Immature vocalizations simplify the speech of Tseltal Mayan and US caregivers

Steven L. Elmlinger
Cornell University
sle64@cornell.edu

Michael H. Goldstein
Cornell University
mhg26@cornell.edu

Marisa Casillas
University of Chicago
mcasillas@uchicago.edu

Abstract

What is the function of immature vocalizing in early learning environments? Previous work on infants in the US indicates that prelinguistic vocalizations elicit caregiver speech which is simplified in its linguistic structure. However, there is substantial cross-cultural variation in the extent to which children's vocalizations elicit responses from caregivers. In the current study we ask whether children's vocalizations elicit similar changes in their immediate caregivers' speech structure across two cultural sites with differing perspectives on how to interact with infants and young children. Here we compare Tseltal Mayan and US caregivers' verbal responses to their children's vocalizations. Similar to findings from US dyads, we found that children from the Tseltal community regulate the statistical structure of caregivers' speech simply by vocalizing. Following the *interaction burst hypothesis*, where clusters of child-adult contingent response alternations facilitate learning from limited input, we reveal a stable source of information facilitating language learning within ongoing interaction.

Keywords: language input; parent-child interaction; language statistics; child-directed speech

Introduction

Across several species, adults' contingent responses to their offsprings' immature vocalizations play a key role in communicative development (Carouso-Peck & Goldstein, 2019; Goldstein & Schwade, 2008; Gultekin & Hage, 2018). When these responses are rare, how might they facilitate learning? Recent findings demonstrate that child-directed talk in both US and non-Western communities is organized around children's vocalizations and predicted by routine activities throughout the day, although the style, context, and source of talk is highly variable between communities (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019; Brown, 2011, 2014; Casillas, Brown, & Levinson, 2020). Per the *interaction burst hypothesis*, predictable clusters of interactive language learning opportunities may maximize learning across contexts that vary widely in their child-directed language style (Casillas et al., 2020, 2021). Learning may be maximized when adult speech within vocal turn-taking bouts with children is structurally simplified and thus, easier to learn from.

What role do children play in structuring the learnability of the ambient language? By 9 months, infants can regulate the complexity of their caregivers' speech simply by vocalizing (Elmlinger, Park, Schwade, & Goldstein, 2021; Elmlinger, Schwade, & Goldstein, 2019a, 2019b). The causal flow from infants' vocalizations to changes in adult speech is established. Contingent and non-contingent speech are spoken in the infant-directed speech register, but caregiver linguistic structure is altered immediately after a child vocalizes (Elmlinger et al., 2019b; Fernald, 1989). Infant vocalizations facilitate the production of more simplified talk from their adult caregivers. The lexical and syntactic structure of caregiver speech is simplified in response to infants' vocalizations. Caregivers utter fewer unique word types, fewer words per utterance and higher proportions of utterances which contained only a single word when talk was contingent on vocalizations. Thus responses to babbling reduce the complexity of caregivers' speech in ways that may facilitate infant learning (Schwab & Lew-Williams, 2016).

The extent to which children across multiple linguistic and cultural backgrounds experience a reduction in speech complexity as a result of their vocalizing is unknown. Differences in caregivers' attitudes about child language development and socialization across cultures may predict whether speech within vocal turn-taking with children is simplified. If the simplification effect of contingent speech relies upon the pedagogical attitudes of the speaker, then Tseltal adult caregivers, who are less likely than US adult caregivers to engage young children in child-centric and pedagogical speech interactions (see below), may not show the simplification effect found in US adults. Alternatively, the simplification effect may be independent of pedagogical attitudes and the contingent simplification found in US caregivers may present as a stable feature of language learnability across multiple languages and cultures. To shed light on these possibilities, we focus on comparing speech simplification of Tseltal Mayan and US caregivers.

Ethnographic background

The Tseltal participants live in a rural Mayan community in the mountains of southern Chiapas, Mexico. Most caregivers in the sample are horticulturalists and most children are raised in multi-generational patrilocal family compounds (i.e., near their nuclear family, paternal grandparents, paternal uncles, etc). Tseltal is the primary language spoken at home.

Young children are carried for much of the first year, and are socialized to attend to the social interactions occurring around them rather than expecting to be the center of adult attention. Longitudinal ethnographic research suggests that speech directed to children is often brief, involves three or

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society.*

more participants, and focuses on appropriate actions and responses rather than words and word meanings (Brown, 2014). In a given day, children under age 3;0 hear an average of 3.6 minutes of speech directed at them per hour (Casillas et al., 2020), which may be comparable to averages from other (e.g., US) communities but includes a greater preponderance of directed speech from other children (Bunce et al., under review). As children become more competent language users, they begin to more effectively engage other children and adults as conversational partners (Brown, 2011, 2014).

Infant-directed speech in Tseltal

Tseltal infant-directed speech is not recognizable as such by naïve Western listeners, who cannot effectively distinguish it from adult-directed speech by the same speakers, despite high reported confidence ratings (Soderstrom, Casillas, Gornik, et al., 2021). Ethnographic report suggests that imperatives, repetitive social routines, and immediate turn repetition are the most common types of speech to infants and young children, while questions and pedagogical talk are much less common (Brown, 2011, 2014). Pye's (1986) in-depth analysis of infant-directed speech in K'iche', a related language community, demonstrates that while there *is* a distinct register for talking to infants, the ways in which pitch, phonology, lexical forms, and morphosyntactic choices are modified are language-specific and do not necessarily involve simplification. For example, there were no appreciable differences in MLU in morphemes between adult- and infant-directed speech.

Present study

We first compared the linguistic structure of parental speech as a function of its contingency on children's vocalizations. The length of parents' utterances to children in everyday learning environments may decrease until children learn specific target words within the utterances (Roy, Frank, & Roy, 2009). Isolated words spoken to children are predictive of the words that are most likely to be produced by children later in development (Brent & Siskind, 2001). By investigating the structure of contingent and non-contingent speech across Tseltal and US caregivers, we can better understand the role that children's vocalizations may play in influencing their own communicative development.

Parental speech structure was quantified in terms of three measures. To measure lexical diversity, we counted the number of unique words (types) in parents' talk to children. The influence of lexical diversity differs by timescale. Short clusters of partially-repetitive speech positively predict language learning (Onnis & Edelman, 2019; Schwab & Lew-Williams, 2016). Variability in lexical input over longer timescales promotes language development (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010). We assessed syntactic complexity of parents' speech by determining the mean length of utterances in words (MLUw) (Parker & Brorson, 2005) and the proportion of utterances which contained only a single word.

Because of the limited comparability across samples included in this study, we treated the Tseltal and US measures as two individual case studies. Direct statistical comparisons of contingent speech simplification were not made due to the differences in participants' age and recording context across sites. The central difference of interest is that of cultural context and whether speech structure is altered when organized around children's vocalizations. To the extent that we see similar patterns of speech structure change across cultures, this suggests strong effects, robust against the study's current limitations.

Methods

Participants

10 Tseltal children between 2 and 36 months were recorded in 2015 during their everyday routines in Chiapas, Mexico. Families were recruited via snowball sampling in the community and were given a small cash gift for their participation in the study.

30 US caregiver-infant pairs participated when infants were 5 and 10 months of age (Table 1). We recruited these subjects from birth announcements in advertisements and local newspapers. As a gift for participation in the study, families received a t-shirt or a bib.

Recordings & procedure

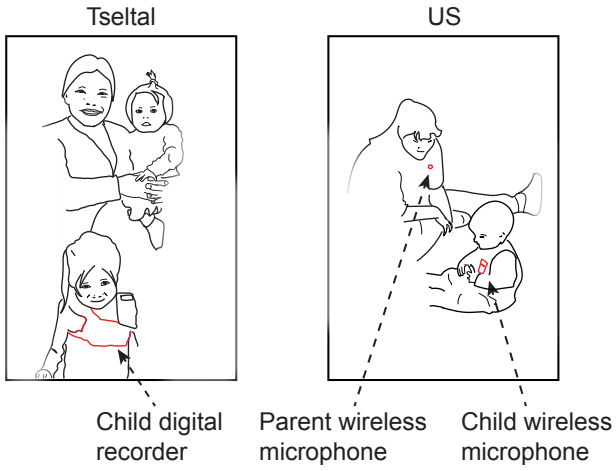

Figure 1: Recording setup across sites.

The Tseltal recordings analyzed here are the same used in Casillas et al. (2020). The recordings were randomly sampled from a total set of 55 to achieve an overall balance in sex, maternal education, and age range between 0 and 36 months (Soderstrom, Casillas, Bergelson, et al., 2021). On the morning of each recording, children donned an elastic vest containing a horizontally stored Olympus WS-832 stereo audio recorder and a small camera on a vertical shoulder strap (images are not analyzed here). Infants too small to comfortably wear both pieces of equipment were outfitted with a onesie

	Dyad (n)	Recording duration (avg)	C voc count (avg)	CG utt count (avg)	C age in months (avg)
TCDS	10	60.00	465.80	149.50	16.10
+CDS	10	60.00	465.80	213.50	16.10
US	60	14.92	38.95	97.33	7.53

Table 1: Recording descriptive statistics. avg = average, C = child, CG = caregiver, TCDS = Tseltal caregivers' target-child-directed speech, +CDS = both Tseltal caregivers' target-child-directed speech and their child-directed speech more generally, utt = utterance, voc = vocalization

shirt that had a horizontal pocket to store the recorder (Figure 1).

Tseltal Children wore the recorder continuously throughout the ∼9 hour recording unless they needed to be bathed or if wearing the equipment during a nap would inhibit their sleep—in this case caregivers were instructed to place the recorder nearby the child. That same evening, the experimenters returned to collect the equipment.

All recordings of the US data took place in a naturalistic environment in a twelve foot by eighteen foot playroom which included a toy box, toys and animal posters. This environment afforded infants the freedom to play and explore around the room as they wished. Three digital cameras were stationed in the room and remote-controlled by experimenters capturing the video recordings. Infants wore overalls which concealed a wireless microphone (Telex FLM-22) paired to a transmitter (Telex USR-100). Before each session, wireless lapel microphones (Telex FLM-22) were affixed to caregivers' shirts. Caregiver microphones were connected to transmitters hidden in a pouch at their waist (Telex USR-100) (Figure 1). Distinct audio channels were utilized in the recording of infants' vocalization and caregiver speech, respectively. See Table 1 for more details of the participants in the study across recording sites.

Each US participant engaged in 15-minute play sessions in the lab. During these sessions, parents were asked to play like they would at home, resulting in unstructured free-play.

Speech transcription

Tseltal parents' speech sample consists of 60 minutes of transcription per recording. 45 of the 60 total minutes were randomly selected 5-minute clips. These speech samples were annotated and transcribed jointly by the visiting Western researcher and a native of the community who knew all the families personally. Annotations included full transcriptions of all hearable speech and to whom the speech was addressed (e.g., to the target child only 'TCDS'; to any child(ren) present 'CDS'; and to adults 'ADS'). 'TCDS' and 'CDS' are examined in the present work. Note that because Tseltal is a mildly polysynthetic language, words typically contain multiple morphemes. The further 15 of 60 total minutes were hand-selected from the remaining, unannotated portions of each recording. Comprehensive review of each audio recording, excluding the original random clips, allowed us to identify the five top one-minute segments of turn-taking between the target child and their interactants, then the five top one-minute segments of target child vocalization from the remaining recording times. The most active interaction captured in those 10 1-minute clips was then expanded a further five minutes, with all additional 15 minutes of clip time per recording fully annotated using the same standards as the random clips. This process resulted in one hour of fully transcribed and annotated recording time from each of the 10 daylong recordings, representing both baseline and high-activity speech periods (i.e., 10 hours of audio in total). Speech in these recordings comes from many speakers; we here focus exclusively on speech from the target child's mother.

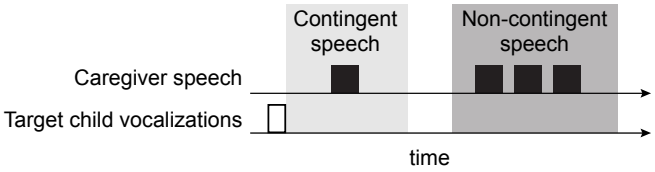

Figure 2: US and Tseltal caregiver utterances were considered contingent when they occured within 2 seconds of the target child's vocalizations. This contingency definition was used for Tseltal TCDS and +CDS analyses.

The speech that US parents produced was completely transcribed. If parents' utterances were separated by silence longer than 2 seconds in duration and/or if their utterance exhibited a terminal pitch contour, they were segmented into separate utterances (Stockman, 2010; Venker et al., 2015). All caregiver utterances were directed to their infant. We excluded caregivers' vocal sound effects and any responses to infant vegetative vocalizations such as coughs, cries, and fusses from the analyses.

When Tseltal and US parents' utterances occurred within 2 seconds of the offset of the target childs' vocalizations, then they were considered contingent utterances (Elmlinger et al., 2019a). Parent utterances which occurred after a 2 second time frame were considered non-contingent (Figure 2). Two seconds was used following previous studies which originally reported on the simplification of contingent speech (Elmlinger et al., 2019b).

Child utterances

The onsets and offsets of all Tseltal infant non-vegetative, communicative vocalizations (i.e., including laughter, fussing, and crying) were annotated, segmented approximately according to breath groups, with some exceptions (e.g.,

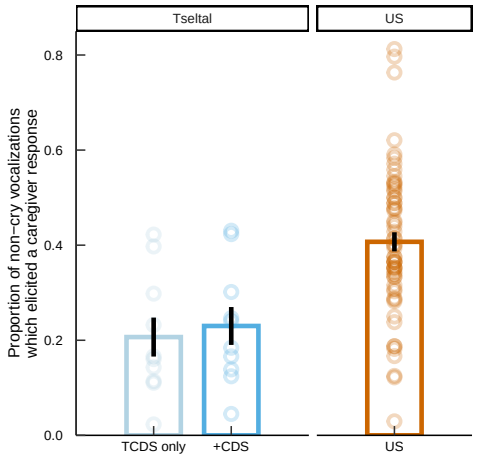

Figure 3: Proportion of child vocalizations which elicited a contingent response. Black lines indicate ± 1 standard error around the mean.

longer bouts of crying). When lexical, vocalizations were transcribed, and were otherwise classified as containing canonical syllables or not, or containing laughter or crying.

US infant non-cry vocalization bout onsets and offsets were annotated in full. Vocalization boundaries were segmented according to breath groups (Oller, 2000; Oller & Lynch, 1992).

Analytic approach

We employed LMMs with the `lme4` package in R (Bates, Mächler, Bolker, & Walker, 2014), to predict linguistic structure from contingency, controlling for target child age, with participants as a random effect. In comparing proportion of child non-cry vocalizations which elicited a response across site and in comparing caregivers' distribution of contingent to non-contingent utterances, we used LMs.

We considered two groups of Tseltal caregiver speech–target-child-directed speech (TCDS), which was caregiver speech directed to the target child being recorded, and all child-directed speech in general (+CDS), which included TCDS and any caregiver speech which was directed at children more generally. Because Tseltal children may treat both TCDS and +CDS as relevant learning cues, here both are considered for changes when they are contingent and not contingent on the target child's non-cry vocalizations.

Results

Levels of contingent responsiveness

Tseltal children's vocalizations elicited their mother's verbal response around a a fifth of the time (M=0.207, SD=0.13). When including all child-directed speech (+CDS) in caregiver responses, vocalizations elicited responses at similar rates (M=0.23, SD=0.126). US infant vocalizations elicited caregiver verbal responses around half of the time (M=0.407, SD=0.156). Non-cry vocalizations of US infants were more likely to elicit a caregiver response than Tseltal children's vocalizations (Estimate=0.177, t=3.41, p=0.0011) (Figure 3).

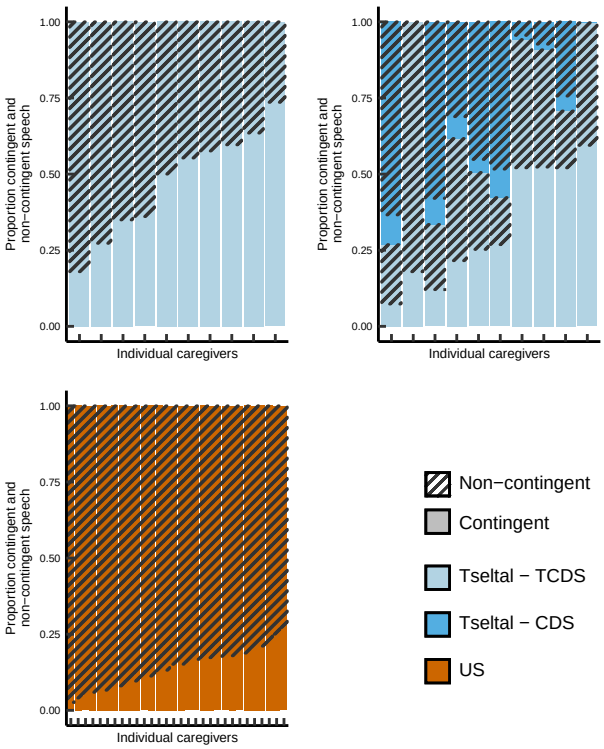

Figure 4: Distribution of caregiver utterance type per site. Each bar along the x-axis represents an individual caregiver.

Caregiver speech: linguistic comparisons

Tseltal caregivers' target-child-directed contingent and non-contingent utterances were equal in number (M^C=73.4, SD^C=61.81, M^{NC}=76.1, SD^{NC}=55.481, Estimate=-2.7, t=-0.1, p=0.9193) (Figure 4). In total, Tseltal caregivers produced 734 contingent and 761 non-contingent target-child-directed utterances. When including general child-directed speech (+CDS) in our measure of Tseltal caregiver speech, there were roughly equal number of contingent than non-contingent utterances (M^C=82.4, SD^C=59.144, M^{NC}=131.1, SD^{NC}=57.855, Estimate=-48.7, t=-1.86, p=0.0791). In total, Tseltal caregivers produced 824 contingent and 1311 non-contingent child-directed utterances. US caregivers produced less contingent than non-contingent speech, with significantly fewer utterances spoken contingently on their infants' vocalizations (M^C=13.65, SD^C=10.88, M^{NC}=83.65, SD^{NC}=32.389, Estimate=-70, t=-15.87, p<.0001) (Figure 4). In total, US caregivers produced 819 contingent and 5019 non-contingent target-child-directed utterances.

To compare lexical diversity across contingent and non-contingent speech, the number of unique words caregivers produced was calculated for contingent and non-contingent utterances (Figure 5A, Table 2). When only considering Tseltal target-child-directed speech (TCDS), contingent and

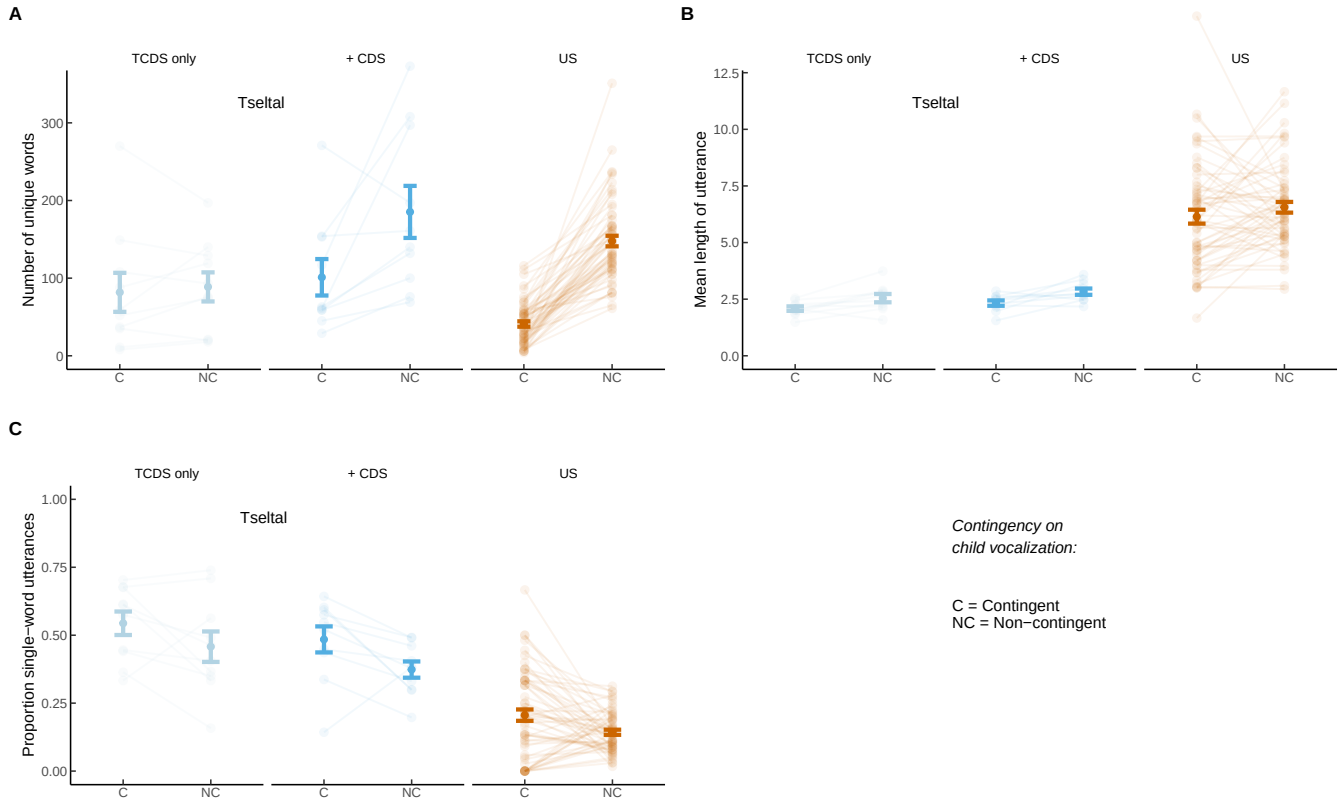

Figure 5: Simplification of contingent speech across Tseltal and US caregivers. (A) Number of unique words per contingent and non-contingent utterances. (B) Mean length of utterance in words per contingent and non-contingent utterances. (C) Proportion of single word utterances per contingent and non-contingent utterances. Transparent dots and connected lines show each caregiver's contingent and non-contingent speech structure. Mean and ± standard error shown in bold.

non-contingent speech contained equal counts of unique words. When including child-directed speech (+CDS) in Tseltal unique word counts, there were significantly fewer unique words spoken contingently. US caregivers produced significantly fewer unique words in their contingent speech relative to non-contingent speech.

To compare syntactic complexity across speech types, caregivers' MLUw was calculated for contingent and non-contingent utterances (Figure 5B, Table 2). Tseltal TCDS and +CDS had significantly shorter contingent utterances than non-contingent utterances. US contingent utterances were significantly shorter than non-contingent utterances.

To further test syntactic complexity, the proportion of utterances which contained a single word was calculated for contingent and non-contingent utterances (Figure 5C, Table 2). Tseltal TCDS and +CDS had a significantly higher proportion of contingent than non-contingent utterances that were a single word (Table 2). US contingent utterances were also more likely to contain only a single word compared to non-contingent utterances (Table 2).

Discussion

We found that US and Tseltal caregivers simplified the statistical and syntactic structure of their speech in response to their child's vocalizations. The simplification pattern generally holds despite cultural differences in the extent to which child vocalizations elicit caregiver responses. In both groups, contingent speech largely contained fewer unique words, contained shorter utterances and was more likely to be a single-word utterance. Together, these characteristics of parents' contingent speech suggest a stable form of influence of children's immature vocalizing on the ambient linguistic environment. Children's vocal behavior may create language learning opportunities by eliciting responses from parents that contain more learnable information.

The lexical and syntactic simplification found in contingent speech may benefit children. Reduced contingent lexical diversity is likely beneficial as clusters of successive word repetitions predict children's learning of the repeated words (Schwab & Lew-Williams, 2016). Shorter caregiver utterances, and single-word utterances in particular, simplify the task of finding word boundaries and facilitates language learning (Lew-Williams, Pelucchi, & Saffran, 2011).

As the pattern of contingent simplification appears largely

	Contingent		**Non-contingent**				
	Unique word count	SD	Unique word count	SD	Estimate	p	95% CI
TCDS	81.7	79.39	88.7	59.25	7	0.6057	-22.61, 36.61
+CDS	101.1	74.21	185.3	106	84.2	0.0296	10.44, 157.96
US	40.93	26.79	147.77	52.79	106.83	<.0001	92.99, 120.68
	Mean length of utterance	SD	Mean length of utterance	SD	Estimate	p	95% CI
TCDS	2.2	1.82	2.51	2.11	0.4	0.0002	0.19, 0.61
+CDS	2.31	1.88	2.88	2.38	0.58	<.0001	0.38, 0.78
US	6.18	6.28	6.47	5.78	0.49	0.0263	0.06, 0.92
	Prop. single-word utterance	SD	Prop. single-word utterance	SD	Estimate	p	95% CI
TCDS	0.52	1.82	0.44	2.11	-0.11	<.0001	-0.16, -0.06
+CDS	0.49	1.88	0.37	2.38	-0.13	<.0001	-0.17, -0.09
US	0.22	6.28	0.14	5.78	-0.08	<.0001	-0.11, -0.05

Table 2: Comparison of contingent and non-contingent speech structure across sites. Estimates derived from the following model structure: `caregiver speech structure ~ contingency + infant age + (1|subject)`. TCDS = Tseltal caregivers' target-child-directed speech, +CDS = both Tseltal caregivers' target-child-directed speech and their child-directed speech more generally

similar across Tseltal and US caregivers, the simplification effect of contingent speech may be independent of the attitudes towards language pedagogy in a given community. Reports demonstrate that adult Tseltal speech to children does not typically include the attention-getting and child-centric features typical of US English infant-directed speech (Brown, 2011, 2014; Soderstrom, Casillas, Gornik, et al., 2021). Thus, the source of the simplification effect of immediate responses to children's vocalizations may have more to do with the strict timing demands, or the immaturity of children's vocalizations, than adults' goals when interacting with young children (see Elmlinger et al., 2021 for further discussion). While the main focus of the present work was to understand the nature of the simplification effect across Tseltal and US cultures, future research will need to extensively examine the underlying mechanism of simplification.

Tseltal children's vocalizations elicited lower rates of caregiver responses than the US infants. This difference may arise from a number of limitations of the present study. Tseltal and US recording durations, contexts, and age differences may have contributed to this difference. Recent research on home recordings of US infants suggests that infant vocalizations' may elicit caregiver responses at a rate of 21 percent, a rate comparable to the Tseltal data presented here (Lopez, Walle, Pretzer, & Warlaumont, 2020). However, because Lopez et al. (2020) relied upon automatic coding of caregiver and infant utterances, which may overestimate the amount of turn-taking in a recording, future work conducted with manual annotation is required to fully understand cross-cultural differences in response rates (Ferjan Ramírez, Hippe, & Kuhl, 2021).

The relative distribution of contingent and non-contingent Tseltal and US caregiver child-directed speech differed. While Tseltal caregivers produced equal amounts of contingent and non-contingent speech, US caregivers produced much more non-contingent speech. It is possible that this pattern reflects the cultural norms in the Tseltal community where talk with children may occur mainly within turn alternations and caregiver talk which extends beyond turn-taking intervals may be more rare than in US caregivers.

Our results suggest that children, via immature vocalizing, play an important role in shaping their own language environment in multiple, distinct cultural contexts. Future research is required to address the comparative limitations in the present work, including the differences in recording duration, interpersonal context, and target child age. Currently, we are adapting the measures in this research to measure the morphemes in caregiver speech to reflect the differing morphological systems of the languages. In spite of these limitations, this work represents a useful advance in understanding how children's real-time interaction with adults facilitates language learnability.

Acknowledgements

We thank and Juan Méndez Girón, Julia Lescht, Rebeca Guzmán López, Antun Gusman Osil, Humbertina Gómez Pérez, Jenna Steins, and Joelle Tancredi, who made the collection and annotation of these recordings possible. We are also enormously grateful to the participating families and to the community at large for their support. Thanks and aknowledgements also to Maartje Weenink and Daphne Jansen for their contribution to manual clip selection.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv Preprint arXiv:1406.5823*.

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure

to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44.

Brown, P. (2011). The cultural organization of attention. In A. Duranti, E. Ochs, & and B. B. Schieffelin (Eds.), *Handbook of Language Socialization* (pp. 29–55). Malden, MA: Wiley-Blackwell.

Brown, P. (2014). The interactional context of language learning in Tzeltal. In I. Arnon, M. Casillas, C. Kurumada, & B. Estigarribia (Eds.), *Language in interaction: Studies in honor of Eve V. Clark* (pp. 51--82). Amsterdam, NL: John Benjamins.

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., ... Casillas, M. (under review). A cross-cultural examination of young children's everyday language experiences.

Carouso-Peck, S., & Goldstein, M. H. (2019). Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. *Current Biology*, *29*(4), 631–636.

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.

Elmlinger, S. L., Park, D., Schwade, J. A., & Goldstein, M. H. (2021). Comparing word diversity versus amount of speech in parents' responses to infants' prelinguistic vocalizations. *IEEE Transactions on Cognitive and Developmental Systems*.

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019a). Babbling elicits simplified caregiver speech: Findings from natural interaction and simulation. In *2019 joint IEEE 9th international conference on development and learning and epigenetic robotics (ICDL-EpiRob)* (pp. 1–6). IEEE.

Elmlinger, S. L., Schwade, J. A., & Goldstein, M. H. (2019b). The ecology of prelinguistic vocal learning: Parents simplify the structure of their speech in response to babbling. *Journal of Child Language*, *46*(5), 998–1011.

Ferjan Ramírez, N., Hippe, D. S., & Kuhl, P. K. (2021). Comparing automatic and manual measures of parent–infant conversational turns: A word of caution. *Child Development*, *92*(2), 672–681.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development*, 1497–1510.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *19*(5), 515–523.

Gultekin, Y. B., & Hage, S. R. (2018). Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys. *Science Advances*, *4*(4), eaar4012.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365.

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, *14*(6), 1323–1329.

Lopez, L. D., Walle, E. A., Pretzer, G. M., & Warlaumont, A. S. (2020). Adult responses to infant prelinguistic vocalizations are associated with infant vocabulary: A home observation study. *PloS One*, *15*(11), e0242232.

Oller, D. K. (2000). *The emergence of the speech capacity*. Psychology Press.

Oller, D. K., & Lynch, M. P. (1992). Infant vocalizations and innovations in infraphonology: Toward a broader theory of development and disorders. *Phonological Development: Models, Research, Implications*, 509–536.

Onnis, L., & Edelman, S. (2019). Local versus global statistical learning in language.

Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (MLUm) and mean length of utterance in words (MLUw). *First Language*, *25*(3), 365–376.

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, *13*(1), 85–100.

Roy, B. C., Frank, M. C., & Roy, D. K. (2009). Exploring word learning in a high-density longitudinal corpus. *Proceedings of the Thirty-First Annual Conference of the Cognitive Science Society*.

Schwab, J. F., & Lew-Williams, C. (2016). Repetition across successive sentences facilitates young children's word learning. *Developmental Psychology*, *52*(6), 879.

Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Alam, F., Warlaumont, A. S., & Bunce, J. (2021). Developing a cross-cultural annotation system and MetaCorpus for studying infants' real world language experience. *Collabra: Psychology*, *7*(1), 23445.

Soderstrom, M., Casillas, M., Gornik, M., Bouchard, A., MacEwan, S., Shokrkon, A., & Bunce, J. (2021). English-speaking adults' labeling of child- and adult-directed speech across languages and its relationship to perception of affect. *Frontiers in Psychology*, *12*, 708887.

Stockman, I. J. (2010). Listener reliability in assigning utterance boundaries in children's spontaneous speech. *Applied Psycholinguistics*, *31*(3), 363–395.

Venker, C. E., Bolt, D. M., Meyer, A., Sindberg, H., Ellis Weismer, S., & Tager-Flusberg, H. (2015). Parent telegraphic speech use and spoken language in preschoolers with ASD. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1733–1746.