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[5,6,*]

[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] Freie Universität Berlin, Faculty of Education and Psychology, Berlin, Germany

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

[*] joint senior author

Author Note

tbd...

The authors made the following contributions. Manuel Bohn: Conceptualization, Methodology, Formal Analysis, Visualization, Writing – original draft, Writing – review & editing; Wilson Filipe da Silva Vieira: Investigation, Writing – review & editing; Marta Giner Torréns: Investigation, Writing – review & editing; Joscha Kärtner: Investigation, Writing – review & editing; Shoji Itakura: Investigation, Writing – review & editing; Lilia Cavalcante: Investigation, Writing – review & editing; Daniel Haun: Investigation, Writing – review & editing; Moritz Köster: Conceptualization, Methodology, Investigation, Writing – review & editing; Patricia Kanngiesser: Conceptualization, Methodology, Investigation, Writing – review & editing.

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

Abstract

Children all over the world learn language, yet, the context in which they do so varies substantially. This variation needs to be systematically quantified in order for theories to accommodate it. The present study compared communicative interactions between parents and their children (N = 99 families) during mealtime across five cultural settings (Brazil, Ecuador, Argentina, Germany, and Japan). In a comparable setup, we coded the amount of talk and gestures as well as their conversational embedding (interlocutors, speech acts, themes). Results showed a comparable pattern of communication across cultural settings, suggesting that children can rely on similar information sources and learning processes. This common pattern was attenuated in ways that reflected particular local norms, values, and beliefs. These findings are the basis for advancing theories of language learning.

Keywords: Language acquisition, Communication, Gesture, Cross-cultural psychology, Parent-child interaction

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). This simple fact poses a serious problem for theoretical accounts of exactly how children accomplish this feat, because social interactions between adults and children are structured by norms, values, and beliefs that vary tremendously across cultures. As a consequence, the social interactions and learning conditions are very different for children growing up in different cultural settings. Theoretical accounts of language learning therefore need to be able to explain how children can reach the same outcome (fluency in their local language) under very different learning conditions (Cristia, 2022; Kidd & Garcia, 2022; Rowe & Weisleder, 2020). An important pre-requisite for solid theorizing about the processes underlying language learning is, therefor, to document variation in learning conditions across cultures. In this paper, we contribute to this effort by reporting on cross-cultural variation in parent-child communicative interactions in a comparable setting: meals involving parents and their 2-year-old child.

A substantial portion of research on language acquisition in recent decades has focused on variation in language input, in particular the number of words children hear in naturalistic settings. This line of work has been sparked by research showing that children who receive more input – in particular 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 is thought to reflect more opportunities for children to learn word-object mappings and build a larger vocabulary (Jones & Rowland, 2017; Kachergis, Marchman, & Frank, 2022; McMurray, Horst, & Samuelson, 2012). This line of

research has gained even more momentum with the introduction of daylong audio recording devices and automized coding algorithms (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 has taken 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 other developmental research (Amir & McAuliffe, 2020; Nielsen, Haun, Kärtner, & Legare, 2017), this line of work has largely been conducted in western affluent settings. As a consequence, it is unclear whether theoretical proposals that built on these results generalize to other cultural settings. Fortunately, the number of studies investigating language input in different cultural settings is constantly increasing (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). This research has already produced very important results: it has shown that there is huge variation in how much direct input children receive in different cultural contexts (Cristia, 2022; see also Sperry, Sperry, & Miller, 2019) while still reaching major developmental milestones at similar ages (Brown & Gaskins, 2014; Casillas, Brown, & Levinson, 2020). Thus, even though the quantity of input plays an important role, theories and models of language learning need to include learning processes that compensate for variation in input (Jones & Rowland, 2017; Kachergis et al., 2022; Meylan & Bergelson, 2022).

These compensatory learning processes are often through to leverage the structure of the social interactions in which language is being 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 leverage contextual information (e.g., Bohn & Frank, 2019;

Tomasello, 2009). Social interactions, especially routines, follow predictable patterns that make it easier for children to infer what speakers are communicating about (Barbaro & Fausey, 2022; Lieven, 1994; Masek, Ramirez, McMillan, Hirsh-Pasek, & Golinkoff, 2021). To illustrate this point, Roy, Frank, DeCamp, Miller, and Roy (2015) found that words are easier to learn when they are primarily used in a distinct spatial and temporal context. The common ground built up during the course of an interaction has a similar effect in that it provides information about the speaker's intention independent of the words that are being used (Bohn & Köymen, 2018; Bohn, Tessler, Merrick, & Frank, 2021). For example, Bohn, Le, Peloquin, Köymen, and Frank (2021) showed that children identify the referent of an ambiguous word by inferring the topic of an ongoing conversation (see also Akhtar, 2002). These findings help to explain why the amount of conversational turn-taking in parent child interactions predicts child language outcomes (Donnelly & Kidd, 2021; Romeo et al., 2018). Turn-taking likely reflects continuous, structured conversations that constitute information-rich learning opportunities.

To our knowledge, there is very little cross-cultural research that systematically compared the structure of communicative interactions between adults and children. Ethnographic descriptions offer important insights into individual cultural settings (see e.g., De León, 2011; Gaskins, 2006) but they do not allow for a direct quantitative comparison. Yet, the example of direct language input discussed above illustrates the importance of direct quantitative comparisons for theory building.

One of the challenges for cross-cultural work on the structure of communicative interactions lies in selecting an appropriate context to study them. Work in western-affluent settings has shown that the amount of language input to children varies substantially across routine activities. For example, Soderstrom and Wittebolle (2013) found that adults speak most during book reading and structured playtime (see also Tamis-LeMonda, Custode, Kuchirko, Escobar, & Lo, 2019). These contexts, however, are very specific to western, affluent settings and less frequent or absent in other cultural

contexts. A particularly promising context for cross-cultural research appears to be 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; Ochs & Shohet, 2006). Furthermore, mealtime has been a fruitful context to study caregiver-child communication (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 in that we a) asked families to record in their home, b) recruited families with a single child between 2 and 3 years of age and c) focused on 10 minute-long episodes during which all three family members were present. Even though one-child families 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 families living in five different cultural 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 research questions that revolve around the quantity of talk and gestures as well as their conversational embedding (interlocutors, speech acts, themes). In the final section of the analysis, we asked how families clustered based on the coded dimensions. The selection of cultural settings offers 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. 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. Excluded because of this were 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 originally obtained for a study by Köster et al. (2022) to study parental teaching behavior across cultures. We refer to this earlier work for a detailed description of each cultural setting. In the following we thus provide only a short overview.

Argentina. Families lived in the metropolitan area of Buenos Aires, Argentina comprising around 15.2 million people. They were recruited via personal contacts of the local experimenter. The family language was Rioplatense Spanish. Compensation included small toys for children and USD 10 for parents. Most parents had completed a university degree (mothers: 74%; fathers: 52%) and engaged in paid professional labor (mothers:

87%; fathers: 78%). The majority of children (91%) either attended kindergarten or were looked after by a nanny or a family member other than the parents.

Brazil. Families lived in villages of around 50 - 300 families in the Amazon region near Apeú, approximately 1.5 hours east of Belém, the capital of the state of Pará. They were recruited with the help of a local health post [M&P: what is a health post?]. The family language was Brazilian Portuguese. Compensation included small toys for children and a certificate of participation for parents. Most parents had completed secondary school (~12 years of schooling, mothers: 50%; fathers: 56%). Mothers worked mainly as housewives (83%) while fathers engaged in paid labor (100%). Some families engaged in traditional subsistence activities such as tapioca farming, livestock breeding, or acaí and fruit harvesting. In line with employment status, the majority of children were looked after by their mothers.

Ecuador. Families lived in villages with 800-5,000 inhabitants located within 1 hour (by car) of the city of Cotacachi in the Imbabura province. They were recruited via personal contacts mediated by the community president. The family language was Ecuadorian Spanish with elements of Quechua. Compensation included food (e.g., rice or oat) and USD 4. Most parents had completed primary school (~10 years of schooling, mothers: 50%; fathers: 56%). Mothers worked mainly as housewives (59%) while fathers engaged in paid labor (77%). [M&P: Info on external child care and, if applicable, subsistence farming activities or so].

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, 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

Families were visited twice. On the first visit, an experimenter (familiar with the local language) instructed parents how to use the video camera and what to record. Families were encouraged to record two instance of the meal they most commonly share 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 completed the two recordings already. In the end, parents provided socio-demographic information and mothers completed an interview (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 the in-depth coding and excluded all families for which we did not find such a section (see above for number of excluded families).

Videos were coded using ELAN (Wittenburg, Brugman, Russel, Klassman, & Sloetjes, 2006) version 6.4. The primary coder was either a native (Germany, Japan, Brazil) or highly fluent (Argentina, Ecuador) speaker of the local language. In Ecuador, sections containing Quechua were first translated into Spanish by a native speaker and then coded by the primary coder.

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. Utterances were defined as a section of continuous talk by one person. If speakers paused for more than 2 seconds, two utterances were coded 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 another. This applies for example to sequences of questions and answers but also to sequences in which the content of an utterance is directly related to the content of the previous utterance. Please note that such themes were coded locally and were not the same as topics. For example, if father and child exchanged four utterances about the child's day in the kindergarten this was coded as one theme. If the same topic (day at the

kindergarten) came up later again, this was coded as a separate theme. Each utterance within a theme was counted as a `round` to capture the sequence and length of a theme. Thus, each utterance was assigned a number for the theme and a number for the round within theme. Themes could have interjections of one or two utterances. After more than two interjections we coded a new theme. For example, if father and child talked about food and the mother made an unrelated comment in between, the mother's comment would be coded as a separate theme while the other theme continued around it:

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

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

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

Child: "No, bread (t1, r3)

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

Speech acts. Each utterance was coded as either being a `question`, `assertion` or `imperative`. Imperatives were only coded if the the utterance was grammatically structured as an imperative. For example "Pass me the salt!" was coded as an imperative while "You should give me the salt." was not.

Referential gestures. We also coded the frequency of two types of referential gestures for each individual. `Points` were coded when someone indicated an object, location or person in the environment, either using a finger (often index finger), the head or an object (e.g., cutlery). Reaches and hold-outs were not coded as points. `Iconic` 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). 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 as the previous utterance or belonged to a new theme[1]. Once again, agreement between coders was high (agreement = 87%, $\kappa = 0.74$).

Analysis and Results

For each of the research questions 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 comparison we visualized the predictions of the winning model and interpret them based

[1] The alternative would have been to ask whether the two coders assign the same theme to an utterance. This would have meant, however, that, if coders disagreed once, all subsequent codes would have disagreed as well.

on their posterior means and 95% Credible Intervals (CrI). We used default priors built into `brms` for all parameters.

How much silience is there?

First, we ask how much time families spent talking as opposed to being silent and how this varied across cultural settings. The dependent variable in this case was the total lengths of all sections coded as silence for each family (modeled as a normal distribution). We compared a null model including only an overall intercept (`silence ~ 1`) to a model including cultural setting (`silence ~ 1 + (1|setting)`).

The model comparison clearly favored the model including cultural setting (WAIC = 338.84, se = 14.93, weight = 1.00) over the null model (WAIC = 362.36, se = 14.97, weight = 0.00). The model predicted an average of 4.95 [95%CrI = 3.80 - 6.07] minutes of silence across cultural settings. Variation across settings was such that families in Ecuador and Brazil had longer sections of silence compared to Argentina and Germany, with Japan falling in the middle (see Figure 1A).

Who is being talked to

Next, we asked who talk was directed to, that is, how much "input" each family member received. The dependent variable was the total lengths of utterances directed at each individual within family. This variable was right-skewed and we therefore modeled it as a skewed normal distribution. Given that the analysis above showed that the amount of overall talk differed across cultural settings, the null model already included a random effect for setting (`input ~ 1 + (1|setting) + (1|family)`). We compared it to two alternative models, one assuming that input additionally differs across recipients (`input ~ recipient + (1|setting) + (1|family)`) and one assuming that this effect in turn 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 not assuming 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 saw that, across settings, more talk was directed at children compared to the two parents with fathers being talked to the least (see Figure 1B).

Who is talking

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 coming from each individual within which was also right-skewed and which we also 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 differs across speakers (`input ~ recipient + (1|setting) + (1|family)`), the second assumed that this effect interacts with setting (`input ~ 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 analyse 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 and 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 varies 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

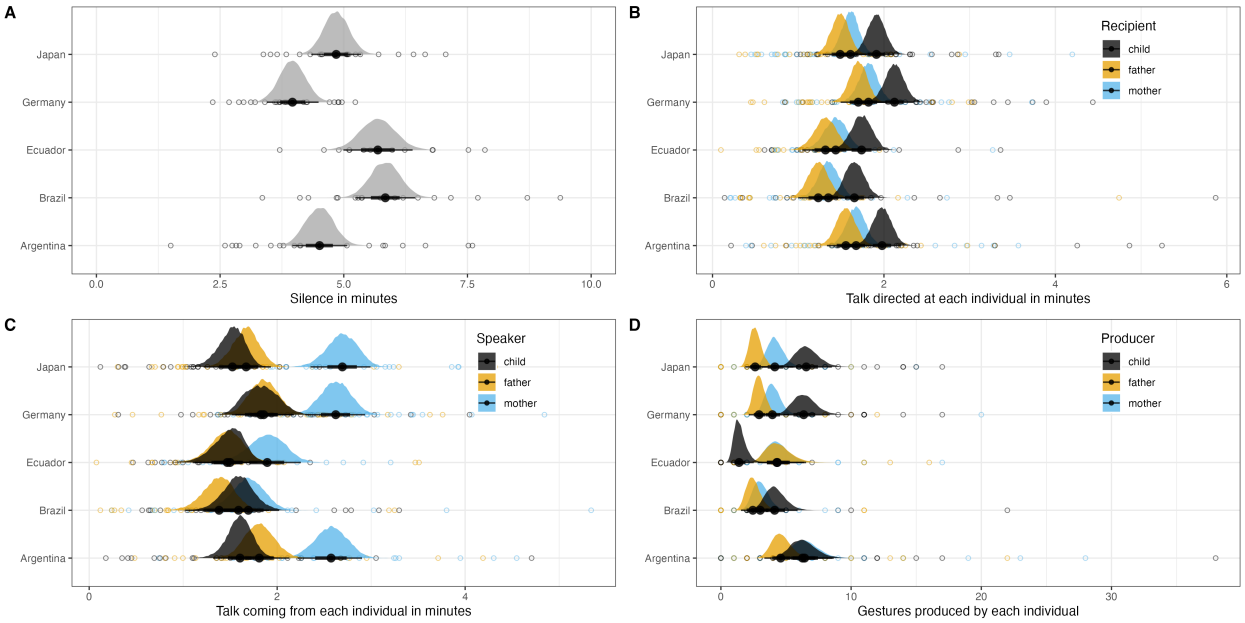

Figure 1. A: Silence across cultural settings. B: Talk directed at the different individuals. C: Time spent talking by the different individuals. D: Number of gestures (points and iconic gestures combined) produced by each individual. in B -D: color denotes the individual. Distributions show the predicted values based on the respective model with solid points and error bars showing the mean with 66% and 95% CrI. Light points show the the aggregated data for each familiy and – whenever applicable – individual.

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

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

Who uses which speech acts?

As the next step, we analysed 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 used varies differently 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 are involved in a theme. To be involved in a theme, an individual had to produce at least one utterance. Please note that themes could have only one

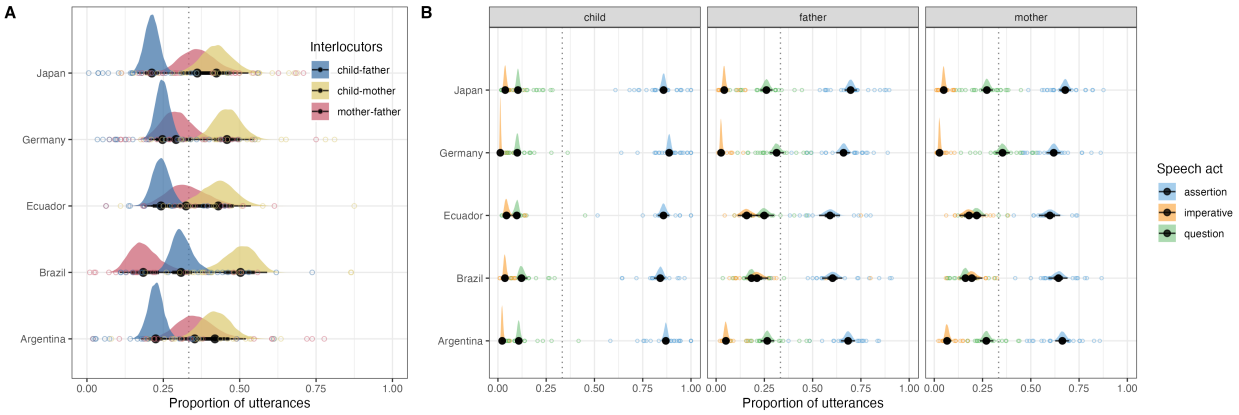

Figure 2. A: Proportion of utterances that are exchanged by a pair of interlocutors. Color shows the interlocutors involved in the utterance regardelss of direction (i.e., identity of speaker and listener). B:Proportion of utterances that belong to a certain class of speech acts. Facets show different 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 familiy.

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 big 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, were 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 how long it was (i.e., the maximum round) and who the main interlocutors were. 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.

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

The model comparison clearly favored the interaction model assuming that the length of a theme varied differently with the interlocutors involved within the cultural settings (WAIC = 11657.48, se = 106.30, weight = 1.00; second best model (without interaction): WAIC = 11671.33, se = 106.59, weight = 0.00). The average predicted length of a theme across interlocutors and settings was 5.71 rounds [95%CrI = 3.95 - 8.35]. Figure 3C indicates a variable pattern across cultural settings. In Japan, themes did not differ in length with the interlocutors involved. In the other settings, conversations between mother and father were shorter compared to conversations between one of the parents and the child. This pattern was less pronounced in Ecuador compared to Germany, Brazil and Argentina. Overall, themes lasted slightly longer in Brazil compared to the other settings.

Family level clustering

In this final analysis, we took a more holistic look at the data and tried to identify patterns across the communicative dimensions analysed above. That is, we asked if there are clusters within our sample that represent different communicative profiles. This allowed us to see a) if families clustered based on cultural settings and b) how the different cultural settings clustered with each other. To construct the data set for this analysis, we computed

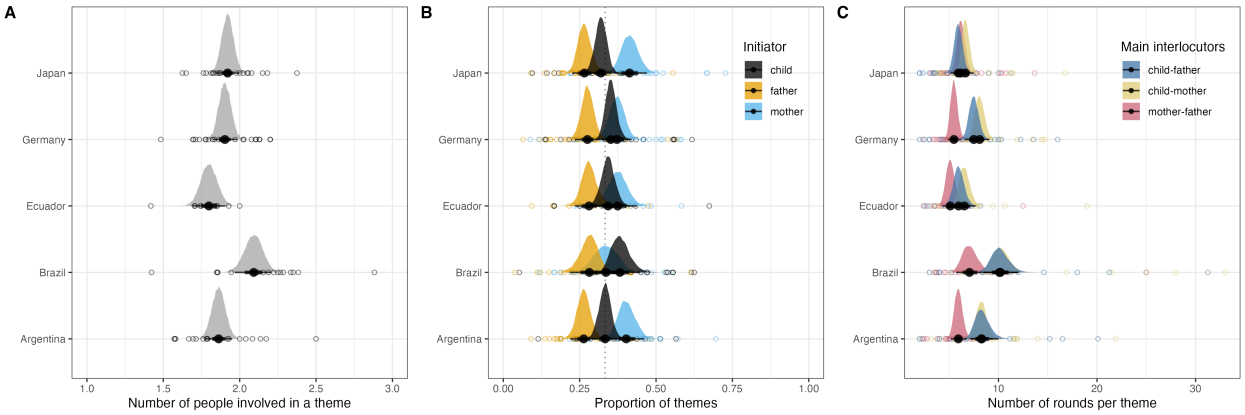

Figure 3. A: Average number of people involved in a theme. B: Proportion of themes as a function of who initiated them. Color shows the initiator. C: Number of rounds per theme depending on the interlocutors involved. Color shows the interlocutors who exchanged the most utterances within a given theme. 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 familiy.

the following dimensions for each family: the amount of `Silence`, the proportion of utterances coming from each individual (`Father speaker`, `Mother speaker`, and `Child 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 be estimated based on less data (resulting in a more noisy estimate), we decided for this 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 (gold), included mainly families from Argentina, Germany and Japan. Within the cluster, there was no further clustering of families by cultural setting. The second cluster (blue), 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 coming from the different individuals) there were some notable associations: more silence went along 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 mealtime in five cultural settings. Each family comprised father, mother and one child and was recorded for 10 minutes. We found that families from Ecuador and Brazil communicated less overall compared to families from Argentina and Germany, with Japan falling somewhere in the middle. Across settings, there was a common pattern in how talk was distributed across

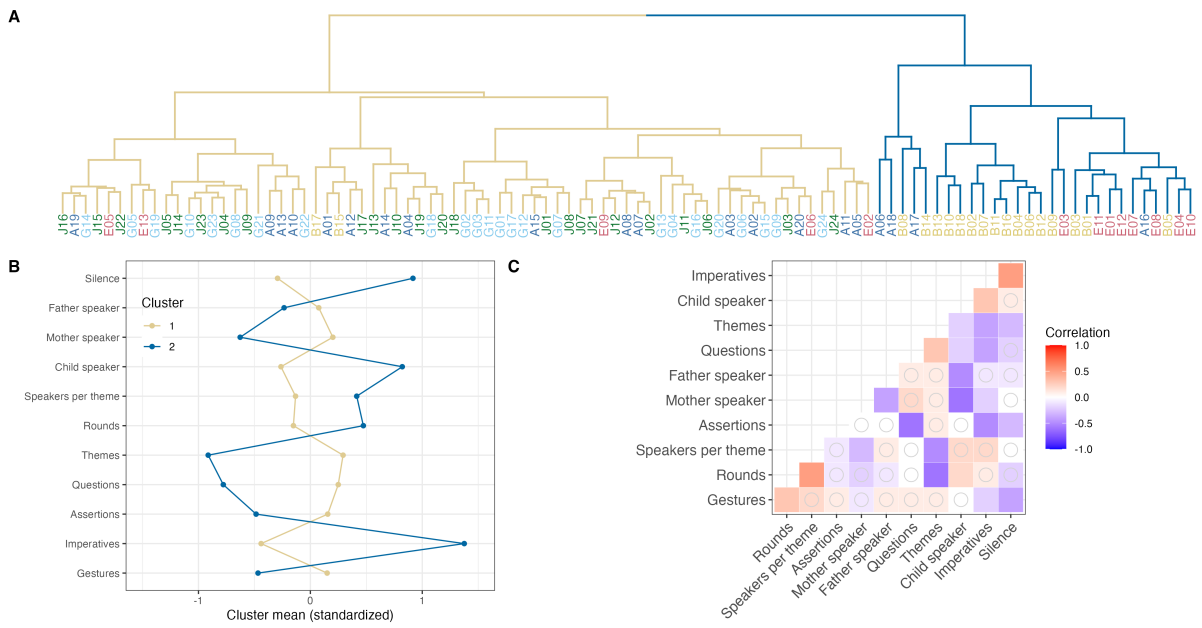

Figure 4. A: Dendogramm visualizing the similarity between families based on a cluster analysis assuming two clusters. Line colors show the two clusters, color of letters for familiy corresponds to the different cultural settings The first letter of the family name denotes the cultural setting (e.g., J = Japan). B: Mean values for the two clusters for each (standardized) dimension on which the cluster analyis 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 crosses show correlations with p-values < 0.05.

family members: mothers talked the most and children were addressed the most. Assertions was 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. 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 a urban-rural split, with families from urban settings (Argentina, Germany, Japan) being more similar to each other compared to families from rural settings (Brazil, Ecuador). These results provide an important first step

towards understanding the variation in communicative contexts in which children learn language through quantitative comparison.

Our findings echo the way Barrett (2020) summarized much of cross-cultural research in the last two decades: *variation on a theme*. For every aspect of communicative interaction we investigated there was a dominant pattern which described behavior in most of the cultural settings and that was attenuated in one or two settings. Attenuation meant that the predicted means for some of the settings were shifted while the distributions of families were largely overlapping. For example, on average, the number of people involved in a theme was around 1.8, with slightly more people involved in Brazil (~ 2.1), yet, the minimum family average in Brazil was 1.40 and the maximum for Ecuador (lowest predicted average) was 2. Or, mothers talked the most in all settings but the difference compared to father an child was less pronounced in Ecuador and Brazil. Thus, we may tentatively say that this overlap in communicative patterns allows for similar learning processes – in particular those that leverage the structure of the communicative context (Casillas et al., 2020; Rogoff et al., 2003; Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016) – to be recruited across settings.

The overall pattern – or *theme* – can be summarized as being child-centered. Across cultural settings, most talk was directed towards the child and themes had more conversational turns (i.e. number of rounds) when the child was involved. The latter finding corresponds well with the idea that children's language learning benefits from coherent and structured interactions (Casillas et al., 2020; Rogoff et al., 2003; Shneidman & Goldin-Meadow, 2012; Shneidman & Woodward, 2016). Mothers seemed to be the driving force behind this pattern: they spoke the most, initiated most themes and most of the themes that involved them also involved the child. This aligns with a recent study by Broesch et al. (2021) who described mothers as the primary interactions partners for 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 live in very different geographical regions and speak very different languages, families living in e.g., Argentina were not more similar to each other than they were to families from Germany or Japan. However, the urban/rural split was by no means perfect 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 pattern was found when the same videos were coded for parental teaching behavior (Köster et al., 2022). Taken together, we take these results to show that variation in communicative interactions does not – at least not primarily – originate from the language that is being spoken, but from norms, values and beliefs prevalent in the respective cultural setting. Below we discuss in more detail how such norms, values and beliefs can explain the variation we found.

Families from Brazil and Ecuador were less communicative in that there was more silence and fewer gestures. This mirrors the results reported by Cristia (2022) who synthesized 29 studies on naturalistic language input and found that infants growing up in rural settings hear less child-directed speech compared to children growing up in an urban setting. For rural Ecuador, Parga (2010) reports a norm that says that meals are supposed to be taken in silence and only the highest authority (usually the father) is allowed to speak. Köster et al. (2022) report on an indigenous norm in the rural Brazilian context according to which children are expected to be silent during meals [M&P: do you have a source for that other than your paper?]. In our sample, such norms seemed to have influenced mothers' communication the most: there was less talk by mothers in Brazil and

Ecuador compared to the other settings while the amount of talk by fathers and children was relatively similar. However, when comparing absolute numbers across settings, it is worth pointing out that such norms were not strictly enforced but rather had an attenuating effect. Fathers in all settings were the least likely to initiate a theme even though the cultural norm in Ecuador reserves the right to speak for them.

Children communicated in very similar ways across settings: they mostly made assertions and rarely asked questions or used imperatives. Parents were also very similar in that they mostly made assertions, asked relatively few questions and hardly used any imperatives. Notably, the rate of imperatives was substantially higher in the rural settings in Brazil and Ecuador. For rural Brazil, Köster, Cavalcante, Vera Cruz de Carvalho, Dôgo Resende, and Kärtner (2016, M&P: you cite Keller et al. 2004 for the same point in your paper but that study is in Costa Rica) reported that mothers assigned tasks to their children in a more assertive way compared to mothers from urban Germany. Furthermore, when Köster et al. (2022) coded teaching behavior in the same families included here, they found a higher rate of direct prompts used to get the child to do something in Brazil and Ecuador. Thus, the higher rate of imperatives likely reflects cultural norms and beliefs about how children should behave and how they learn (Keller, 2007).

Limitations

Even though we see the comparable setup in which we studied communicative interactions among family members as a strength of the current study, it comes with important limitations. The specific constellation of mother, father and one child is probably more representative for the urban contexts of Germany and Japan than the rural settings. As a consequence, the majority of children growing up in these contexts might actually be confronted with different interaction patterns. In any case, it would be interesting to see how the patterns we observed would change when more people (especially more children) would be involved. Based on the data at hand, we expect similar rates of change across

cultural settings. For example, we would expect that the presence of a second child lowers the rate of talk addressed to the other child in a similar way in all cultural settings.

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 associated with living in the rural settings studied here produced the patterns we observed.

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

Conclusions

The results reported here offer important insights into the variability in children's language learning environment. For all aspects of communication we investigated, there was a common pattern 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 reflected particular local norms, values and beliefs. This exemplifies the importance of quantitative cross-cultural research for theory building.

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*.

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

Bohn, M., & Köymen, B. (2018). Common ground and development. *Child Development Perspectives*, *12*(2), 104–108.

Bohn, M., Le, K. N., Peloquin, B., Köymen, B., & Frank, M. C. (2021). Children's interpretation of ambiguous pronouns based on prior discourse. *Developmental Science*, *24*(3), e13049.

Bohn, M., Tessler, M. H., Merrick, M., & Frank, M. C. (2021). How young children integrate information sources to infer the meaning of words. *Nature Human Behaviour*, *5*(8), 1046–1054.

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67.

Broesch, T., Carolan, P. L., Cebioğlu, S., Rueden, C. von, Boyette, A., Moya, C., ... Kline, M. A. (2021). Opportunities for interaction: Natural observations of children's social behavior in five societies. *Human Nature*.

Brown, P., & Gaskins, S. (2014). Language acquisition and language socialization. In N. J. Enfield, P. Kockelman, & J. Sidnell (Eds.), *Cambridge handbook of linguistic anthropology* (pp. 187–226). Cambridge University Press.

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

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Migdalek, M., et al.others. (2020). *A cross-cultural examination of young children's everyday language experiences*.

Bürkner, P.-C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software*, *80*(1), 1–28.

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.

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.

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.

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.

Parga, J. S. (2010). Puerilizado y adulterado: Representaciones institucionales de la infancia. *Universitas-XXI, Revista de Ciencias Sociales y Humanas*, (13), 95–130.

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.

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