<div style="border:1px solid">
This article is part of the topic "Best of Papers from the 2022 Cognitive Science Society Conference," Andrea Bender (Topic Editor).
</div>

Immature Vocalizations Simplify the Speech of Tseltal Mayan and U.S. Caregivers

Steven L. Elmlinger,[a] Michael H. Goldstein,[a] Marisa Casillas[b]

[a]*Department of Psychology, Cornell University*
[b]*Department of Comparative Human Development, University of Chicago*

Abstract

What is the function of immature vocalizing in early learning environments? Previous work on infants in the United States 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 U.S. caregivers' verbal responses to their children's vocalizations. Similar to findings from U.S. 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 that may facilitate language learning within ongoing interaction.

Keywords: Language input; Parent–child interaction; Language statistics; Child-directed speech

1. Introduction

Across several species, adults' contingent responses to their offspring's immature vocalizations play a key role in communicative development. Vocal learning studies on humans (Goldstein & Schwade, 2008), songbirds (Carouso-Peck & Goldstein, 2019), and marmoset monkeys (Gultekin & Hage, 2017; 2018; Takahashi, Fenley, & Ghazanfar, 2016) demonstrate that adults who coordinate their vocalizations with those of their offspring create contingent social feedback that facilitates advancements in vocal communication. However, contingent verbal responses may be rare in human infants' experience (Fagan & Doveikis, 2017; Goldstein & Schwade, 2008; Goldstein, King, & West, 2003). When these responses are rare, how might they facilitate learning? Recent findings demonstrate that child-directed talk in both U.S. 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 are 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. Simpler instances of utterance structure may be particularly useful for children when first breaking into the structure of their ambient language.

What role do children play in structuring the learnability of ambient language? By 9 months, English-learning infants in the United States regulate the complexity of their caregivers' speech by vocalizing (Elmlinger, Park, Schwade, & Goldstein, 2021; Elmlinger, Schwade, & Goldstein, 2019a, 2019b). Maternal utterances that are child-directed and contingent on infant vocalizations are lexically and syntactically simplified compared to utterances that are similarly child-directed but not contingent on infant vocalizations. The contingent speech was found to have fewer unique word types, fewer words per utterance, and higher proportions of one-word utterances (Elmlinger, Schwade, & Goldstein, 2019b). Thus, responses to babbling reduce the complexity of caregivers' speech in ways that may facilitate learning (Schwab & Lew-Williams, 2016). Presently, however, our understanding of children's effects on caregiver speech is limited to a North American cultural context. The cross-cultural extent of this effect is unknown. Why would it vary?

Differences in caregivers' attitudes about child language development and socialization across cultures—specifically about when and how to respond to child vocalizations—may predict whether contingent responses to infants' vocalizations are 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 U.S. adult caregivers to engage infants in child-centric and pedagogical speech interactions (see below), may not show the simplification effect found in U.S. adults. Alternatively, the simplification effect may be independent of pedagogical attitudes, and the contingent simplification found in U.S. 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 the speech simplification of Tseltal Mayan and U.S. caregivers.

1.1. Ethnographic background

The Tseltal participants in the present study 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 multigenerational patrilocal family compounds (i.e., typically near their nuclear family, paternal grandparents, and paternal uncles' nuclear families). Tseltal is the primary language spoken at home; Spanish is acquired in primary school, after the first couple of years in Tseltal.

Young children are carried for much of their 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 more participants, and focuses on appropriate actions and responses rather than words and word meanings (Brown, 2014). Across a waking day, children under age 3;0 hear an average of 3.6 min of speech directly addressed to them per hour (Casillas et al., 2020), which may be comparable to averages from other (e.g., the United States) 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).

1.2. Infant-directed speech in Tseltal

Tseltal infant-directed speech is not recognizable as such by naïve Western listeners, who cannot effectively distinguish it from an adult-directed speech by the same speakers, despite high reported confidence ratings (Soderstrom et al., 2021). The 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 different Mayan language, 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 (mean length of utterance) in morphemes between adult- and infant-directed speech.

1.3. Present study

We focused on three aspects of variation in caregiver linguistic structure—lexical diversity, syntactic complexity, and one-word utterances—that may influence the language acquisition process via simplification. We first measured lexical diversity, which has effects on learning that differ by timescale: partially repetitive speech positively predicts language learning in the short term (Onnis & Edelman, 2019; Schwab & Lew-Williams, 2016), while variability in

lexical input predicts outcomes the same in the long term (Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010). To measure lexical diversity, we counted the number of unique words (types) in parents' talk to children. We next measured syntactic complexity using utterance length as a proxy; the length of parents' utterances to children in everyday learning environments is sensitive to when those words emerge in children's own production, getting shorter just before and then increasing again after a word emerges (Roy, Frank, & Roy, 2009). We assessed the syntactic complexity of parents' speech by determining the mean length of utterances in words (MLUw) (Parker & Brorson, 2005). Finally, we measured one-word utterances because 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). We measured the proportion of utterances composed of a single word. By investigating the structure of contingent and non-contingent speech across Tseltal and U.S. caregivers, we can better understand the role that children's vocalizations may play in influencing their own communicative development.

Because of the limited comparability across samples included in this study, here we treated the Tseltal and U.S. measures as two individual case studies. We did not make direct statistical comparisons across sites due to differences in participants' age and recording context. The central question of interest is whether speech structure is altered in qualitatively similar ways when organized around children's vocalizations across different cultural contexts. The extent to which we see similar patterns of speech structure change across cultures serves as evidence for or against a generalized mechanism driving simplification in contingent caregiver talk.

2. Methods

2.1. Participants

Ten 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.

Thirty U.S. 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.

2.2. Recordings and procedure

The Tseltal recordings analyzed here are the same used in Casillas et al. (2020). The recordings were 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 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 shirt that had a horizontal pocket to store the recorder (Fig. 1).

Table 1
Recording descriptive statistics

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

Abbreviations: Avg, average; C, child; CG, caregiver; TCDS, Tseltal caregivers' target-child-directed speech; +CDS, all Tseltal child-directed speech (i.e., target-child-directed speech and other-child-directed speech); utt, utterance; voc, vocalization.
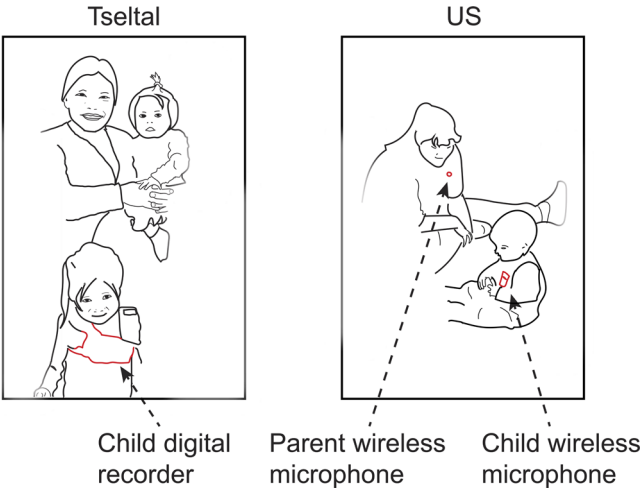

Fig. 1. Recording setup across sites.

Tseltal children wore the recorder continuously throughout the ∼9-h 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 U.S. data took place in a naturalistic environment in a 12-foot by 18-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) (Fig. 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 U.S. participant engaged in 15-min play sessions in the lab. During these sessions, caregivers were asked to play like they would at home, resulting in unstructured free play.

2.3. Speech transcription

Tseltal parents' speech sample consists of 60 min of transcription per recording. Forty-five of the 60 total minutes were randomly selected 5-min clips. These speech samples were annotated and transcribed jointly by the visiting Western researcher and a local member of the community who natively spoke Tseltal and knew all the recorded families personally. Annotations included full transcriptions of all hearable speech, with an indication for each utterance regarding to whom it 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 the 60 total minutes were hand-selected from the remaining, unannotated portions of each recording. A comprehensive review of each audio recording, excluding the original random clips, allowed us to identify the five top 1-min segments of turn-taking between the target child and their interactants, then the five top 1-min segments of target child vocalization from the remaining recording times. The most active interaction captured in those 10 1-min clips was then expanded a further 5 min, with all additional 15 min of clip time per recording fully annotated using the same standards as the random clips. This process resulted in 1 h 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 h of audio in total). Speech in these recordings comes from many speakers; here, we focus exclusively on the speech from each target child's mother for better comparison to the U.S. data.

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 plus any caregiver speech which was directed at other children. Because Tseltal children may treat both TCDS and +CDS as relevant learning cues, we analyze both for changes when they are contingent and not contingent on the target child's noncry vocalizations.

The speech that U.S. caregivers produced was completely transcribed. If caregivers' utterances were separated by silence longer than 2 s 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 from the analyses.

When Tseltal and U.S. caregivers' utterances occurred within 2 s of the offset of the target children's vocalizations, then they were considered contingent utterances (Elmlinger et al., 2019a). Caregiver utterances which occurred after a 2-s time frame were considered non-contingent (Fig. 2). Two seconds were used following previous studies which originally reported on the simplification of contingent speech (Elmlinger et al., 2019b). Responses to infant vegetative vocalizations, such as coughs, cries, and fusses, were excluded from the analysis.
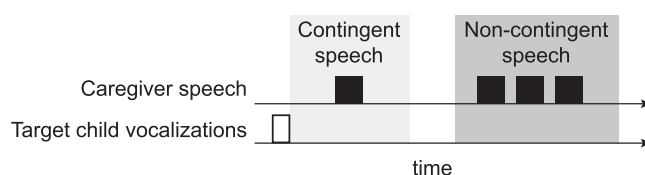
Fig. 2. U.S. and Tseltal caregiver utterances were considered contingent when they occurred within 2 s of the target child's vocalizations. This same contingency definition was used for Tseltal TCDS and +CDS analyses.

2.4. *Child utterances*

The onsets and offsets of all Tseltal infant nonvegetative, communicative vocalizations (i.e., including laughter, fussing, and crying) were annotated and segmented approximately according to breath groups, with some exceptions (e.g., longer bouts of crying). When lexical, vocalizations were transcribed, and were otherwise classified as containing canonical syllables or not, or containing laughter or crying.

U.S. infant noncry vocalization bout onsets and offsets were annotated in full. Vocalization boundaries were segmented according to breath groups (Oller, 2000; Oller & Lynch, 1992).

2.5. *Analytic approach*

We employed linear and logistic mixed regressions with the lme4 package in R (Bates, Mächler, Bolker, & Walker) to predict linguistic structure from contingency, controlling for target child age, with participants as a random effect. We used linear regression in comparing the proportion of child noncry vocalizations which elicited a response across the site and in comparing caregivers' distribution of contingent to non-contingent utterances.

3. Results

3.1. *Levels of contingent responsiveness*

Tseltal children's vocalizations elicited their mother's verbal response around 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 ($M = 0.23$, $SD = 0.126$). U.S. infant vocalizations elicited caregiver verbal responses more than a third of the time ($M = 0.407$, $SD = 0.156$). Noncry vocalizations of U.S. infants were more likely to elicit a caregiver response than Tseltal children's similar vocalizations (Estimate $= 0.177$, $t = 3.41$, $p = .0011$) (Fig. 3).

3.2. *Caregiver speech: Linguistic comparisons*

Tseltal caregivers' target-child-directed contingent and non-contingent utterances were equal in number ($M^C = 73.40$, $SD^C = 61.81$, $M^{NC} = 76.10$, $SD^{NC} = 55.48$, Estimate $=$
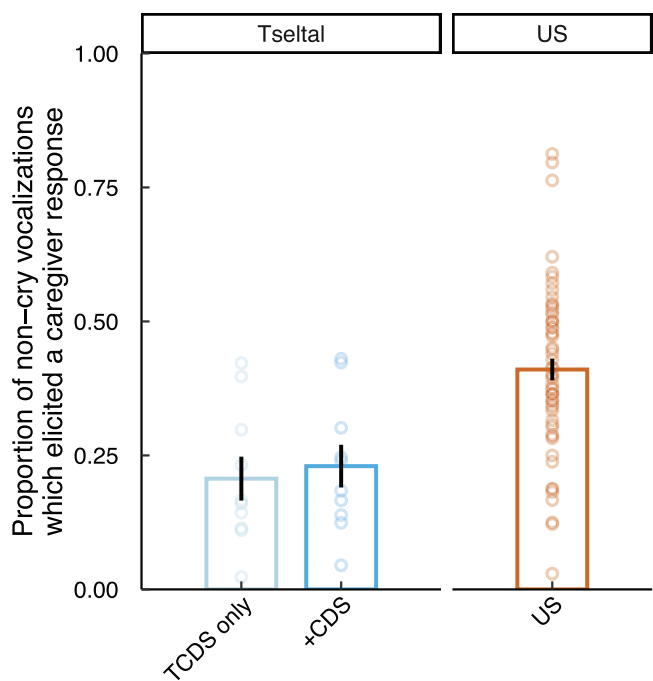
Fig. 3. Proportion of child vocalizations which elicited a contingent response. Black lines indicate $\pm$ 1 standard error around the mean.

–2.7, $t = –0.1$, $p = .9193$) (Fig. 4). In total, Tseltal caregivers produced 734 contingent and 761 non-contingent target-child-directed utterances. On average, Tseltal caregivers did not produce significantly more non-contingent than contingent +CDS utterances ($M^C = 82.40$, $SD^C = 59.14$, $M^{NC} = 131.10$, $SD^{NC} = 57.86$, Estimate $= –48.7$, $t = –1.86$, $p = .0791$). In total, Tseltal caregivers produced 824 contingent and 1311 non-contingent child-directed utterances. U.S. caregivers produced less contingent than non-contingent speech, proportionally producing contingent utterances significantly less often ($M^C = 13.67$, $SD^C = 10.86$, $M^{NC} = 83.67$, $SD^{NC} = 32.39$, Estimate $= –70$, $t = –15.87$, $p < .0001$) (Fig. 4). In total, U.S. 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 (Fig. 5a and Table 2). When only considering Tseltal target-child-directed speech (TCDS), contingent and non-contingent speech contained equal counts of unique words. When including all child-directed speech (+CDS) in Tseltal unique word counts, there were significantly fewer unique words spoken in contingent speech. U.S. 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 (Fig. 5b and Table 2). Tseltal TCDS and +CDS
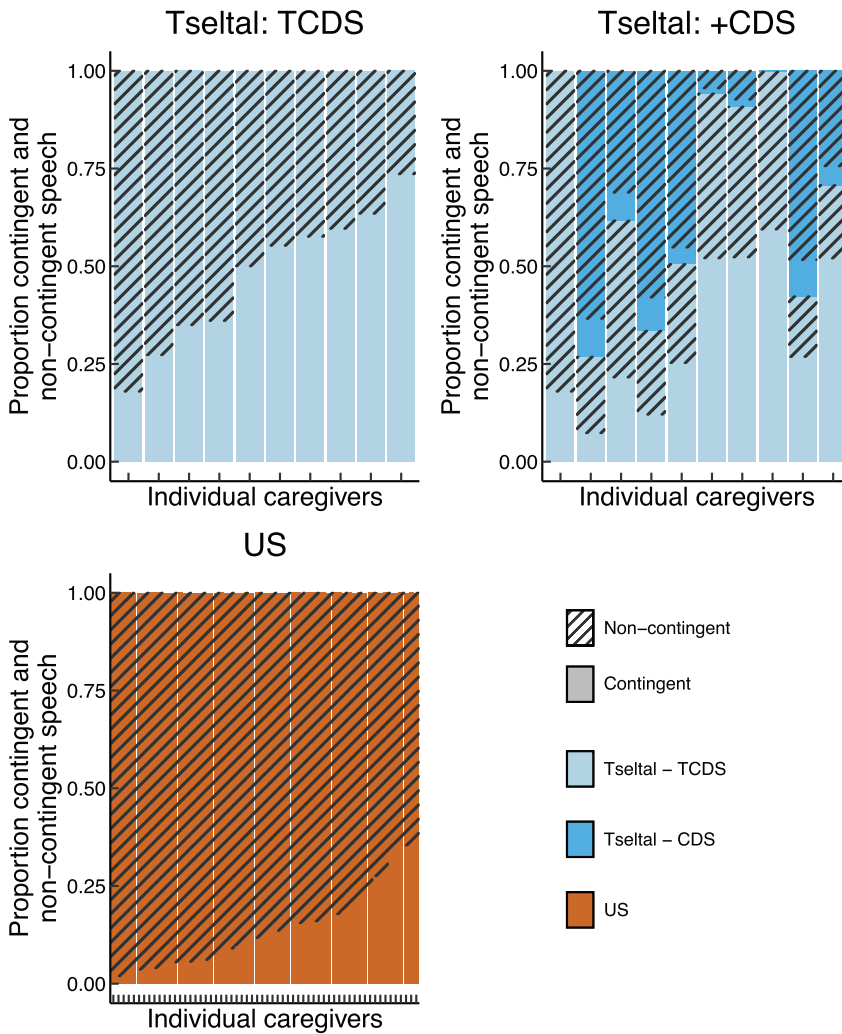
Fig. 4. Distribution of caregiver utterance type per site. Each bar along the x-axis represents an individual caregiver.

had significantly shorter contingent utterances than non-contingent utterances. U.S. 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 (Fig. 5c and Table 2). Tseltal TCDS and +CDS had a significantly higher proportion of contingent than non-contingent utterances that were a single word (Table 2). U.S. contingent utterances were also more likely to contain only a single word compared to non-contingent utterances (Table 2).
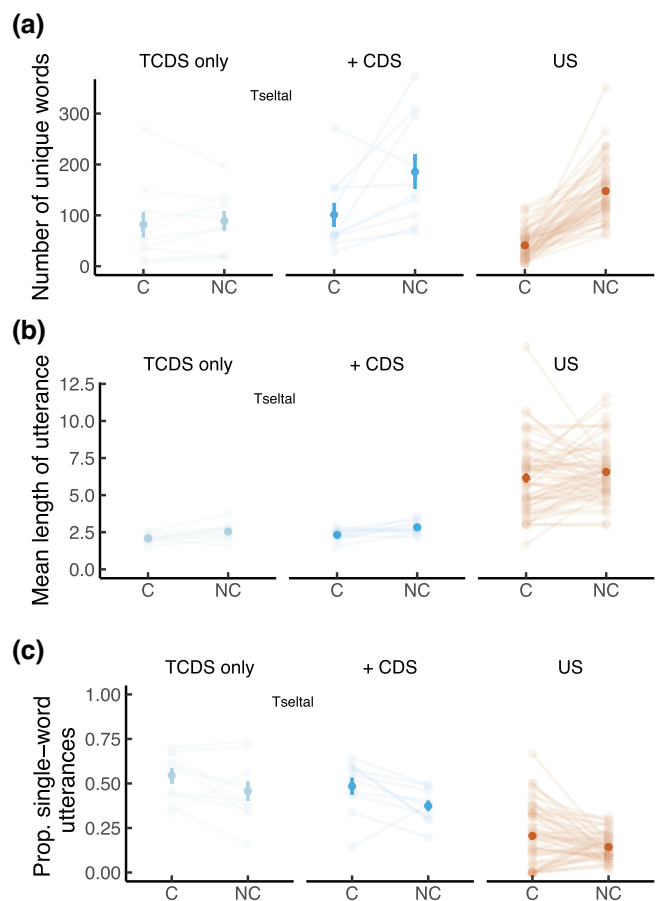
Fig. 5. Simplification of contingent speech across Tseltal and U.S. caregivers. Each of the three dependent measures is given per contingent (left) and non-contingent (right) utterances: (A) number of unique words, (B) mean length of utterance in words (MLUw), and (C) proportion of single-word utterances. Line-connected data points indicate data from an individual caregiver. Mean and ± standard error are shown in bold. Abbreviations: C, contingent; NC, non-contingent.

4. Discussion

U.S. 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 how caregivers talk to children, as documented in prior ethnographic work (Brown, 2011, 2014). Here, we observed the simplification pattern despite differences in the extent to which child vocalizations elicited caregiver responses and in the relative frequency of non-contingent utterances. 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 caregivers' contingent speech suggest a stable

Table 2

Comparison of contingent and non-contingent speech structure across sites

	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	.6057	–22.61, 36.61
+CDS	101.1	74.21	185.3	106	84.2	.0296	10.44, 157.96
US	41.03	26.67	147.77	52.79	106.73	<.0001	92.88, 120.59
	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	.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	0.48	.0276	0.05, 0.91
	Prop. single-word utterance	SD	Prop. single-word utterance	SD	Estimate	p	95% CI
TCDS	0.52	1.82	0.44	2.11	–0.46	<.0001	–0.68, –0.24
+CDS	0.49	1.88	0.37	2.38	–0.55	<.0001	–0.74, –0.36
US	0.22	6.28	0.14	5.78	–0.57	<.0001	–0.76, –0.37

Note: Estimates derived from the following model structure: caregiver speech structure ~ contingency + infant age + (1|subject).

Abbreviations: TCDS, Tseltal caregivers' target-child-directed speech; +CDS, all Tseltal child-directed speech (i.e., target-child-directed speech and other-child-directed speech).

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 caregivers that contain more learnable information.

The lexical and syntactic simplification found in contingent speech may benefit children at the beginning of their language-learning process. 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 thereby facilitate language learning (Lew-Williams, Pelucchi, & Saffran, 2011).

As the pattern of contingent simplification appears largely similar across Tseltal and U.S. caregivers, the simplification effect of contingent speech may occur independently of the attitudes toward 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 U.S. English infant-directed speech (Brown, 2011, 2014; Soderstrom et al., 2021). Thus, the source of the simplification effect of immediate responses to children's vocalizations may have more to do with the immaturity of children's vocalizations than adults' goals when interacting with young children. The language proficiency of one interaction partner may predict the extent of linguistic simplification in the other's contingent responses (Elmlinger et al., 2021). Future work may investigate this further with TCDS versus +CDS; here, we found similar but nonidentical simplification effects in Tseltal. If children's learning benefits from simplification, they may benefit indirectly from speech directed contingently to their peers. While the main focus of the present work was to test whether similar contingent speech simplification effects emerge across Tseltal and U.S. cultures, future research will need

to extensively examine the underlying mechanisms of simplification, including the extent to which they are similarly or differently employed both with and across cultural groups.

Tseltal children's vocalizations elicited lower rates of caregiver responses than the U.S. infants. This difference may arise from a number of limitations of the present study. Tseltal and U.S. recording durations, contexts, and age differences may have contributed to this difference. Recent research on home recordings of U.S. infants suggests that infant vocalizations' may elicit caregiver responses at a rate of 21%, 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).

Our results suggest that children, via immature vocalizing, play an important role in shaping their own language environment in 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 also adapting the primary measures so that they are based on lemmas (lexical diversity) and morphemes (MLU) rather than words. This more morphosyntactically informed approach gives us a more accurate view of simplification effects considering the differing morphological systems of the languages. In spite of these limitations, this work substantially advances our understanding of how children's real-time interaction with adults changes their linguistic experiences and thereby may facilitate the process of language learning.

Acknowledgments

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4, *67*(1), 1-48.

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, & 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., Migdalek, M. J., & Casillas, M. (2021). A cross-cultural examination of young children's everyday language experiences. *PsyArXiv*.

Carouso-Peck, S., & Goldstein, M. H. (2019). Female social feedback reveals non-imitative mechanisms of vocal learning in zebra finches. *Current Biology*, *294*, 631–636.

Casillas, M., Brown, P., & Levinson, S. C. (2020). Early language experience in a Tseltal Mayan village. *Child Development*, *915*, 1819–1835.

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language*, *484*, 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). 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*, *465*, 998–1011.

Fagan, M. K., & Doveikis, K. N. (2017). Ordinary interactions challenge proposals that maternal verbal responses shape infant vocal development. *Journal of Speech, Language, and Hearing Research*, *6010*, 2819–2827.

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*, *922*, 672–681.

Goldstein, M. H., & Schwade, J. A. (2008). Social feedback to infants' babbling facilitates rapid phonological learning. *Psychological Science*, *195*, 515–523.

Goldstein, M. H., King, A. P., & West, M. J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences*, *10013*, 8030–8035.

Gultekin, Y. B., & Hage, S. R. (2017). Limiting parental feedback disrupts vocal development in marmoset monkeys. *Nature Communications*, *81*, 1–9.

Gultekin, Y. B., & Hage, S. R. (2018). Limiting parental interaction during vocal development affects acoustic call structure in marmoset monkeys. *Science Advances*, *44*, eaar4012.

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *614*, 343–365.

Lew-Williams, C., Pelucchi, B., & Saffran, J. R. (2011). Isolated words enhance statistical language learning in infancy. *Developmental Science*, *146*, 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*, *1511*, 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. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 509–536). Timonium, MD: York Press.

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*, *253*, 365–376.

Pye, C. (1986). Quiché Mayan speech to children. *Journal of Child Language*, *131*, 85–100.

Roy, B. C., Frank, M. C., & Roy, D. K. (2009). Exploring word learning in a high-density longitudinal corpus. Proceedings of the 31st 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*, *526*, 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*, *71*, 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*, *313*, 363–395.

Takahashi, D. Y., Fenley, A. R., & Ghazanfar, A. A. (2016). Early development of turn-taking with parents shapes vocal acoustics in infant marmoset monkeys. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *3711693*, 20150370.

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*, *586*, 1733–1746.