

Learning a typologically unusual reduplication pattern: An artificial language learning study of base-dependent reduplication

Adam Ussishkin, Jason Haugen, Colin Dawson, Andrew Wedel

Abstract We report on an artificial language learning experiment testing the learnability of a typologically rare pattern of reduplication. Our model comes from syllable-copy reduplication in Hiaki (aka Yaqui, Uto-Aztecan), a base-dependent pattern wherein the shape of reduplication depends crucially on syllabification in the base: coda consonants can copy in reduplication if and only if there is a corresponding coda in the base. Using a controlled artificial language experiment with a forced-choice paradigm, we show that native English speakers who have no prior exposure to any language with a grammar employing syllable-copy reduplication are in many cases able to learn a variable CV or CVC syllable-copying rule as measured by eventual above-chance selection of the correct form. However, compared to participants learning either a consistent CV or CVC copying rule, the performance of participants tasked with learning such a variable syllable-copying rule improves more slowly, and these participants make more errors overall. We suggest that this difference in learnability may be one of a number of factors helping to explain the typological rarity of certain morphological patterns.

Introduction

Typological asymmetries raise a variety of questions regarding cross-linguistic tendencies and can point to the relative likelihood of a particular type of pattern coming into existence or persisting over time within a language. Following a currently fruitful approach involving the application of experimental techniques to study such typological asymmetries, here we investigate whether learners in an artificial language learning experiment are able to extract generalizations within a well-controlled dataset. Our aim is to investigate whether cross-linguistically common generalizations may be easier to learn in a laboratory setting than cross-linguistically rare generalizations. The morphological pattern under investigation is the so-called “syllable copy” reduplication found in Hiaki (also known as Yaqui, a Uto-Aztecan language of Sonora, Mexico and Arizona, USA). This pattern shows variable copying of a CV or CVC syllable, depending on the syllable structure of the base. Specifically, coda consonants can copy in reduplication if and only if there is a corresponding coda in the base.

Using a controlled artificial language experiment with a forced-choice paradigm, we show that native English speakers with no prior exposure to a language with a grammar employing reduplication are in many cases able to learn such a variable CV or CVC syllable-copying rule as measured by eventual above-chance selection of the correct form. However, compared to participants learning either a consistent CV or CVC copying rule, the performance of participants tasked with learning such a variable syllable-copying rule improves more slowly, and these participants make more errors overall. We suggest that such a difference in learnability may be one of a number of factors helping to explain the typological rarity of certain morphological patterns.

The methods used here stem from work done in other artificial grammar learning studies, which have usually focused on phonology or syntax. For studies focusing on phonology, see Aslin, Saffran, and Newport (1998), Peperkamp (2003), Wilson (2006), Kapatsinsky (2013), and LaCross (2015), among others; for syntax, see Culbertson, Smolensky, and Legendre (2012) and Culbertson, Smolensky, and Wilson (2013). An important difference between these earlier studies and the current one is our focus on morphology. The phenomenon of interest in this paper concerns patterns of partial morphological reduplication. In particular, we examine the relative learnability of a cross-linguistically uncommon pattern of base-dependent reduplication involving the copying of a syllable.

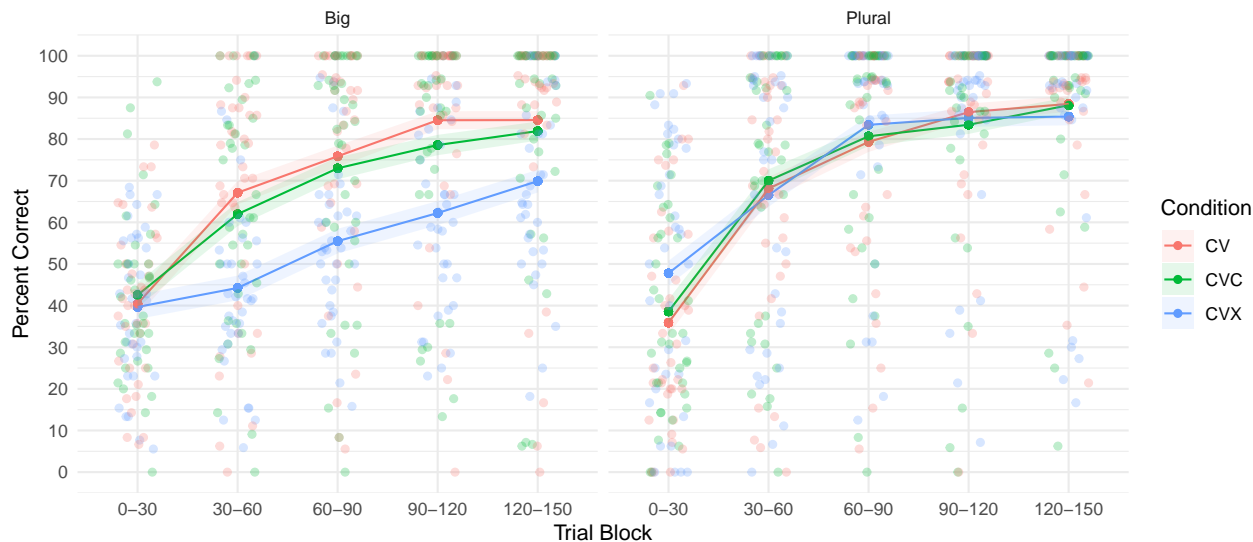
Results

Participants

One participant who was assigned to the CVC condition had zero correct responses and was excluded from the analysis. The remaining 108 participants consisted of 36 in each of the three morphological conditions (CV, CVC, CVX).

Accuracy by Condition and Prompt

Each participant responded to 75 trials with the total reduplication (“Plural”) prompt, and 75 trials with the partial reduplication (“Big”) prompt, interleaved in a different random order for each participant. The proportion of correct responses in each 25 trial segment, separated by Condition and Prompt type and averaged over participants, is shown in Fig. XX. Individual participants’ accuracy rates by trial bin are plotted as light circles, with random horizontal perturbations added to facilitate visibility. Note that because each participant responded to both prompts, each of these individual accuracy rates was obtained from only half of the trials in each 25 trial bin, on average. Confidence intervals for the mean percent correct are shown for each trial bin, and are calculated based on a Binomial distribution as if the collection of responses each bin were independent trials. This is unrealistic, however, due to heterogeneity across participants, items, and trials. To better model the dependencies and heterogeneities across observations, a generalized linear mixed model was fit to the data. This model is described in the following section.



Statistical Analysis

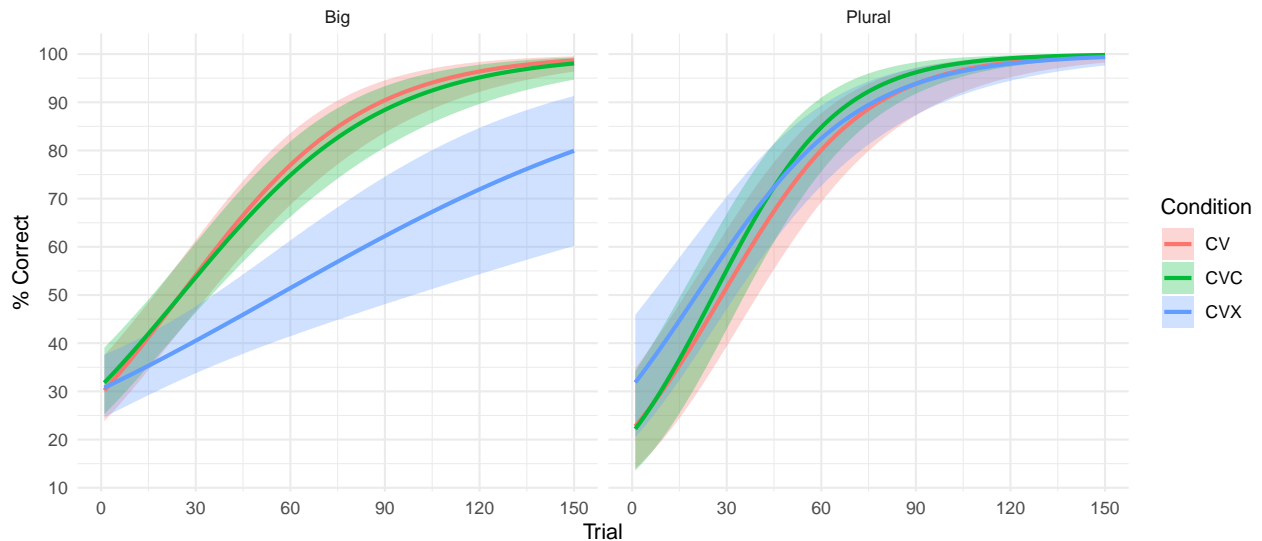
Accuracy was modeled for individual trials using a generalized linear mixed model with a logistic link and binomial likelihood. Condition (3 levels: CV, CVC, CVX) and Prompt (2 levels: Plural, Big) were included as categorical predictors, and Trial (1-150) was included as a quantitative predictor. All interactions were included as well. Random effects by Participant were included for Prompt, Trial and their interaction, and crossed random effects by Item were included for Condition¹. Maximum likelihood parameter estimates were obtained using the BOBYQA algorithm, as implemented by the `glmer()` function in the `lme4` R package.

To aid in numerical stability and interpretability of the lower order coefficients, the predictors were recoded as follows: **Trial** was centered and scaled so that zero represented the mid-way point in the sequence. **Condition** was coded using planned Helmert contrasts, based on the theoretical predictions of the study: the first coefficient represented the difference in log odds of a correct responses between the CVC and CV conditions, and the second coefficient represented the difference between the CVX condition and the other

¹In the syntax used by the `lme4` package, the model formula is `Accuracy ~ Condition * Trial * Prompt + (Condition | Item) + (Trial * Prompt | Participant)`

two combined. These two contrasts were centered so that zero represented the average result. **Prompt** was reference coded, with partial reduplication (the “Big” rule) as the reference category. These choices should not affect the estimated probabilities; however, we found that coding the predictors in this way helped improve the numerical stability of the matrix operations involved in the iterative estimation algorithm.

Fitted Learning Curves To obtain 95% confidence bands for the predicted probability of a correct response at each combination of trial, condition, and prompt, a parametric bootstrap procedure was employed, using the **bootMer** algorithm in the **lme4** R package. Both fixed and random effects were resampled as part of the bootstrap algorithm, and fitted response curves by trial, condition and prompt, based on the fixed effects components only, were computed for each bootstrap sample. A total of 3200 bootstrap samples were used to calculate 95% confidence bands for the predicted accuracies by trial, which are shown in Fig. XX. Individual participants’ fitted learning curves (adjusted for pseudoword) are overlaid for reference.



Participants improved their accuracy over the course of the experiment in all three groups for both the full and partial reduplication rules. As predicted, for trials using the total reduplication (**Plural**) rule, all three groups achieved near-ceiling performance by the end of the trial sequence². For trials using the partial reduplication (**Big**) rule, the CV and CVC groups still achieved near-ceiling performance by the end of the experiment, but the CVX group made significantly slower progress: the upper endpoint of the confidence interval for the success probability at the last trial is barely above 0.90 for this group.

Hypothesis Tests The central prediction was that the CVX reduplication pattern would be more difficult to learn than either the CV or CVC pattern in the context of the **Big** rule, but not in the context of the **Plural** rule, whereas the CV and CVC reduplication patterns would be equally learnable for both types of rule.

To assess whether the pattern of results across conditions differed according to the **Prompt** type (**Big** vs **Plural**), a reduced model was fit in which the three way interaction and the two way interaction between **Prompt** and **Condition** were omitted from the fixed effects structure. The other two-way interactions were retained and the random effects structure was the same³. In context, this reduced model allowed both the

²The discrepancy between the empirical mean accuracy curves in Fig. XX and the predicted probabilities of a correct response from the mixed model shown in Fig. XX – particularly the fact that the empirical means are so much lower than the modeled probabilities – bears explanation. The random effects structure allows for individual participants and individual items to have higher or lower predicted accuracy than the aggregate prediction. On average, these perturbations from the aggregate prediction are zero; however, this is only true in the log odds space. A given perturbation in the log odds corresponds to a larger difference in probabilities when the aggregate probability is near 0.5; hence a symmetric distribution of perturbations of the log-odds corresponds to a skewed distribution in probability space, such that the fitted probability is closer to 0 or 1 than the empirical average of the perturbed probabilities.

³In **lme4** syntax, the model formula is `Accuracy ~ Trial * (Condition + Prompt) + (Condition | Item) + (Trial * Prompt | Participant)`

level and rates of increase over trials in the odds of a correct response to differ between trial types, and to differ between conditions as well, but the differences among conditions in learning rate and level were required to be the same for both trial types.

A likelihood ratio test comparing the full model with this reduced model revealed significant evidence of an interaction between **Prompt** and **Condition** ($\chi^2(4) = 14.52, p = 0.006$). As such, all subsequent testing and estimation of the effect of **Condition** was done separately for each **Prompt** type (the original model was employed for these analyses, but tests and confidence intervals involved simple effects).

Simple effects of **Condition** were obtained from the model for each **Prompt** type at 30-trial intervals (0, 30, 60, 90, 120, 150). Using Wald tests with an asymptotic χ^2 approximation of the sampling distribution of the standardized squared norm of the vector of fixed effects for **Condition** at each combination of **Prompt** and **Trial**, there is no significant evidence of a difference across conditions for either **Big** or the **Plural** trials at the beginning of the experiment (**Plural**: $\chi^2(2) = 1.68, p = 0.43$; **Big**: $\chi^2(2) = 0.11, p = 0.94$). For the trials with the **Plural** reduplication prompt, there is no significant evidence of any differences across conditions at any later trial either (all $\chi^2(2) \leq 1.52$; all $p \geq 0.467$). For the trials with the **Big** reduplication prompt, there is significant evidence of a difference across conditions at each trial marker from 30 on (all $\chi^2(2) \geq 8.96$; All $p \leq 0.011$).

Planned Helmert contrasts for the **Condition** factor were tested for the **Big** prompt trials, every 30 trials starting at trial 30. The first compared accuracy in the CV condition to the CVC condition, and the second compared accuracy in the CVX condition to the other two combined. Asymptotic Wald tests reveal significant evidence of lower accuracy in the CVX condition than in the other two at all five reference trials (all $z > 2.99$; all $p < 0.003$). There is no significant evidence of a difference in accuracy between the CV and CVC conditions at any trial (all $z < 0.55$; all $p > 0.585$). Odds ratios and confidence intervals for each contrast at each reference trial are given in Table XX.

Contrasts for the interaction between **Prompt** and **Condition** were tested by comparing the coefficients for each of the **Condition** contrasts for the **Plural** prompt with their counterparts in the **Big** prompt at each of the reference trials. No evidence of a difference of differences was found at Trial 0 for either the CVC vs CV contrast (all $z < 0.24$; all $p > 0.811$) or the CVX vs others contrast (all $z < 1.15$; all $p > 0.252$). For the CVC vs CV contrast, no evidence of a difference of differences was found at any subsequent reference trial either (all $z < 1.49$; all $p > 0.136$); however for the CVX vs others contrast, the odds ratios were significantly different between the two Prompt types (all $z > 2.61$; all $p < 0.009$). Odds ratios and confidence intervals for these interaction coefficients at each reference trial are given in Table XX.

Appendices

Model Coefficients and Confidence Intervals

Prompt	Contrast	Trial	OR	95% CI	z	P-value	
Big	CVC vs CV	0	1.08	(0.69 , 1.69)	0.33	0.738	
		30	0.98	(0.65 , 1.48)	-0.09	0.927	
		60	0.89	(0.5 , 1.59)	-0.39	0.698	
		90	0.81	(0.35 , 1.89)	-0.49	0.625	
		120	0.74	(0.23 , 2.3)	-0.53	0.598	
		150	0.67	(0.16 , 2.84)	-0.55	0.585	
	CVX vs rest	0	1.02	(0.47 , 2.19)	0.04	0.968	
		30	0.34	(0.17 , 0.69)	-2.99	0.003	**
		60	0.11	(0.04 , 0.31)	-4.29	< 0.001	***
		90	0.04	(0.01 , 0.16)	-4.47	< 0.001	***
		120	0.01	(0 , 0.09)	-4.45	< 0.001	***
		150	0.00	(0 , 0.05)	-4.41	< 0.001	***
Plural	CVC vs CV	0	0.96	(0.41 , 2.24)	-0.09	0.929	
		30	1.16	(0.58 , 2.31)	0.41	0.680	
		60	1.39	(0.62 , 3.14)	0.79	0.428	
		90	1.67	(0.54 , 5.17)	0.89	0.373	
		120	2.01	(0.44 , 9.18)	0.90	0.369	
		150	2.41	(0.35 , 16.83)	0.89	0.374	
	CVX vs rest	0	2.64	(0.61 , 11.48)	1.29	0.196	
		30	1.62	(0.49 , 5.38)	0.78	0.434	
		60	0.99	(0.24 , 4.03)	-0.01	0.989	
		90	0.61	(0.09 , 4.21)	-0.51	0.613	
		120	0.37	(0.03 , 5.01)	-0.75	0.456	
		150	0.23	(0.01 , 6.31)	-0.87	0.383	

Prompt Contrast	Condition Contrast	Trial	OR	95% CI	z	P-value	
Plural / Big	CVC vs CV	0	0.89	(0.35 , 2.29)	-0.24	0.811	
		30	1.18	(0.6 , 2.31)	0.48	0.632	
		60	1.56	(0.79 , 3.07)	1.29	0.198	
		90	2.06	(0.8 , 5.35)	1.49	0.136	
		120	2.73	(0.72 , 10.43)	1.47	0.142	
		150	3.61	(0.62 , 21.18)	1.42	0.155	
	CVX vs rest	0	2.60	(0.51 , 13.29)	1.15	0.252	
		30	4.77	(1.48 , 15.41)	2.61	0.009	**
		60	8.77	(2.75 , 28.02)	3.66	< 0.001	***
		90	16.12	(3.22 , 80.6)	3.39	0.001	**
		120	29.62	(3.1 , 283.3)	2.94	0.003	**
		150	54.44	(2.77 , 1070.45)	2.63	0.009	**

Individual Modeled Learning Curves

