(Non-)effects of linguistic environment on early stable consonant production: A cross cultural case study

Bram Peute Radboud University Nijmegen Houtlaan 4, 6525 XZ Nijmegen Leiden University Rapenburg 70, 2311 EZ Leiden

aakvs@xs4all.nl

Marisa Casillas

Max Planck Institute for Psycholinguistics

Wundtlaan 1, 6525 XD Nijmegen

University of Chicago

Rosenwald Hall, 3rd Floor, 1101 E. 58th

St., Chicago, IL, 60637 USA

mcasillas@uchicago.edu

Word count: 7501

Abstract

Recent evidence shows that children reach expected basic linguistic milestones in two rural Indigenous communities, Tseltal and Yélî, despite their infrequent exposure to child-directed speech from adults. However, those results were partly based on measures that are fairly robust to environmental variation, e.g. the onset of babbling. By contrast, directed speech input is typically linked to lexical development, which is environmentally sensitive. We investigate whether these children's vocal motor schemes—a motor-phonological measure of stable consonant production related to early vocabulary—show a similar "expected" developmental pattern to what has been found for typically developing children in urban "child-centered" linguistic communities. We also compare development between the two languages, whose phonological inventories differ greatly in size and complexity. Using spontaneous speech from clips sampled across children's waking days at home, we find that children's canonical babble and stable consonant production is comparable to previous work with typically developing children. We find no evidence for difference between the languages

Keywords: language acquisition; consonant production; vocal motor schemes; canonical babble; child directed speech

1. Introduction

Children's acquisition of spoken language, from early vocalizations to first words and beyond, offers insight into the evolutionary origins and environmental sensitivity of the human language learning system. The trajectory of early vocal development is well documented across a diverse set of language communities: infants start producing protophones soon after birth, then around 6 to 8 months of age they begin to babble, and then some time around the first birthday their first words appear (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021). Canonical babble serves as an important step along the road to first words. Canonical babble is composed of well-formed syllables, typically reduplicated consonant-vowel structures, like "mama" or "da" (Lee et al. 2017). This kind of early babble is a milestone in the child's motor development that also marks the beginning of their ability to reliably and distinguishably produce phones similar to the phonemes of their home language(s) (McGillion et al. 2017).

The robustness of this overall developmental trajectory is underscored in recent work documenting similar developmental patterns despite cross-cultural variability in children's linguistic input. For example, while children in the Tseltal and Yélî Dnye speech communities are directly addressed at a rate similar to middle-class children in urban contexts (e.g., in the US), a substantial portion of their directed speech input (henceforth CDS; "child-directed speech") comes from other children—not adults (Bunce et al., in under review). Diverging caregiver ideologies between these two communities about how to talk to young children also mean that typical CDS input differs greatly in style, everyday activities, topics under discussion, and likely interlocutors (Brown & Casillas, in press). Despite these differences, Tseltal and Yélî Dnye-children's single- and multi-word lexical vocalizations showed a similar onset to each other and to children in urban, child-centric talk settings where a greater portion of CDS comes from adults (Bunce et al., in under review; Casillas et al. 2020a; b). This result appears, at face value, to run counter to other work clearly showing that "high quality" adult CDS is strongly associated with faster-growing receptive and productive vocabularies (e.g., Hart & Risley 1995; Hoff 2003; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013). Critically, however, the investigations by Casillas and colleagues were not focused on vocabulary development. The authors instead focused on a few broad categories of vocalization type, several of which are known to be fairly robust to environmental variation (e.g., Oller et al. 1995; Lee et al. 2017; Cychosz et al. 2021).

A stronger test of the impact of CDS input type on early vocal development would instead be to look at patterns in production that link children's pre-lexical behavior to their eventual lexical development. If there were still no evidence for delayed development with measures that relate to the lexicon, it would more strongly suggest that these children indeed have some means of gathering language information on the same timescale as children in child-centric environments, and would thus justify further work investigating how children pick up on linguistic information from these environments (e.g., via sibling-provided input, third-party input, or other means; Rogoff et al. 1993; Shneidman & Goldin-Meadow 2012; see Casillas et al. 2020a for a discussion).

While past work on early vocal development has been linguistically diverse (de Boysson-Bardies & Vihman 1991; Kunnari 2003; Fikkert & Levelt 2008; Lee et al. 2010; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013; Lee et al. 2017) links between early phonological development, caregiving practices, and everyday language use have yet to be drawn for any rural, Indigenous context. Especially considering that many languages spoken in rural and traditional or Indigenous contexts have typological features underrepresented in developmental research (Kidd & Garcia, 2021), exploring the predictive relationship between productions at the pre-lexical and early lexical stages of development in these communities is a key next step for research on language development.

Caregiving practices and everyday activities certainly shape CDS input differently in the two language communities studied here. While Tseltal children are socialized to become expert observer-participants in the interactions around them, Yélî children are more often placed at the center of social interactions (Brown 2011; 2014). CDS in both languages maintains some lexical and prosodic features that distinguish it from adult-directed speech, but Yélî Dnye CDS has been noted to be—relative to Tseltal—more overtly affect-laden and is more often heard from men and boys compared to the Tseltal context (Brown & Casillas in press; Bunce et al. under review). Infants in the Yélî context are more often passed between multiple caregivers throughout the day and, once they are walking, often spend their afternoon in larger child playgroups than what is typically found in the Tseltal community (Brown & Casillas in press). In brief, while both communities have a large amount of CDS input from other children (e.g., compared to the US), Yélî Dnye speaking children hear much more CDS from other children and from males compared to Tseltal speaking children.

In addition to these differences in CDS, the two language phonologies differ greatly, with Yélî Dnye using three times the number of contrastive consonants as Tseltal, and within relatively packed phonological parameters (e.g., no contrastive voicing). Further, little is known about the acquisition of phonemes in each language (e.g., Tseltal: ejectives, Yélî Dnye: doubly articulated consonants), some of which are rare across the world's languages. For these reasons, similar developmental trajectories between the two languages would suggest substantial robustness in early vocal development.

In addition to canonical babble, which was investigated in prior work with these communities, we here investigate children's "vocal motor schemes" (henceforth VMS), which are motor-plans for phone production developed through babbling. VMSs provide a bridge between pre-lexical babble and early productive vocabulary. They provide the child with the means of producing "auditory approximations" to the phonemes present in the target words produced by adults, but without the child having to deduce the exact phonemes involved (Vihman 1993). This capacity for consistent phonetic patterning is theorized to help prepare the child for producing stable consonants, described as "a generalized action plan that generates consistent phonetic forms... a formalized pattern of motor activity that does not require heavy cognitive resources to enact" (McCune & Vihman 2001: 152). McCune and Vihman (2001) operationalize VMS in their longitudinal dataset of at-home play sessions as follows: If a child produces 10 realizations of a given consonant in 3 or 4 contiguous monthly sessions, the child is said to have acquired the VMS for that consonant (see Figure 1 for an illustration of our similar counting process). On average, the typically developing American English-learning children in McCune and Vihman's (2001) study acquired two VMS consonants at 12 months (SD = 1.78, range = 9-14). A larger dataset of at-home play sessions with typically developing British English-learning children by DePaolis and colleagues (2011) showed a slightly earlier average achievement of 2 VMS: 10 months (SD = 1.35, range = 9–15). Though we note that this dataset used an additional criterion for counting a VMS (50 uses of a consonant within 30 minutes). While these datasets do not provide sufficient evidence for a clear linguistic milestone (e.g., as we have for the onset of canonical babble; Oller 2000; Lee et al. 2017; Cychosz et al. 2021), they do set a preliminary range of typical early VMS development: 2 VMS between 9 and 15 months, with children reaching 2 VMS by 10–12 months on average.

Variation in VMS development is meaningful for later lexical development. McCune and Vihman (2001) first linked VMS to children's shift to referential word use, defined in their study as use of a word across two or more contexts. Most children achieved 2 VMS just before making this referential transition, and most of these stable early word forms included the child's acquired VMS consonants. More recent work also supports the idea that variation in VMS development predicts multiple aspects of later lexical development and language processing including: the subsequent appearance of first words (McGillion et al. 2017), early expressive vocabulary size (Majorano et al. 2014), and attention to auditory presentations of familiar and unfamiliar words, or words that do and don't include the child's acquired VMS consonants (DePaolis et al. 2011; Majorano et al. 2014). Although VMS consonants do not directly reflect frequency of use in caregiver input (DePaolis et al. 2011; Majorano et al. 2014), the established link between variance in VMS acquisition and variance in later lexical development—the latter of which is known to be very sensitive to environmental variation in CDS (Hart & Risley 1995; Hoff 2003; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013)—suggests the possibility that broader characteristics of VMS development may vary across diverse CDS input contexts (e.g., more adult CDS relating to average earlier VMS acquisition in addition to average larger vocabulary size later on). If so, we might expect that children in language communities with less adult-produced CDS, such as the Tseltal and Yélî groups studied here, show somewhat later VMS development than children in environments with more frequent adult CDS and that, similarly, children learning Yélî Dnye, who hear more of their directed input from other children, would show somewhat later VMS development than Tseltal children.

1.1 The present study

The present study focuses on the acquisition of consonants (both in onset and coda position) by children born into two speech communities: Tenejapan Tseltal and Yélî Dnye. Tseltal is a Mayan language spoken in the highlands of Chiapas in Southern Mexico. This language, which has five main dialects, is in vigorous use with estimates of over 400,000 to 500,000 speakers who are bilinguals in Spanish, and an estimated 40,000 to 50,000 monolingual speakers (Polian 2013; Eberhard et al. 2020). Yélî Dnye is an isolate spoken on Rossel Island, in the Milne Bay Province of Papua New Guinea. The language, which has two main dialects, is in vigorous use as well, with a loosely estimated 5000-7000 native speakers, most of whom are at least partly bilingual in English, Tok Pisin, and/or other languages of Papua New Guinea (Levinson in press; Eberhard et al. 2020). Many aspects of daily life, as well as the overall rate of CDS, are similar between these two patrilocally organized, swidden horticulturalist communities (Casillas et al. 2020a; b). Before age 3;0, children in both communities are directly addressed around 3-3.5 minutes per hour during waking days at home (Casillas et al. 2020a; b). While the quantity of directed linguistic input is relatively similar, differences in early caregiving and ideas about speaking to young children create differences in who speaks to young children and how, with more input from men and children in the Yélî community compared to Tseltal. Additionally, the consonant inventories of the two languages are highly distinct from each other. Yélî Dnye has 60 contrastive consonants compared to Tseltal's 20, with the additional consonants including both single- and doubly-articulated stops with nasal, labial, and/or palatal releases and pre-nasalization (Levinson in press; Polian 2013). The fact that these two language communities use similar rates of directed input but have highly divergent phonological inventories makes the comparison of children's early consonant development particularly interesting.

If Tseltal and Yélî Dnye-acquiring children show similar VMS development to each other, and to same-aged English acquiring children, it would suggest that even lexically-related early phonological development is cross-culturally robust to variation in CDS use, and that children may use other cues from their linguistic environment to spur their early linguistic growth. For example, many scholars focusing on language development in non-urban contexts have highlighted the role of observable third-party behavior for learning (e.g., Rogoff et al. 1993; Chavajay & Rogoff 1999; Gaskins & Paradise 2010). In addition to VMS, we also attempt to replicate the finding that canonical babble onset shows no delayed development in these two communities, this time using a larger sample and independent annotations from the original study data (Casillas et al. 2020a; b).

1.2 Predictions

Based on prior findings, we hypothesize that the onset of canonical babble occurs between 6 and 8 months in both datasets (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021). We also predict that VMS acquisition by Yélî and Tseltal children will be comparable to that of Western children; that is, children will reach 2 VMS consonants between 9 and 15 months, typically between 10 and 12 months (McCune & Vihman 2001; DePaolis et al. 2011). We make this prediction in line with the prior finding that the onset of single- and multi-word utterances is not delayed in these communities (Casillas et a. 2020a; b), and despite the fact that VMS is related to early lexical development, which is known to be sensitive to ambient adult CDS rates. Finally, given the differences in phonological inventories between Tseltal and Yélî Dnye, we predict that Yélî children acquire their VMS consonants slightly later than Tseltal children because the Yélî consonant inventory is large, complex, and

acoustically packed (Levinson in press; see also Cristia & Casillas under review). That said, differences based on inventory effects may not emerge with children this young (Jakobson 1968; Vihman & de Boysson-Bardies 1994; Monnin & Lævenbruck 2010).

2. Methods

2.1 The communities

We analyzed a total of 15 hours and 45 minutes of Tseltal and 4 hours and 30 minutes of Yélî Dnye audio recordings for the phonological content of children's spontaneous vocalizations. The data consist of annotated segments from one daylong recording per child. From each recording we extract a single estimate for canonical proportion (i.e., the proportion of vocalizations containing at least some canonical babble) and a single estimate for VMS acquisition.

This approach differs from most other studies on VMS and early babble, where multiple recordings or longitudinal data are used to establish stable evidence of phonological development (e.g., Vihman & de Boysson-Bardies 1994; McCune & Vihman 2001; DePaolis et al. 2011; Majorano et al. 2014; Oller et al. 1995; Lee et al. 2017). Laing and Bergelson (2020) established VMS counts for their participants (mostly) at single timepoints by adopting DePaolis and colleagues' (2011) criterion of 50 productions of a consonant within 30 minutes. They then used automated output from the LENA software accompanying their daylong recordings to identify the top 30-minute clips of infant vocalization from the long (up to 16 hours) recording, manually choosing among these clips and sometimes recombining subparts of them to estimate VMS from 30 minutes of high-quality infant vocalization.

In the remote fieldwork context in which these data were collected, recordings could only be made during a few weeks' visit every 1–2 years, rendering longitudinal data collection impractical. The original dataset aimed to capture one recording from as many children as possible, leaving little time to make multiple recordings with individual children. Lacking longitudinal data, repeated recordings, or automated annotation output, we are forced to adapt the VMS measure to our present dataset. We do, however, use a secondary measure to establish that the resulting estimates reflect stable consonant production (see below). That said, the lack of multiple samples per child, longitudinal or not, increases the likelihood that some of the resulting estimates reflect production patterns that aren't representative of the child's true abilities at that time; thus the patterns we find should be considered at the population level, not at the individual level.

The recordings were collected in 2015 (Tseltal) and 2016 (Yélî Dnye), and can be accessed via the Casillas HomeBank corpus (Casillas et al. 2017). Participant consent processes and data collection were conducted in accordance with ethical guidelines approved by the Radboud University Social Sciences Ethics Committee. We focused on children between 5 and 20 months because children are expected to begin canonical babble production around 6 months, sometimes earlier, and by 20 months nearly all children would be expected to have started producing recognizable words (Oller et al. 1998; McGillion et al. 2017; Casillas et al. 2020a).

The Tseltal speaking children come from a farming community in the highlands of Chiapas in Southern Mexico, where they are typically raised in patrilineally organized, multigenerational households. During the day, infants are carried on their mother's back while she goes about her business, or they are left at home with other family members while the mother works elsewhere (e.g., in the field). The majority of children in this community grow up monolingually until they go to school (Casillas et al. 2020a), and their linguistic environments have been characterized as non-child-centered and non-object-centered (Brown 1998; 2011; 2014). CDS from adults is typically limited until the infants themselves start to

seek out interaction (Brown 2014) and continues to be relatively infrequent through 3;0 (Casillas et al. 2020a). The Tseltal data used in the current study include 20 children (M = 10 months; median = 9; range = 5–19), including four children whose recordings were used in the Casillas et al. (2020a) study.

The Yélî Dnye speaking children live in a collection of small settlements on the northeastern shore of Rossel Island, which is located 250 nautical miles off the south coast of mainland Papua New Guinea. Children grow up in hamlets with patrilocally organized household clusters, where there is often a shared open space between households. During the day, children are carried in their caregivers' arms, and they are frequently passed around between community members—even those from far outside the natal hamlet—who return the child to the mother for feedings (Casillas et al. 2020b; Brown & Casillas in press). Yélî children mostly grow up speaking Yélî Dnye at home, although English, Tok Pisin, and other regional languages are often spoken by adults and school-aged children (Brown & Casillas in press; Casillas et al. 2020b). Children begin to learn English once they start school. The linguistic environment of Yélî children can be characterized as child-centered (Brown & Casillas in press; Ochs & Schieffelin 1984); children are considered a shared responsibility, as well as a source of joy and entertainment for caregivers, and, as such, interaction with infants and young children on Rossel Island is initiated by women, men, girls, and boys alike (Casillas et al. 2020b). The Yélî data we use in the current study include all 12 children in the 5-20-month age range who were reported as acquiring Yélî Dnye monolingually in the 2016 Yélî Dnye Casillas HomeBank corpus (M = 12.4 months; median = 12.5; range = 8–17), including four children whose recordings were used in the Casillas et al. (2020ab) study.

Table 1. Descriptive statistics of the age in months of the children in the present dataset.

Language	N	Mean age in months (Median, SD, range)
Tseltal	20	10 (9, 4.0, 5–19)
Yélî Dnye	12	12.4 (12.5, 2.6, 8–17)

2.2 The data

For both datasets, the recordings were made using an audio recorder (Olympus WS-832 or WS-853) and photo camera (Narrative Clip 1) strapped to the child's chest during most of a waking day at home. For young infants and very small children, the primary caregiver wore the photo camera (see Casillas et al., 2020a for details). The recordings give a natural picture of language use over the course of multiple home activity contexts. We do not use the image data in this study—only the audio recordings.

We base our analyses on a subsample of each child's spontaneous vocalizations from the day. We selected short random clips from each recording, following the random sampling procedure used for the eight children whose recordings were previously annotated for Casillas et al. (2020a) and (2020b). That is, we focused our analyses on spontaneous vocalizations by the target child within nine randomly selected and non-overlapping clips for each recording; clips were 5 minutes long for the Tseltal data and 2.5 minutes long for the Yélî Dnye data. The reason for this disparity in clip duration is that the number of speakers and amount of background noise in the Yélî dataset makes it particularly time consuming to annotate, limiting what can be feasibly accomplished during researcher visits to the island (Casillas et al. 2020b). We take this disparity into account in the analysis, as explained below. Considering that, for Yélî Dnye, we also have fewer recordings from children in the target age range (12 vs. 20 for Tseltal) and fewer recordings from young children (Table 1), the Yélî Dnye data should be regarded as very preliminary, especially at the younger end of the studied age range.

Each target child vocalization in each clip was classified and broadly phonetically transcribed. For the eight recordings with existing annotations, phonetic transcriptions were added to the existing speech annotations (completely independently from their previously added vocal maturity classification). For all other recordings, the child's vocalizations within each clip were first diarized, and subsequently classified and phonetically transcribed. Vocalization annotation followed the scheme shown in Table 2; non-canonical vocalizations were labeled as such when a vocalization was not classifiable as laughing or crying but also did not include canonical babble (e.g., vowel-only vocalizations).

Table 2. Vocalization annotation scheme.

Annotation	Meaning
hamuwa	example of a transcribed vocalization with canonical
	babble
N	non-canonical vocalizations
L	laughing
Y	crying

The first author—a linguist with training in phonetics who lacked previous experience with either Tseltal or Yélî Dnye-transcribed children's canonical babble using the International Phonetic Alphabet (International Phonetic Association, 2020) on the basis of the phoneme inventory of each language, as it is spoken by adults. Instances in which children produced phones that are not present in the adult language were transcribed as perceived by the first author. We note that phone production by infants and very young children is typically limited to a subset of phones (Prather et al. 1975; Ingram 2008; DePaolis et al. 2016) that commonly appear across the world's languages (Moran & McCloy 2019), including simple labial, coronal, and velar oral and nasal stops with occasional fricatives, all of which occur in the languages native and highly familiar to the annotator. It is of course possible that children learning Tseltal and Yélî Dnye produce other phones outside of this set; phonological development data for these languages is highly limited. We were unable to have our annotations completed or corrected by native speakers of each language given the impedance of travel caused by the COVID-19 pandemic. Instead, the second author, who is more familiar with the two languages and is also a linguist with phonetic training, independently annotated a random two clips (22%) of each child's recording data and found high agreement with the first annotator (see Appendix for details).

The phones used in transcribing the Tseltal data are based on Polian's (2013) grammar, the inventory for which is shown in Tables 3 and 4.

Table 3. Tseltal consonants (non-bold, no line), with the additional, non-native phones we found in Tseltal children's spontaneous vocalizations (bold, underlined).

	Bilab	ial	Alveo	dental	Pal	atoalveolar	Velar	Glottal
Plosives	р	b	t	<u>d</u>			k	3
	p'		ť'				k'	
Affricates			ts		tſ	<u>d3</u>		
			ts'		tʃ'			
Fricatives			S		ſ	3	X	h
Nasals	m		n					
Laterals			1					
Rhotics			r					
Approximants	β				j		W	

Table 4. Tseltal vowels (non-bold, no line), with the additional, non-native phones we found in Tseltal children's spontaneous vocalizations (bold, underlined).

	Front	Central	Back
Close	i		u
Close-mid	e	<u>ə</u>	0
Open		a	

The phones used in transcribing the Yélî Dnye data are based on Levinson's (in press) grammar, the inventory for which is shown in Tables 5 and 6. The majority of the phones were not attested in the child vocalizations (e.g., we found no doubly articulated consonants).

Table 5. Yélî Dnye consonants (non-bold and no line) with additional, non-native phones found in Yélî children's spontaneous vocalizations (bold, underlined). SBC stands for 'simultaneous bilabial closure'.

		Bila	babial		Alveolar		Alv	Alv.+SBC		t-Alv.	Po	st-		Velar	
											Alv	-SBC			
		Pal.	Lab.	Both		Pal.		Pal.		Pal.		Pal.		Pal.	Lab.
Plosives	p	\mathbf{p}^{j}	p ^w	p^{jw}	t	t∫	tp	tp ^j	ţ	ţj	ţp	ţp ^j	k g	\mathbf{k}^{j}	k ^w
Prenasalized plosives	mb	mb ^j	mb ^w	mb ^{jw}	nt	ndʒ	nmdb	nmdb ^j	ņḍ	ņḍ ^j	ņmḍb	ņmḍb ^j	ŋg		ŋg ^w
Nasally- released plosives						ſ			ţņ	ţņ ^j	ţpņm	ţpņm ^j	kŋ		kŋ ^w
Nasals	m	m^{j}	m^{w}	m^{jw}	n	\mathbf{n}^{j}	nm	nm ^j	ņ	ņ ^j	ņm	ņm ^j	ŋ		$\mathfrak{y}^{\mathrm{w}}$
Rhotics					<u>r</u>										
Continuants	w~β	β^{j}							j l	l j	1β ^j				

	Vela	r+SBC	Glottal
		Pal.	
k	p	kp ^j	3
ŋm	ıgb		
kp	ŋm		
			<u>h</u>
ŋı	m		
γ	X		

Table 6. Yélî Dnye vowels.

	Front		Central		Back			
	Oral	Nasal	Oral	Nasal	Oral	Nasal		
Close	i	ĩ	i		u	ũ		
Near-close	e		э	õ				
Open-mid	ε	ε̃			э	õ		
(Near-)open	æ		B	ã	a	ã		

2.3 The framework

In order to compare the onset and development of canonical babble produced by Tseltal and Yélî children to each other and also to other children discussed in previous literature (e.g., Lee et al. 2017; Cychosz et al. 2021), we calculated their canonical proportion (CP; see formula

below). To make this calculation, we looked per child at what percentage of their vocalizations included canonical babble. Lee et al. (2017) used a similar measure to quantify the amount of canonical babble, but looked at canonical productions by syllables, a measurement they called the canonical babbling ratio (CBR). In contrast, we looked at canonical babble by whole vocalization, in line with the approach of Cychosz et al. (2021). To calculate CP in the present study, any vocalization containing one or more canonical syllables was counted as an instance of canonical babble, and we subsequently divided the number of canonical babble-containing vocalizations by the total number of vocalizations, excluding vocalizations comprised solely of laughing or crying (Table 2).

Formula for calculating CP:

#voc's with canonical babble

#voc's with canonical babble + #non-cry/non-laugh voc's without canonical babble

We deployed VMS as a measure of consonant development, though we made some adjustments to McCune and Vihman's (2001) original measure in order to accommodate our crosssectional, daylong recording dataset. Our lack of longitudinal data removes our ability to observe individual children's development. However, the daylong recording collection affords us more data per individual and the production setting is highly natural. Laing and Bergelson (2020) define VMS as achieved at a single time point if the child produces 50+ tokens of a phone within the highest 30-minutes of vocal activity in the day. However, without automated annotation output like they had, we were unable to replicate this method. Instead we created a new adaptation of VMS, which examines children's consonant production across the day. We define VMS as follows: If a Tseltal child produced 10+ realizations of a consonant within their total 45 randomly sampled minutes, or if a Yélî child produced 5+ realizations of a consonant within their total 22.5 randomly sampled minutes, then we consider the child to have acquired the VMS for that consonant. A visual aid of this process using hypothetical data is shown in Figure 1. On the basis of vocalization rates in the data from Casillas et al. (2020a) and (2020b) we had anticipated that our random clip sampling would typically yield 100+ vocalizations per child, a lower estimate for what has previously been required to secure stable VMS scores (Vihman et al. 1985; Vihman et al. 1994; Vihman personal communication). As noted below, there were a few children for whom we found fewer than 100 vocalizations in their random clips, namely 1 Yélî Dnye- and 3 Tseltal-acquiring children.

While a departure from prior definitions of VMS, our present measure capitalizes on children's productions over the course of the day, which affords a new perspective on consonant production stability. We verified that our adapted VMS scores highly correlate with a second stability measure: the number of consonants each child produced across four or more of their nine clips ("cross-clip consonants"; see Appendix). In addition to the correlation between VMS and cross-clip consonant production, this secondary measure produces the same pattern of statistical results reported below with VMS. That said, our direct comparisons to previous VMS outcomes in what follows should be taken with a pinch of salt—the best comparison is between these two communities, where the method of data collection and VMS measurement is the same.

Following McCune and Vihman (2001), we collapsed voiced and voiceless variants of produced consonants, as this distinction is often not yet mastered by young children (Eilers et al. 1984) and is also not contrastive in either language. We also only counted supraglottal consonants toward children's VMS score, following prior work: Glottals and glides already occur frequently in the early part of children's first year while supraglottal consonants only begin to occur around 6–8 months (McCune & Vihman 2001). Notably, the majority of non-

native phones attested in the data fall into these two categories (i.e., voiced equivalents of native voiceless phones and subglottal consonants), supporting the omission of these distinctions in the analysis and in related past work.

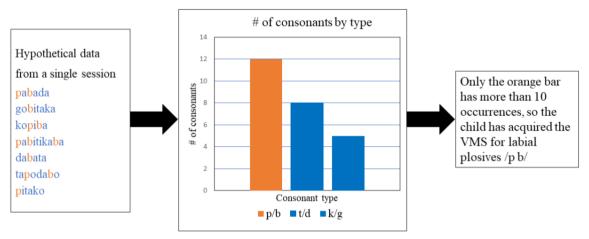


Figure 1. Illustration of how VMSs are derived from the data. Note again that, because of the difference in the amount of annotated data between corpora, we define 5 occurrences of a consonant as a VMS in the case of Yélî Dnye, instead of 10 occurrences as we do for Tseltal.

3. Results

We modeled our two dependent measures—canonical proportion and the number of VMSs produced by each child—using a linear regression with fixed effects of age (numeric, in months), language (Tseltal/Yélî Dnye), and their interaction (i.e., measure ~ age * language). Because we only had one datapoint per child (e.g., the number of VMSs that child displayed), we were unable to include a random effect of child. The analysis was conducted in R (R Core Team, 2018) using lme4 (Bates et al. 2015), and all plots were generated with ggplot2 (Wickham, 2009). We first review findings regarding canonical proportion, then findings for VMS acquisition.

3.1 Canonical proportion

All Tseltal and Yélî speaking children 8 months and older used canonical babbles, consistent with previous findings (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017). Furthermore, and also consistent with other cross-linguistic data, both populations of children had a CP greater than 0.15 after age 0;10 (Lee et al. 2017; Cychosz et al. 2021), where CP less than 0.15 might indicate delay in development (Oller et al. 1995). Differences in CP between the youngest and oldest children in each corpus were apparent (Figure 2), with a gradual increase more apparent among the Tseltal children. Note, however, that the youngest children in our Yélî sample are older than the youngest children in the Tseltal sample, and as such already produce vocalizations with canonical babble quite frequently.

A linear regression of CP revealed a significant positive effect of age (p<.01), no significant effect of language (p=.15) and no age-by-language interaction (p=.15). In other words, while the proportional use of canonical babble increased with age overall, there is no evidence for differing developmental rates per language group, neither overall nor specifically for age-related increase between languages.

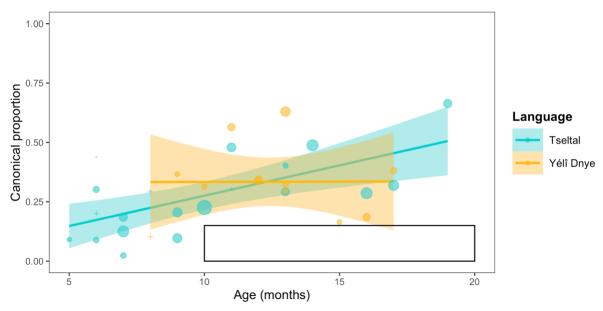


Figure 2. CP of Tseltal and Yélî Dnye speaking children. Scores falling in the white box would be unexpected for typically developing children based on the benchmark of 0.15 CP at age 0;10 (Oller et al. 1995; Lee et al. 2017; Cychosz et al. 2021). Points indicate the number of vocalizations (range: 64–643), and those shaped '+' indicate fewer than 100 vocalizations found.

Table 7. Output of the CP regression analysis.

Coefficients:				
Term	Estimate	Std. Error	t-value	p-value
(Intercept)	0.02	0.082	0.243	0.81
Age in months	0.026	0.026	3.339	<.01
Language	0.311	0.211	1.474	0.152
Age in months:Language	-0.025	0.017	-1.478	0.151

3.2 VMS consonants

Prior work suggests that American and British English-acquiring children typically acquire 2 VMS consonants between 9 and 154 months, on average reaching this benchmark between 10 and 12 months (McCune & Vihman 2001; DePaolis et al. 2016). We see that most Tseltal and Yélî Dnye children showed VMS scores aligning with results based on English, though 1 Tseltal-learning child and 3 Yélî Dnye-learning produced fewer VMS consonants than expected on this basis after 12 months (Figure 3). These findings are not necessarily anomalous: in McCune & Vihman's (2001) study, 6 out 20, or 30%, of the English children did not reach 2 VMS before 14 months, which is comparable to the 27.3% of Tsetal and 45.4% of Yêlî children found here. In contrast, in the study by DePaolis et al. (2016) only 2 out of 59, or 3.4%, of the children did not achieve 2 VMSs before 14 months. An overview of all VMS consonants found in the Tseltal and Yélî Dnye data is shown in Table 8 (the more detailed data used to generate Figure 3 and Table 8 are shown in Appendix Table 1 and Appendix Table 2).

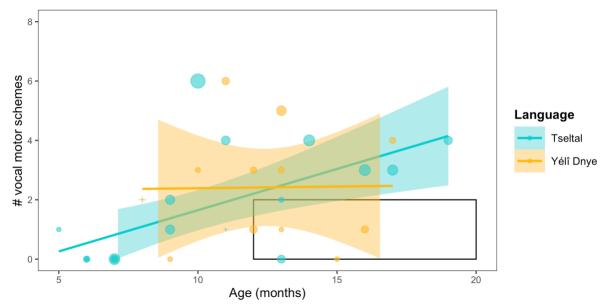


Figure 3. VMS count of Tseltal and Yélî Dnye speaking children; the white box indicates <2 VMSs beyond 12 months. Points indicate number of vocalizations (range: 64–643), and those shaped '+' indicate fewer than 100 vocalizations found.

In both Tseltal and Yélî Dnye, all phones reaching VMS were part of the native phonological inventory, voicing contrasts aside. The most common VMS consonants produced by children in both samples were [t/d] and [m], followed by [p/b] (Table 8). The [m] VMS is more prevalent here than in prior studies, while [p/b] and [t/d] are well-attested early VMS consonants in English, Italian, and Welsh (DePaolis et al. 2011; Majorano et al. 2014; McCune & Vihman 2001). The difference in distribution of the plosives across these two languages shows an interesting pattern: While in both cases [t/d] is the most prevalent type of plosive (in line with the hypothesis that the alveolar place of articulation is universally less marked, (Paradis & Prunett 1991; Tsuji et al. 2015), in Tseltal the second most prevalent is [p/b] with few instances of [k/g], while in Yélî Dnye [k/g] is the second most prevalent plosive type and there are few instances of [p/b]. The graphs for the total number of phones used across tokens analyzed in each language are shown in Appendix Figure 4.

Table 8. Distribution of all VMS consonants acquired by language sample.

	p/b	t/d	k/g	m	n	1
Tseltal	7	8	2	8	5	3
Yélî Dnye	3	7	6	7	2	4

A linear regression of VMS counts revealed a significant positive effect of age (p<.01), no significant effect of language (p=.20) and no age-by-language interaction (p=.21). In other words, while the number of VMSs acquired increased with age overall, there is no evidence for different patterning by language group, neither overall nor specifically for age-related increase between languages.

Table 9. Output of the VMS regression analysis.

Coefficients:				
Term	Estimate	Std. Error	t-value	p-value
(Intercept)	-1.128	1.002	-1.125	0.27

Age in months	0.278	0.093	2.987	<.01
Language	3.404	2.565	1.327	0.195
Age in months:Language	-0.266	0.208	-1.281	0.211

4. Discussion

Children living in the Tseltal and Yélî communities are infrequently directly spoken to by adults, yet prior work shows no apparent delay in their early linguistic development (Casillas et al. 2020a; b). This finding seems counterintuitive, as adult caregiver CDS has previously been strongly associated with faster-growing receptive and productive vocabularies (Hart & Risley 1995; Hoff 2003; Shneidman & Goldin-Meadow 2012; Weisleder & Fernald 2013). This apparent discrepancy may lie in the fact that prior work on Tseltal and Yélî Dnye used vocal maturity measures that are, in fact, fairly robust to environmental variation. In the current study, we investigated whether there was evidence of delay on a measure of early phonological development that has been shown to relate to early lexical development—stable early consonant production—via an adaptation of traditional vocal motor scheme measures. We also tested whether the prior finding of non-delayed canonical babble onset would hold up with a larger sample of children than was studied previously.

We predicted that, replicating prior work, canonical babble development would show no delays in either corpus (Casillas et al. 2020a; b; Cychosz et al. 2021). Following prior results showing no delays in the emergence of single- and multi-word utterances (Casillas et al. 2020a; b), we further hypothesized that VMS acquisition by Yélî and Tseltal children would be on par with previous results from English-acquiring children, who typically acquire at least 2 VMS consonants around age 10 to 12 months, almost certainly doing so by 15 months. We also tentatively predicted that Yélî children would acquire their VMS consonants slightly later than Tseltal children because the Yélî consonantal inventory is large, complex, and acoustically packed (Levinson in press).

4.1 Canonical babbling

Prior work has shown that most children start using canonical babble-after 8 months, with canonical babble making up at least 0.15 of their total syllables by 0;10 (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021). Using a related measure, canonical proportion (CP), we found that both Tseltal and Yélî children have indeed already surpassed this benchmark of 0.15 at 7 months, in line with the cross-linguistic findings of Cychosz et al. (2021). This is significantly earlier than the 10 months reported by Lee et al. (2017), which might be because CP is a less fine-grained measure than the canonical babbling ratio they use (CBR). These results are consistent with prior work on these populations (Casillas et al. 2020a; b) and from other language communities, despite the relatively infrequent rate of directed speech, overall and from adult speakers in particular, in the present communities of study.

4.2 Vocal motor schemes

Prior work on typically developing children learning English suggests that children typically use two stable consonants in their pre-lexical babble (i.e., vocal motor schemes) beginning between 10 and 12 months of age, almost certainly reaching this milestone by 15 months (McCune & Vihman 2001; DePaolis et al. 2016). Most of the Tseltal and Yélî Dnye-learning children followed this pattern, and the consonants qualifying as having reached VMS status are consistent with past work (i.e., [t/d] > [p/b], [k/g] > [m], [n], [l]). Importantly, 1 Tseltal-learning child (1;1) and 3 Yélî Dnye-learning children (1;1, 1;3, and 1;4) produced fewer than two VMS consonants after 12 months (27.3% and 45.4% of those samples, respectively). Our review of individual variation among typically developing children in the prior studies on English

suggests that this pattern comparable to between-individual variation reported by McCune and Vihman (2001; 30% of the sample) but represents a larger than expected share of the sample with low VMS counts compared to DePaolis et al. (2016; 3.4% of the sample). Because our adapted VMS measure and our cross-sectional data differ from these past studies on English, direct comparison of this milestone is noisy at best. However, we take the overall similarity in patterning across age as providing no support for the idea that these children, who hear less adult CDS, are delayed in their early linguistic development, in this case with a pre-lexical measure that has been shown to connect to later lexical development.

Our best comparisons can be made between these two communities because the methods of data collection, sampling, and VMS estimation were the same. We predicted, on the basis of differences in the complexity and acoustic packedness of the languages' phonological inventories, that Yélî Dnye-acquiring children might have a slightly slower development of VMS consonants. While relatively more Yélî children than Tseltal children did not reach the 2-VMS milestone by 12 months, we found no statistical evidence for an effect of language on VMS consonant counts, nor any interaction effect of language and age on VMS consonant counts. While the inventory of early VMS consonant types was similar between the two language groups, we saw preliminary evidence for differences in the distribution of those consonant types (i.e., the prevalence of labials for Tseltal speakers and velars for Yélî Dnye speakers). Cross-linguistic differences in VMS consonants have been documented in prior work (DePaolis et al. 2011; Majorano et al. 2014; McCune & Vihman 2001), and future work should capitalize on cross-linguistic datasets to more closely examine what drives these variations, if not frequency of those phones in the child's input (DePaolis et al. 2011; Majorano et al. 2014).

4.3 Infrequent CDS and phonological development: canonical babble vs vocal motor schemes

Keeping in mind that (a) our results are preliminary, (b) some of our VMS scores are based on fewer vocalizations than hoped for, and (c) 2 VMS consonants by 10–12 months is a less robustly attested benchmark than the one we use for canonical babble (particularly with respect to daylong recording data) and is not in any way related to clinical language delay, we now tentatively discuss the possibility that VMS is indeed sensitive to environmental variation and that, therefore, some children are likely to show slower VMS development in these two communities where adult CDS is relatively less frequent.

Prior research suggests that canonical babble onset is more a meaningful benchmark for motor development than it is for language development (Vihman et al. 2009). Under this view, canonical babble is a hallmark of the development of rhythmic motor tools that develop in the first year of life (Iverson, Hall, Nickel & Wozniak 2007), and develops as a result of both proprioceptive and auditory experience with syllable production (Westermann & Miranda 2004; Guenther & Vladusic 2012), but doesn't itself require stable articulatory or phonological representations. This perspective—that canonical babble demonstrates a motor skill that helps prepare the child for language, rather than an early step in language development—aligns with findings showing its cross-cultural and cross-linguistic developmental robustness (Oller 1980; Oller et al. 1998; Oller 2000; Lee et al. 2017; Cychosz et al. 2021).

In contrast, VMS gives us insight into early, stable consonant productions that approximate consonant categories in the ambient language (Vihman 1993), and so may be expected to link more tightly to early linguistic representations and, thereby, early productive vocabulary. Over the first year of life, we know that children's ambient language environment comes to shape how they perceive speech (Werker & Tees 1984; Monnin & Lœvenbruck 2010). In fact, adult CDS (and not just ambient speech in general) has been proposed to facilitate early phone discrimination (Kuhl et al. 2003; Kuhl 2007), and may thereby help children acquire

phonological categories sooner. In production, we can also imagine that the elicitation of childproduced speech by caregivers engaging in interactive CDS may result in children getting more frequent practice in targeting more adult-like phonological forms (Kuhl & Meltzoff 1996; Kuhl 2007). If we understand VMS as reflecting something about the initial stabilization of phonological categories rather than a simple practiced motor skill—e.g., via the coupling of articulatory parameters and auditory perception (Westermann & Miranda 2004)—we can predict that it is sensitive to the child's exposure to adult CDS and, thereby, linked to early lexical development (Vihman 1993; McCune & Vihman 2001; McGillion et al. 2017), which is also sensitive to adult CDS. We do not find evidence for this possibility in the present dataset, suggesting that VMS, like CP, may be fairly robust to environmental variation despite its relationship with early lexical productions. The findings raise multiple questions: at what point does environmentally driven variation in early production milestones begin to emerge?; is it limited to lexical phenomena?; and is the scope of variation primarily within and not across populations? Investigations addressing these questions shed new light on the unresolved question of how it is that children, in such variable developmental contexts, come to acquire the linguistic representations and language practices appropriate to their local communities.

Returning to the issue of input quantity, we note that total linguistic input in these communities is not at all sparse: On the basis of past work by Casillas and colleagues (2020a; b), we can say that there is a great deal of other ambient speech present in these two communities for children to learn from beyond adult CDS, much of it directed to other children within earshot of the target children or to the target children from other children. While direct comparisons of VMS counts between the present work and past work are limited by different measurement and data collection approaches, the patterns of VMS count were also similar between the two communities studied here, which have *parallel* data collection and measurement approaches. Because children in these two communities hear differing quantities of CDS from adults vs. children (Casillas et al. 2020a; b; Bunce et al. under review), we take the comparative findings between these two communities as evidence that VMS acquisition is fairly robust to variation in input type (adult CDS, child CDS, and overhearable other speech).

While we can only draw very limited conclusions with the present data, we hope that future work will investigate stable consonant production in more language communities with diverse phonologies and caregiving contexts; together, e.g., in a meta-analysis, these studies could do much to illuminate the role of linguistic input in shaping the development of stable pre-lexical consonant production and, later, lexical development. For example, if future work consistently finds earlier VMS acquisition in linguistic communities where children experience not just a lot of linguistic input, but a lot of directed linguistic input from adults, it would stand as evidence that mature speech directed to the target child, and not exposure to speech in general (nor CDS in general), critically shapes phonological development in this transitional period between babble and first words. If no such pattern emerges, it would suggest early robustness to environmental variation in this transitional period, pinpointing lexical development as deeply different from phonological development in its sensitivity to linguistic input, raising questions about what mechanisms would drive such a difference. Of course, until we gain greater clarity on the relationship between VMS and language environment from further work along the lines suggested above, the current findings should only be taken as preliminary.

4.4 Further research

The current study has a number of important limitations. While our sample size is an improvement over previous work done in these communities (Casillas et al. 2020a; b), it is still quite limited considering our cross-sectional design and broad age span (0;5–1;8). Future work should aim to gather larger and more symmetrical samples across languages than we managed here in order to more comparably explore within-population variability and age-based

differences. We were also unable to achieve at least 100 spontaneous vocalizations for every child with our random sampling technique, which means some individual estimates may not be as stable as hoped. Relatedly, our lack of longitudinal data or multiple recordings per child may have yielded less stable individual estimates of canonical babbling and VMS than those presented in prior work (though see our second measure of stable consonant production in the Appendix). Because we had no measure of CDS input rate for each child in the study, we were also only able to compare individual production patterns to community-wide CDS input patterns. Individual input rate estimates would allow us to investigate whether VMS production relates to directed input *within* each community, regardless of the benchmark for English and other languages.

Finally, while phonetically trained, our transcriber did not speak the target languages and so did not benefit from the contextual and lexical information that could have rendered the transcriptions closer to what would be heard by a native annotator. Future work linking VMS to individualized input rate estimates and vocabulary size in these and comparably designed corpora from other languages and cultural communities are needed to clarify the results presented in the current study.

5. Conclusion

We find that, Tseltal and Yélî Dnye-learning children show a similar pattern of CP use and VMS acquisition, compared to each other and compared to past work on English-learning children. Although some children did not reach the typical milestone, they represented a comparable share to one prior study of English (and a larger-than-expected share of another). In addition to replicating prior work on CP development with a larger and independently annotated sample, we take the current findings as providing no support for the idea that these children, who hear less adult CDS than English-learning children, are delayed in their early linguistic development. Future work should continue to examine stable consonant production in diverse developmental environments, and should ideally also more closely examine individual differences within each community, including any links to the lexical development of the same children. Such data will be key to illuminating the ways in which children's linguistic environments influence their transition from pre-lexical to lexical productions cross-linguistically.

Acknowledgements

We are deeply grateful to the participating communities, families, and children represented in these datasets. We thank Rebeca Guzmán López, Humbertina Gómez Pérez, Juan Méndez Girón, Taakêmê Ńamono, Ndapw:éé Yidika, and Y:aaw:aa Pikuwa for their help in corpus creation and initial annotation. We also thank the National Research Institutes of PNG and Milne Bay Province administration for their support. This work was funded by an NWO Veni Innovational grant to MC (275-89-033). Last but not least, we thank Dr. Paula Fikkert for her supervision and support.

Appendix

Table 1. Overview of VMS consonant use in the Tseltal dataset.

Aclew ID	Age in months	p/b	t/d	k/g	m	f	n	1	r	s/z	J/3	ŋ	x/y	Total VMS	Total consonants	Total vocalizations	Total cross- clip	tʃ/dʒ	W	j	3	h
6964	5	_	10	_	_	_	_	_	_	_		_		1	28	100	consonants 0	_	_	_		_
4935	6	_	-	_	_	-	-	-	-	_	-	-	-	0	33	54	0	_	_	_	-	19
6107	6	-	-	-	-	-	-	-	-	-	-	-	-	0	92	152	0	-	11	-	23	53
1991	6	-	-	-	-	-	-	-	-	-	-	-	-	0	34	81	1	-	-	-	-	-
8179	6	-	-	-	-	-	-	-	-	-	-	-	-	0	19	127	0	-	-	-	-	-
1188	7	-	-	-	-	-	-	-	-	-	-	-	-	0	61	233	0	-	15	-	-	39
456	7	-	-	-	-	-	-	1	-	-	-	-	-	0	30	381	0	-	-	-	1	-
2109	7	-	-	-	-	-	-	-	-	-	-	-	-	0	6	128	0	-	-	-	-	-
3214	8	-	-	-	10	-	10	-	-	-	-	-	-	2	34	68	0	-	-	-	-	-
7273	9	14	-	-	-	-	-	-	-	-	-	-	-	1	44	279	0	-	-	-	-	14
7439	9	-	10	-	11	-	-	-	-	-	-	-	-	2	80	284	5	-	-	-	-	14
3591	10	15	31	40	12	-	21	11	-	-	-	-	-	6	223	643	5	-	13	20	-	52
909	11	10	-	-	-	-	-	-	-	-	-	-	-	1	46	61	1	-	10	-	-	14
8787	11	64	107	14	16	-	-	-	-	-	-	-	-	4	260	260	4	-	15	13	-	10
684	13	21	-	-	14	-	-	-	-	-	-	-	-	2	85	116	0	-	-	-	-	23
9415	13	-	-	-	-	-	-	-	-	-	-	-	-	0	82	218	1	-	-	-	14	51
7326	14	17	120	-	-	-	86	52	-	-	-	-	-	4	374	392	5	-	15	23	14	26
6028	16	-	15	-	74	-	11	-	-	-	-	-	-	3	183	391	6	-	10	-	10	26
5147	17	-	24	-	18	-	-	15	-	-	-	-	-	3	198	335	4	-	-	16	-	89
9592	19	32	107	-	35	-	21	-	-	-	-	-	-	4	266	235	5	-	11	12	-	20

Table 2. Overview of VMS consonant use in the Yélî Dnye dataset.

Aclew ID	Age in months	p/b	t/d	k/g	m	f	n	1	r	s/z	J/3	ŋ	x/y	Total VMS	Total consonants	Total vocalizations	Total cross- clip	tʃ/dʒ	W	j	3	h
																	consonants					
1143	8	-	6	-	-	-	-	13	-	-	-	-	ı	2	24	99	0	-	-	-	ı	-
8018	9	-	-	-	-	-	-	-	-	-	-	-	ı	0	43	109	0	-	-	6	ı	21
8274	10	-	15	-	22	-	-	19	-	-	-	-	1	3	81	118	1	-	-	-	ı	11
5623	11	22	28	5	54	-	20	11	-	-	-	-	ı	6	209	197	6	-	23	21	5	17
3446	12	-	-	-	50	-	-	-	-	-	-	-	ı	1	91	208	1	-	-	-	ı	29
4624	12	-	16	6	-	-	5	-	-	-	-	-	-	3	72	153	1	-	-	8	-	31
2977	13	-	10	10	5	-	-	-	-	-	-	-	-	3	55	148	1	-	-	6	-	14
5139	13	27	40	11	8	-	-	12	-	-	-	-	-	5	384	303	5	-	83	23	12	16
																						5
7035	13	-	-	-	9	-	-	-	-	-	-	-	-	1	33	104	1	-	-	-	-	14
1504	15	-	-	-	-	-	-	-	-	-	-	-	-	0	15	107	0	-	-	-	-	12
9562	16	-	-	7	-	-	-	-	-	-	-	-	-	1	43	193	0	-	10	-	-	13
6934	17	21	19	9	14	-	-	-	-	-	-	-	-	4	114	146	4	-	16	-	15	13

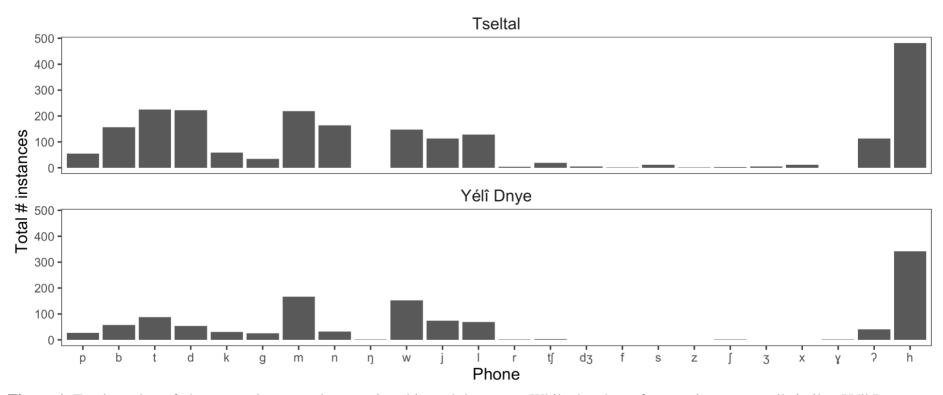


Figure 1. Total number of phones used across tokens analyzed in each language. While the phone frequencies are overall similar, Yélî Dnye appears to have an overrepresentation of /m/ and /w/, and underrepresentation of /?/ compared to Tseltal.

Reliability annotations.

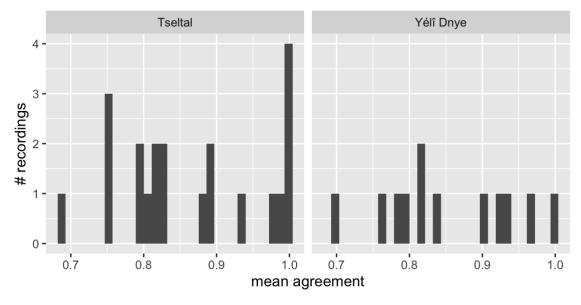
The second author, who has basic working field knowledge of the two languages of study and is a trained linguist, randomly selected two of the nine annotated clips for all recordings in the dataset and independently annotated each vocalization using the same workflow as the first author

(i.e., phonetic transcription for vocalizations with canonical babble). Her annotations only differed from his in that she counted all vocalizations without canonical babble as "N" and she didn't pay close attention to vowel transcription, which is not analyzed here.

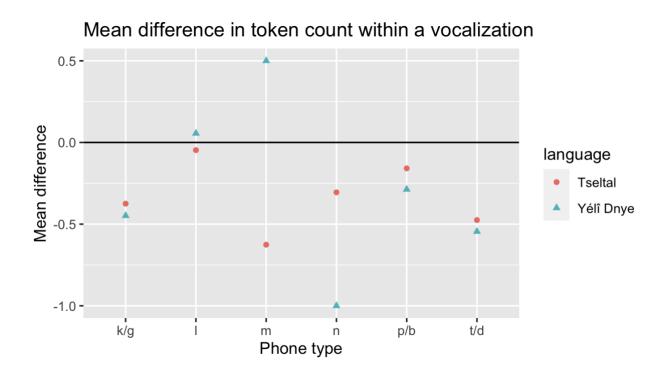
This reliability process resulted in 27,496 vocalizations over which we can make reliability estimates. We focus our analysis here on two facets of this reliability dataset: (1) canonical babble vs. other and (2) the subset of specific VMS consonants that appeared in the main analysis (p/b, t/d, k/g, m, n, and l):

- Canonical babble vs. other: Agreement of whether a vocalization contains canonical babble
 - Mean agreement over all recordings was 86% (Tseltal: 87%, Yélî Dnye: 85%)
 - On average, the first annotator was more likely to label vocalizations as having canonical babble compared to the second annotator (resulting in an average of 3.9% more vocalizations per recording transcribed than the second annotator would have done)

canonical babble vs. other



- VMS phone agreement: Difference in the number of tokens of a given phone (e.g., "m") that appear in a vocalization, only including vocalizations where at least one annotator indicated that the phone was used (i.e., not inflating agreement by including vocalizations in which neither annotator indicated the phone was present).
 - \circ p/b
 - 1166 vocalizations were indicated by one or both coders to contain a p or b phone. Average difference was -0.21 tokens of a p or b consonant in a vocalization, meaning that the first author very slightly undercounted p/b tokens relative to the first author (Tseltal: -0.16, Yélî Dnye: -0.29)
 - \circ t/d
- 1785 vocalizations were indicated by one or both coders to contain a t or d phone. Average difference was -0.5 tokens of a t or d consonant in a vocalization, meaning that the first author slightly undercounted t/d tokens relative to the first author (Tseltal: -0.48, Yélî Dnye: -0.55)
- \circ k/g
 - 684 vocalizations were indicated by one or both coders to contain a t or d phone. Average difference was -0.4 tokens of a k or g consonant in a vocalization, meaning that the first author slightly undercounted k/g tokens relative to the first author (Tseltal: -0.37, Yélî Dnye: -0.45)
- \circ m
- 1679 vocalizations were indicated by one or both coders to contain an m phone. Average difference was -0.25 tokens of an m consonant in a vocalization, meaning that the first author overall slightly undercounted m tokens relative to the first author. However when broken down by language we see that he undercounted for Tseltal but overcounted for Yélî Dnye (Tseltal: -0.63, Yélî Dnye: 0.5)
- \circ n
- 1299 vocalizations were indicated by one or both coders to contain an n phone. Average difference was -0.45 tokens of an n consonant in a vocalization, meaning that the first author overall slightly undercounted n tokens relative to the first author, with the substantial average difference for Yélî Dnye of one whole token on average (Tseltal: -0.31, Yélî Dnye: -1)
- 0 1
- 1310 vocalizations were indicated by one or both coders to contain an l phone. Average difference was -0.01 tokens of an l consonant in a vocalization, meaning that there was near perfect agreement (Tseltal: -0.05, Yélî Dnye: 0.06)
- Overall, these results suggest that the primary annotator overall slightly undercounted tokens of the phones that resulted in VMS counts relative to the second annotator (see visual summary below).



In sum, there was very good overall agreement between the two coders, both in distinguishing canonical babble from other vocalization types and in phonetically transcribing the consonants present.

Cross-clip consonant production.

To check whether our adapted VMS estimates likely reflect children's stable consonant production, we leverage our day-long data to build a second, novel measure of stable consonant production: cross-clip consonant production. The logic of this secondary measure is that, if children stably produce a consonant, they will be found to do so spontaneously across much of their waking day. Because our clips are derived from random points throughout a very long (typically 8–11-hour) audio recording, we could create a fairly strict estimate in hopes of reconstructing which consonants children consistently produce across multiple contexts (vs. those they temporarily produce or produce in limited contexts). A consonant was defined as having cross-clip stability if it appeared at least once in at least four

of the nine annotated clips. Using this definition, we counted the number of cross-clip consonants per child (Appendix Tables 1 and 2). This secondary measure of stability highly correlates with our adapted VMS score (Pearson's r(30) = 0.82 [0.66, 0.91], p = <.001) suggesting that the primary analysis, based on VMS, likely reflects information about children's stable early consonant productions. Further, when we run the same statistical models as reported for VMS in the main text, only using cross-clip consonant count as our dependent measure rather than VMS, we find qualitatively identical results as we do our VMS measure (Appendix Table 3). We take this analysis as encouraging evidence that the adapted VMS scores and analyses reported in the main text do reflect children's early consonant productions, despite our limited sample and departure from past VMS-estimation methods.

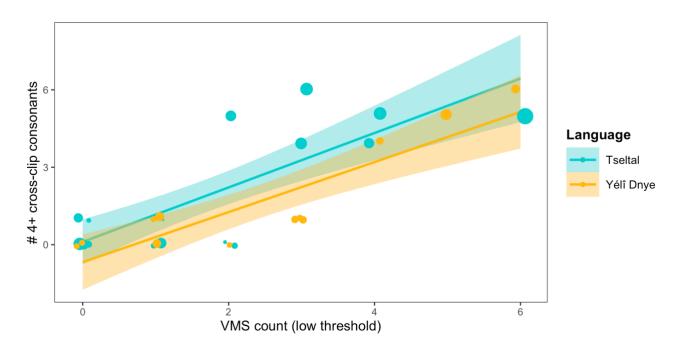


Table 3. Output of the cross-clip count regression analysis (same model structure as the VMS regression).

Coefficients:										
Term	Estimate	Std. Error	t-value	p-value						
(Intercept)	-2.101	1.127	-1.863	0.073						
Age in months	0.395	0.105	3.777	<.001						
Language	2.13	2.884	0.739	0.466						
Age in months:Language	-0.263	0.234	-1.125	0.270						

References

- Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1). 1–48. doi:10.18637/jss.v067.i01.
- Brown, Penelope. 1998. Conversational structure and language acquisition: The role of repetition in Tzeltal adult and child speech. *Journal of Linguistic Anthropology*. 8(2). 197–221.
- Brown, Penelope. 2011. The cultural organization of attention. In Alessandro Duranti, Elinor Ochs & Bambi B. Schieffelin (eds.), *Handbook of language socialization*, 29–55. Malden, MA: Wiley-Blackwell.
- Brown, Penelope 2014. The interactional context of language learning in Tzeltal. In Inbal Arnon, Marisa Casillas, Chigusa Kurumada & Bruno Estigarribia (eds.), *Language in interaction: Studies in honor of Eve V. Clark*, 51–82. Amsterdam, NL: John Benjamins.
- Brown, Penelope & Marisa Casillas. In press. Childrearing through social interaction on Rossel Island, PNG. In Alicia J. Fentiman & Mary Goody (eds.), *Esther Goody revisited:* Exploring the legacy of an original inter-disciplinarian. New York, NY: Berghahn.
- Casillas, Marisa, Penelope Brown & Stephen C. Levinson. 2020a. Early language experience in a Tseltal Mayan village. *Child Development* 91(5). 1819–1835.
- Casillas, Marisa, Penelope Brown & Stephen C. Levinson. 2020b. Early language experience in a Papuan village. *Journal of Child Language*, 48(4). 792–814.
- Casillas, Marisa, Penelope Brown & Stephen C. Levinson. 2017. *Casillas HomeBank corpus*.
- Chavajay, Pablo & Barbara Rogoff. 1999. Cultural variation in management of attention by children and their caregivers. *Developmental Psychology* 35(4). 1079–1090. doi:10.1037/0012-1649.35.4.1079.
- Cristia, Alejandrina, Marisa Casillas. Under review. Non-word repetition in children learning Yélî Dnye.
- Cychosz, Margaret, Alejandrina Cristia, Elika Bergelson, Marisa Casillas, Gladys Baudet, Anne S. Warlaumont, Camila Scaff, Lisa Yankowitz & Amanda Seidl. 2021. Vocal development in a large-scale cross-linguistic corpus. *Developmental Science*, 24:e13090.
- de Boysson-Bardies, Bénédicte & Marilyn May Vihman. 1991. Adaptation to language: Evidence from babbling and first words in four languages. *Language*. 67(2). 297–319.
- DePaolis, Rory A., Marilyn May Vihman & Tamar Keren-Portnoy. 2011. Do production patterns influence the processing of speech in prelinguistic infants?. *Infant Behavior and Development* 34(4). 590–601.
- DePaolis, Rory A., Tamar Keren-Portnoy, & Marilyn May Vihman. 2016. Making sense of infant familiarity and novelty responses to words at lexical onset. *Frontiers in psychology* 7. 715.
- Eberhard, David. M., Gary. F. Simons & Charles D. Fennig (eds.). 2020. *Ethnologue: Languages of the World.* Twenty-third edition. Dallas, Texas: SIL International. Online version: www.ethnologue.com.ru.idm.oclc.org.
- Eilers, Rebecca E., D. Kimbrough Oller & Carmen R. Benito-Garcia. 1984. The acquisition of voicing contrasts in Spanish and English learning infants and children: A longitudinal study. *Journal of Child Language* 11(2). 313–336.
- Fikkert, Paula & Claartje Levelt. 2008. How does Place fall into Place?. *Contrast in phonology: Theory, perception, acquisition* 231–268.
- Gaskins, Susan & Ruth Paradise. 2010. Learning through observation in daily life. In David F. Lancey, John Bock & Susan Gaskins (eds.), *The anthropology of learning in childhood*, 85–117. Walnut Creek, CA: Rowman AltaMira Press.

- Guenther, Frank H. & Tony Vladusich. 2012. A neural theory of speech acquisition and production. *Journal of neurolinguistics* 25(5). 408–422.
- Hart, Betty & Todd R. Risley. 1995. *Meaningful differences in the everyday experience of young American children*. Baltimore, MD: Paul H. Brookes.
- Hoff, Erika. 2003. The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development* 74(5) 1368–1378
- International Phonetic Association. 2020. doi: www.internationalphoneticassociation.org/content/full-ipa-chart
- Ingram, David. 2008. Cross-Linguistic Phonological Acquisition. *The handbook of clinical linguistics*. 626–640
- Iverson, Jana M., Amanda J. Hall, Lindsay Nickel. & Robert H. Wozniak. 2007. The relationship between reduplicated babble onset and laterality biases in infant rhythmic arm movements. *Brain and Language* 101(3). 198–207.
- Jakobson, Roman. 1968. Child language, aphasia and phono-logical universals. The Hague, The Netherlands: Mouton.
- Kidd, Evan & Rowena Garcia. 2021. How diverse is child language acquisition?. *PsyArXiv*. September 6. doi:10.31234/osf.io/jpeyq.
- Kuhl, Patricia K. & Andrew N. Meltzoff. 1996. Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America* 100(4). 2425–2438.
- Kuhl, Patricia K., Feng-Ming Tsao & Huei-Mei Liu. 2003. Foreign-language experience in infancy: Effects of short-term exposure and social interaction on phonetic learning. *Proceedings of the National Academy of Sciences* 100(15). 9096–9101.
- Kunnari, Sari. 2003. Consonant inventories: a longitudinal study of Finnish-speaking children. *Journal of multilingual communication disorders* 1(2) 124–131.
- Laing, Catherine & Elika Bergelson. 2020. From babble to words: Infants' early productions match words and objects in their environment. *Cognitive Psychology* 122.
- Lee, Chia-Cheng, Yuna Jhang, George Relyea, Li-mei Chen & D. Kimbrough Oller. 2017. Babbling development as seen in canonical babbling ratios: A naturalistic evaluation of all-day recordings. *Infant Behavior and Development* 50. 140-153.
- Lee, Sue Ann S., Barbara Davis & Peter MacNeilage. 2010. Universal production patterns and ambient language influences in babbling: A cross-linguistic study of Korean-and English-learning infants. *Journal of Child Language* 37(2). 293–318.
- Levinson, Stephen C. In press. A grammar of Yéli Dnye, the Papuan language of Rossel Island.
- Majorano, Marinella, Marilyn May Vihman & Rory A. DePaolis. 2014. The relationship between infants' production experience and their processing of speech. *Language Learning and Development* 10(2). 179–204.
- McCune, Lorriane & Marilyn May Vihman. 2001. Early phonetic and lexical development. *Journal of Speech, Language, and Hearing Research* 44. 670–684.
- McLeod, Sharynne & Kathryn Crowe. 2018. Children's consonant acquisition in 27 languages: A cross-linguistic review. *American Journal of Speech-Language Pathology* 27(4). 1546–1571.
- McGillion, Michelle, Jane S. Herbert, Julian Pine, Marilyn May Vihman, Rory A. DePaolis, Tamar Keren-Portnoy & Danielle Matthews. 2017. What paves the way to conventional language? The predictive value of babble, pointing, and socioeconomic status. *Child Development* 88(1) 156–166.

- Monnin, Julia & Hélène Lœvenbruck. 2010. Language-specific influence on phoneme development: French and Drehu data. *Proceedings of Interspeech 2010* 1882-1885. Makuhari, Japan, September 26–30, 2010.
- Moran, Steven & Daniel McCloy (eds.) 2019. PHOIBLE 2.0. Jena: Max Planck Institute for the Science of Human History. (Available online at http://phoible.org, Accessed on 2021-08-13.)
- Ochs, Elinor & Bambi B. Schieffelin. 1984. Language acquisition and socialization: Three developmental stories and their implications. *Culture theory: Essays on mind, self, and emotion* 276–322. Cambridge University Press.
- Oller, D. Kimbrough. 1980. The emergence of the sounds of speech in infancy. *Child Phonology* 1, 93–112.
- Oller, D. Kimbrough. 2000. *The emergence of the speech capacity*. Lawrence Erlbaum Associates.
- Oller, D. Kimbrough, Sharyse Levine, Alan B. Cobo-Lewis, Rebecca E. Eilers & Barbara Zurer Pearson. 1998. Vocal precursors to linguistic communication: How babbling is connected to meaningful speech. *Exploring the speech-language connection* 8. 1–25.
- Oller, D. Kimbrough, Rebecca E. Eilers, Devorah Basinger, Michelle L. Steffens & Richard Urbano. 1995. Extreme poverty and the development of precursors to the speech capacity. *First Language* 15(44). 167–187.
- Paradis, Carole, & Jean-François Prunet. 1991. The special status of coronals: Internal and external evidence. *Phonetics and Phonology* 2. 125-57.
- Polian, Gilles. 2013. *Gramática del tseltal de Oxchuc*. Centro de Investigaciones y Estudios Superiores en Antropología Social, México.
- Prather, Elizabeth M., Dona Lee Hedrick, Carolyn A. Kern 1975. Articulation development in children aged two to four years. *Journal of Speech and Hearing Disorders* 40(2). 179-191.
- R Core Team. 2018. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/
- Rogoff, Barbara, Jayanthi Mistry, Artin Göncü & Christine Mosier. 1993. Guided participation in cultural activity by toddlers and caregivers. *Monographs of the Society for Research in Child Development*, 58(8). I–179. doi:10.2307/1166109.
- Shneidman, Laura A. & Susan Goldin-Meadow. 2012. Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science* 15(5). 659–673.
- Tsuji, Sho, Reiko Mazuka, Alejandrina Cristia & Paula Fikkert. 2015. Even at 4 months, a labial is a good enough coronal, but not vice versa. *Cognition* 134. 252–256.
- Vihman, Marily May. 1993. Vocal motor schemes, variation and the production-perception link. *Journal of Phonetics* 21. 163–169.
- Vihman, Marilyn May & Bénédicte de Boysson-Bardies. 1994. The nature and origins of ambient language influence on infant vocal production and early words. *Phonetica*, 51(1–3). 159–169.
- Vihman, Marilyn May, Rory A. DePaolis & Tamar Keren-Portnoy. 2009. A dynamic systems approach to babbling and words. *British Journal of Psychology* 108(1). 1–27.
- Vihman, Marilyn May, Marlys A. Macken, Ruth Miller, Hazel Simmons & Jim Miller. 1985. From babbling to speech: A re-assessment of the continuity issue. *Language* 61(2). 397–445.
- Vihman, Marily May, Edwin Kay., Bénédictie de Boysson-Bardies, Catherine Durand, & Ulla Sundberg. 1994. External sources of individual differences? A cross-linguistic

- analysis of the phonetics of mothers' speech to 1-yr-old children. *Developmental Psychology*, 30(5), 651–662.
- Wickham, Hadley. 2009. *Ggplot2: Elegant graphics for data analysis*. New York, NY: Springer-Verlag. Retrieved from http://ggplot2.org
- Werker, Janet F. & Richard C. Trees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development* 7(11). 49–63.
- Westermann, Gert & Eduardo Reck Miranda. 2004. A new model of sensorimotor coupling in the development of speech. *Brain and language* 89(2). 393–400.
- Weisleder, Adriana & Anne Fernald. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science* 24(11). 2143–2152.