo. *Child Development Perspectives.*

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

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a papuan community. *Journal of Child Language*, *48*(4), 792–814.

Choi, S. (2000). Caregiver input in english and korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language*, *27*(1), 69–96.

Clark, E. V. (2009). *First language acquisition.* Cambridge: Cambridge University Press.

Cristia, A. (2022). A systematic review suggests marked differences in the prevalence of infant-directed vocalization across groups of populations. *Developmental Science*, e13265.

Cristia, A., Dupoux, E., Gurven, M., & Stieglitz, J. (2019). Child-directed speech is infrequent in a forager-farmer population: A time allocation study. *Child Development*, *90*(3), 759–773.

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., . . . Bergelson, E. (2021). A thorough evaluation of the language environment analysis (LENA) system. *Behavior Research Methods*, *53*(2), 467–486.

De León, L. (2011). Language socialization and multiparty participation frameworks. In A. Duranti, E. Ochs, & B. B. Schieffelin (Eds.), *The handbook of language socialization* (pp. 81–111). Malden, MA: Wiley-Blackwell.

Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turn-taking and vocabulary growth in early language development. *Child Development*, *92*(2), 609–625.

Fjellström, C. (2004). Mealtime and meal patterns from a cultural perspective. *Scandinavian Journal of Nutrition*, *48*(4), 161–164.

Gaskins, S. (2006). Cultural perspectives on InfantCaregiver interaction. In N. J. Enfield & S. Levinson (Eds.), *Roots of human sociality: Culture, cognition and interaction* (pp. 279–298). Oxford, UK: Berg.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models.* Cambridge university press.

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531.

Greenwood, C. R., Thiemann-Bourque, K., Walker, D., Buzhardt, J., & Gilkerson, J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly*, *32*(2), 83–92.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young american children.* Paul H Brookes Publishing.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *28*(1), 100–108.

Hewlett, B. S., Lamb, M. E., Shannon, D., Leyendecker, B., & Schölmerich, A. (1998). Culture and early infancy among central african foragers and farmers. *Developmental Psychology*, *34*(4), 653.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236.

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, *98*, 1–21.

Kachergis, G., Marchman, V. A., & Frank, M. C. (2022). Toward a "standard model" of early language learning. *Current Directions in Psychological Science*, *31*(1), 20–27.

Kanngiesser, P., Schäfer, M., Herrmann, E., Zeidler, H., Haun, D., & Tomasello, M. (2022). Children across societies enforce conventional norms but in culturally variable ways. *Proceedings of the National Academy of Sciences*, *119*(1), e2112521118.

Kärtner, J., Schuhmacher, N., & Torréns, M. G. (2020). Culture and early social-cognitive development. *Progress in Brain Research*, *254*, 225–246.

Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and visualize the results of multivariate data analyses.* Retrieved from https://CRAN.R-project.org/package=factoextra

Keller, H. (2007). *Cultures of infancy.* Lawrence Erlbaum Associates.

Keller, H., Yovsi, R., Borke, J., Kärtner, J., Jensen, H., & Papaligoura, Z. (2004). Developmental consequences of early parenting experiences: Self-recognition and self-regulation in three cultural communities. *Child Development, 75*(6), 1745–1760.

Kidd, E., & Garcia, R. (2022). How diverse is child language acquisition research? *First Language*, 01427237211066405.

Köster, M., Cavalcante, L., Vera Cruz de Carvalho, R., Dôgo Resende, B., & Kärtner, J. (2016). Cultural influences on toddlers' prosocial behavior: How maternal task assignment relates to helping others. *Child Development, 87*(6), 1727–1738.

Köster, M., Torréns, M. G., Kärtner, J., Itakura, S., Cavalcante, L., & Kanngiesser, P. (2022). Parental teaching behavior in diverse cultural contexts. *Evolution and Human Behavior.*

Lavechin, M., Bousbib, R., Bredin, H., Dupoux, E., & Cristia, A. (2020). An open-source voice type classifier for child-centered daylong recordings. *arXiv Preprint arXiv:2005.12656.*

Levinson, S. C., & Holler, J. (2014). The origin of human multi-modal communication. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1651), 20130302.

Lieven, E. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition* (pp. 56–73). New York: Cambridge University Press.

Loukatou, G., Scaff, C., Demuth, K., Cristia, A., & Havron, N. (2021). Child-directed and overheard input from different speakers in two distinct cultures. *Journal of Child Language*, 1–20.

Masek, L. R., Ramirez, A. G., McMillan, B. T., Hirsh-Pasek, K., & Golinkoff, R. M.

  (2021). Beyond counting words: A paradigm shift for the study of language acquisition.

  Child Development Perspectives, *15*(4), 274–280.

McElreath, R. (2018). *Statistical rethinking: A bayesian course with examples in r and

  stan.* Chapman; Hall/CRC.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the

  interaction of online referent selection and slow associative learning. *Psychological

  Review*, *119*(4), 831.

Meylan, S. C., & Bergelson, E. (2022). Learning through processing: Toward an integrated

  approach to early word learning. *Annual Review of Linguistics*, *8*(1), 77–99.

  https://doi.org/10.1146/annurev-linguistics-031220-011146

Nielsen, M., Haun, D., Kärtner, J., & Legare, C. H. (2017). The persistent sampling bias in

  developmental psychology: A call to action. *Journal of Experimental Child Psychology*,

  162, 31–38.

Ochs, E., & Shohet, M. (2006). The cultural structuring of mealtime socialization. *New

  Directions for Child and Adolescent Development*, *2006*(111), 35–49.

Rogoff, B. et al. (2003). *The cultural nature of human development.* Oxford University

  Press.

Rogoff, B., Paradise, R., Arauz, R. M., Correa-Chávez, M., & Angelillo, C. (2003).

  Firsthand learning through intent participation. *Annual Review of Psychology*, *54*(1),

  175–203.

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L.,

  & Gabrieli, J. D. (2018). Beyond the 30-million-word gap: Children's conversational

  exposure is associated with language-related brain function. *Psychological Science*,

  29(5), 700–710.

Rowe, M. L., & Weisleder, A. (2020). Language development in context. *Annual Review of

  Developmental Psychology*, *2*, 201–223.

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663–12668.

Sánchez-Parga, J. (2010). *El oficio de antropólogo. Crítica de la razón (inter) cultural.* Ediciones Abya-Yala, Quito.

Shneidman, L., & Goldin-Meadow, S. (2012). Language input and acquisition in a mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673.

Shneidman, L., & Woodward, A. L. (2016). Are child-directed interactions the cradle of social learning? *Psychological Bulletin*, *142*(1), 1.

Snow, C. E., & Beals, D. E. (2006). Mealtime talk that supports literacy development. *New Directions for Child and Adolescent Development*, *2006*(111), 51–66.

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS One*, *8*(11), e80646.

Sperry, D. E., Sperry, L. L., & Miller, P. J. (2019). Reexamining the verbal environments of children from different socioeconomic backgrounds. *Child Development*, *90*(4), 1303–1318.

Swingley, D., & Humphrey, C. (2018). Quantitative linguistic predictors of infants' learning of specific english words. *Child Development*, *89*(4), 1247–1267.

Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child Development*, *90*(6), 2135–2152.

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of english, italian, and mandarin. *Journal of Child Language*, *24*(3), 535–565.

Tomasello, M. (2009). *Constructing a language.* Cambridge, MA: Harvard University Press.

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes

based on early language production and socioeconomic factors. *Child Development*,

65(2), 606–621.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language

experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11),

2143–2152.

Wittenburg, P., Brugman, H., Russel, A., Klassman, A., & Sloetjes, H. (2006). ELAN: A

professional framework for multimodality research. *Proceedings of LREC 2006, Fifth*

International Conference on Language Resources and Evaluation. Retrieved from

https://archive.mpi.nl/tla/elan