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

Appendix

Bram Peute & Marisa Casillas

Table A1. Overview of VMS consonant use in the Tseltal dataset. Our alternative measure, cross-clip consonant count, is shown for each recording in the sixth-to-last column.

ACLEW ID	Age in months	p/b	t/d	k/g	m	f	n	l	r	s/z	ʃ/ʒ	ŋ	x/ɣ	Total VMS	Total consonants	Total vocalizations	Total cross-clip consonants	tʃ/dʒ	w	j	ʔ	h
6964	5	-	10	-	-	-	-	-	-	-	-	-	-	1	28	100	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
0456	7	-	-	-	-	-	-	-	-	-	-	-	-	0	30	381	0	-	-	-	-	-
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
0909	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 A2. Overview of VMS consonant use in the Yélî Dnye dataset. Our alternative measure, cross-clip consonant count, is shown for each recording in the sixth-to-last column.

ACLEW ID	Age in months	p/b	t/d	k/g	m	f	n	l	r	s/z	ʃ/ʒ	ŋ	x/ɣ	Total VMS	Total consonants	Total vocalizations	Total cross-clip consonants	tʃ/dʒ	w	j	ʔ	h
1143	8	-	6	-	-	-	-	13	-	-	-	-	-	2	24	99	0	-	-	-	-	-
8018	9	-	-	-	-	-	-	-	-	-	-	-	-	0	43	109	0	-	-	6	-	21
8274	10	-	15	-	22	-	-	19	-	-	-	-	-	3	81	118	1	-	-	-	-	11
5623	11	22	28	5	54	-	20	11	-	-	-	-	-	6	209	197	6	-	23	21	5	17
3446	12	-	-	-	50	-	-	-	-	-	-	-	-	1	91	208	1	-	-	-	-	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	165
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

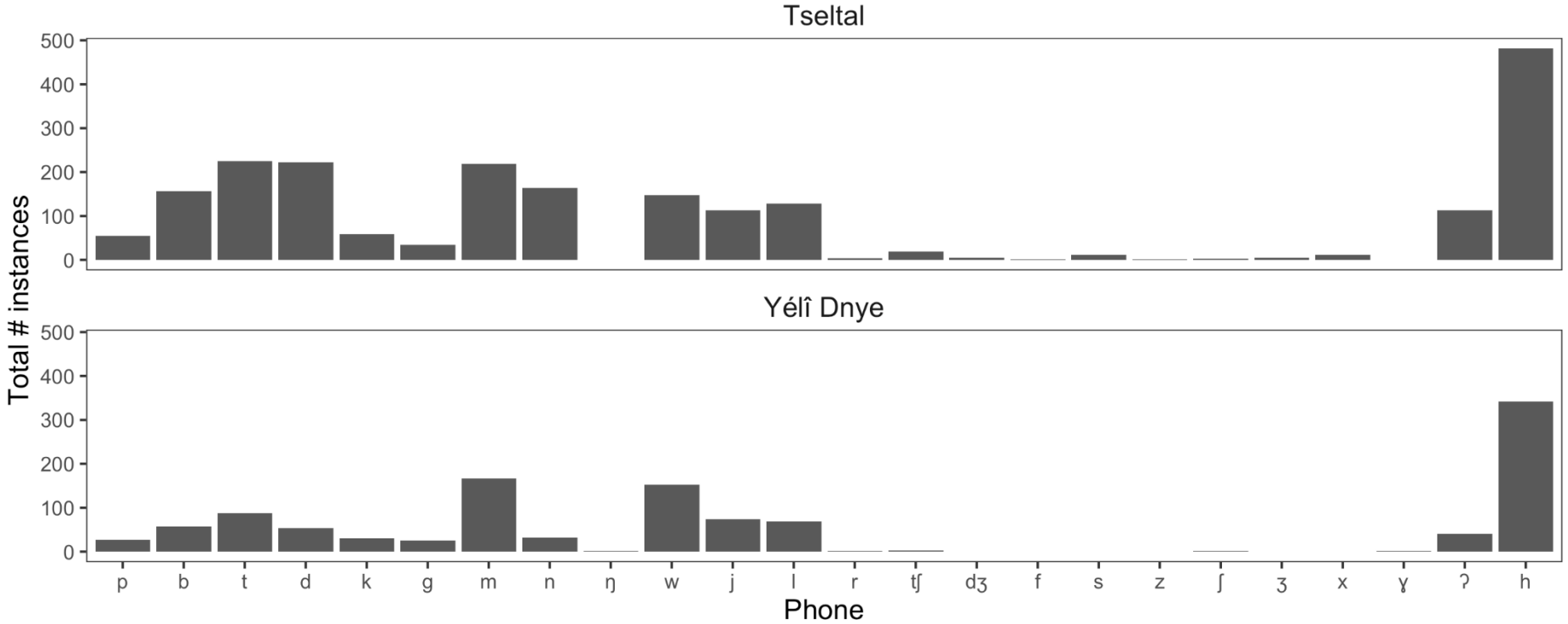

Figure A1. Total number of phones used across tokens from all children, separated by language group.

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 in the present paper.

This 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 vocalizations 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)
    - Shown in the figure below is the mean agreement between the annotators for the two double-annotated clips for each recording; 1.0 indicates perfect agreement on whether vocalizations contained canonical babble or not.

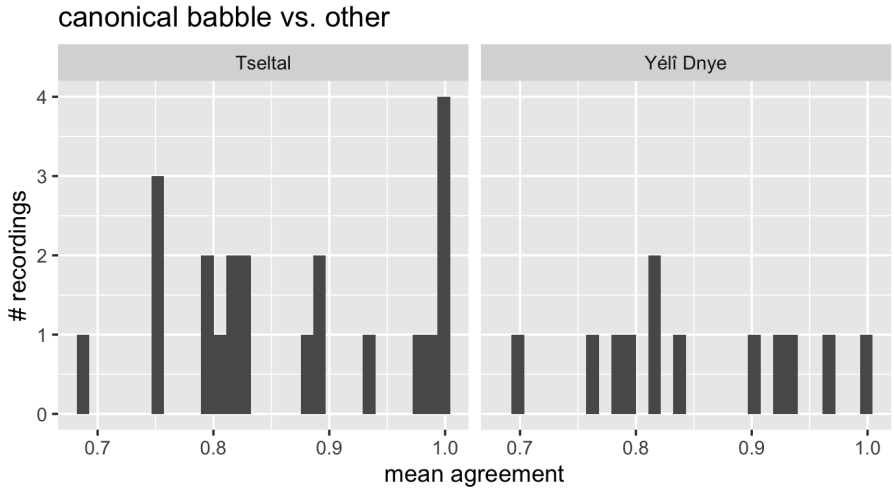

Figure A2. Distribution of agreement between the two annotators, for Tseltal (left) and Yélî Dnye (right).

- **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).
    - 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 second author (Tseltal: -0.16, Yélî Dnye: -0.29)
    - 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 second author (Tseltal: -0.48, Yélî Dnye: -0.55)
    - 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 second author (Tseltal: -0.37, Yélî Dnye: -0.45)
    - 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 second author. However when broken down by language we see that he relatively undercounted for Tseltal but overcounted for Yélî Dnye (Tseltal: -0.63, Yélî Dnye: 0.5)
    - 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 second author, with the substantial average difference for Yélî Dnye of one whole token on average (Tseltal: -0.31, Yélî Dnye: -1)
    - l
        - 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).

- Shown in the figure below is the mean difference in phone counts within a given vocalization between the first and second annotator, for both Tseltal and Yélî Dnye. A mean difference of 0 would indicate that the first and second annotator agreed perfectly on the number of tokens of a given phone in each vocalization. Negative values indicate that the first annotator (the primary annotator) undercounted tokens of that given phone per vocalization, relative to the second annotator. Positive tokens indicate the that the first annotator overcounted tokens relative to the second annotator. Over each of the major phone types examined (p/b, t/d, k/g, m, l, and n), the first annotator typically underestimated tokens per vocalization, with very small differences overall and even smaller differences by phone type across languages.

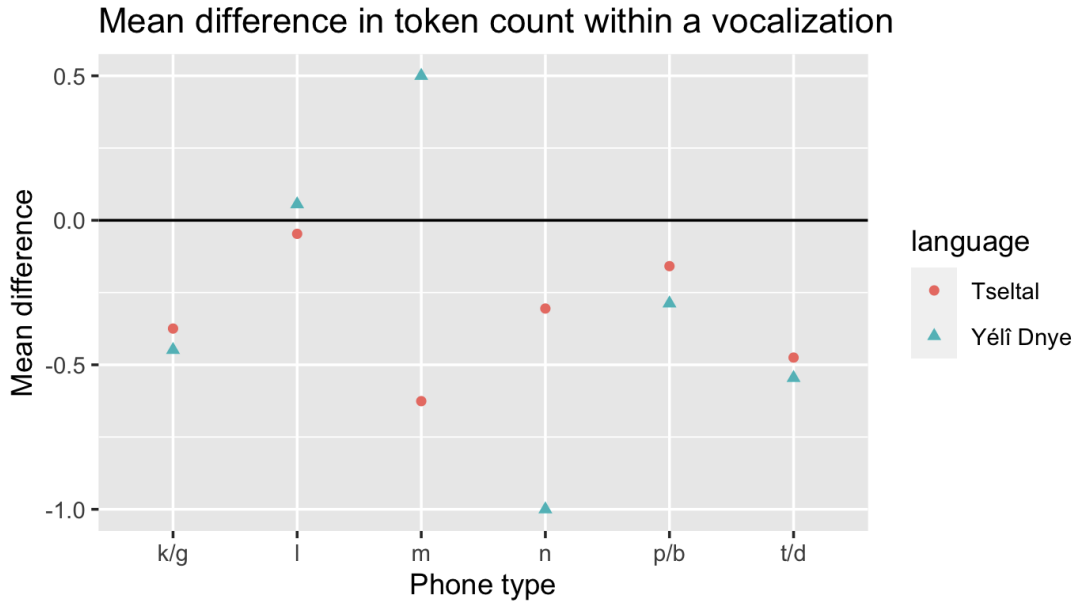

Figure A3. Mean difference in measured number of tokens of different phone types and languages (red circles = Tseltal; blue triangles = Yélî Dnye) between thw two annotators.

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 actually 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 different hours 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 scheme for 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 this time 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. The figure below shows the correlation between number of cross-clip consonants (y-axis) and VMS count (x-axis) for each recording in the dataset, separated by language group (Tseltal in blue, Yélî Dnye in yellow).

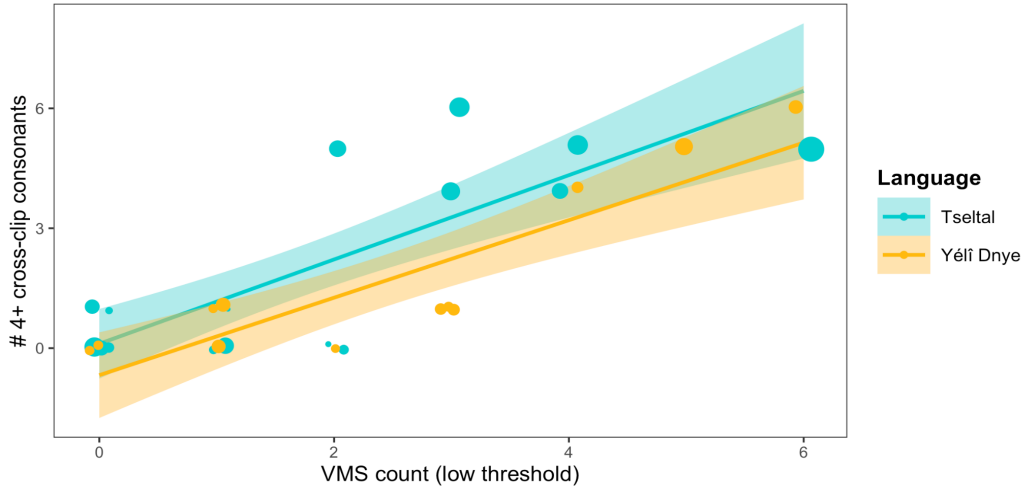

Figure A3. Number of cross-clip consonants (y axis) and their correlation with VMS count (x axis) for each child in the corpora. Each child's recording is a datapoint, with estimated correlations shown separately for the Tseltal (blue) and Yélî Dnye (yellow) data.

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