Pattern Structure Modulates Learning of Lexically Conditioned Morphology

Work in first language acquisition and artificial grammar learning (AGL) indicates that language learners can extract systematic regularities from inconsistent language data. In some cases, learners reproduce the inconsistencies in the data (frequency match), while in other cases they over-extend (regularize) the most frequent patterns when generalizing to novel examples. Studies with children have generally observed regularization, while adults in AGL studies tend to frequency match (e.g., Hudson Kam & Newport 2005, 2009). However, some adult AGL studies (e.g., Schumacher & Pierrehumbert 2021: SP) instead demonstrate regularization. We replicate and extend one such case (SP) to better understand why learners regularize in some cases but not others. We find evidence that adult learners' regularization depends on the structure of the morphological system they are exposed to.

SP's AGL languages all involve morphological reversal: some stems consistently use a suffix to mark the plural and others use that same suffix to mark the singular. To reconcile the regularization observed by SP but not others, we hypothesize that learners regularize more with reversal than other inconsistent systems, such as plural allomorphy, where two distinct suffixes are used by different stems to mark the plural. We tested this hypothesis by constructing languages that differ only in whether lexical exceptions involve REVERSAL (1c) or ALLOMORPHY (1a). To determine whether regularization is specifically driven by reversal or by exceptional marking of singulars, we also constructed a third DOMINANT language condition (1b), identical to reversal except that singulars are marked with a distinct suffix. Stems and images were balanced phonotactically, semantically, and visually to ensure none of these properties were predictive of suffix choice. 20 native English speakers, recruited via Prolific, were trained for 288 trials on each of these languages and then tested on 24 new stems.

Fig. 1 shows the proportion of minority (exceptional) marking produced by participants in the test phase for each condition. ALLOMORPHY is roughly frequency matching (median near the proportion of exceptions in training), REVERSAL is more regularized (median well below the proportion of exceptions in training), and the generalization rate for DOMINANT is in between. As predicted by our hypothesis, a logistic mixed effects regression (Table 2) indicates there is significantly more regularization in REVERSAL than in the other two conditions (β = -0.72, z = -2.20, p = 0.028). We do not observe more regularization in DOMINANT than in ALLOMORPHY (β = -0.33, z = -0.78, p =0.44). Our results provide a possible explanation for the varying results found in previous AGL studies with adults. Since we do not observe regularization in ALLOMORPHY, this suggests that the regularization observed by SP is not due to a general regularization pressure. Rather, regularization depends on the structure of the pattern being learned, with more pressure to regularize in the presence of reversal. We discuss implications of our results for first and second language learning.

	Regulars: 18/24 (75%) of stems		Exceptions: 6/24 (25%) of stems	
	Pattern	Example Stem/Trial	Pattern	Example Stem/Trial
1a) ALLOMORPHY	Sg: stem + ∅ Pl: stem + **fi**	gragol gragolfi	Sg: stem + ∅ Pl: stem + **ku**	drokra drokraku
1b) DOMINANT	Sg: stem + ∅ Pl: stem + **fi**	gragol gragolfi	Sg: stem + **ku** Pl: stem + ∅	drokraku drokra
1c) REVERSAL	Sg: stem + ∅ Pl: stem + **fi**	gragol gragolfi	Sg: stem + **fi** Pl: stem + ∅	drokrafi drokra

Table 1. Distribution of number marking in each condition. Images represent singular and plural meanings for each stem. There were 24 stems, each associated with a different image. Each trial presented either a singular image (e.g.) or a plural (e.g.) and asked participants to select the correct form for it (e.g. "gragol" or "gragolfi").

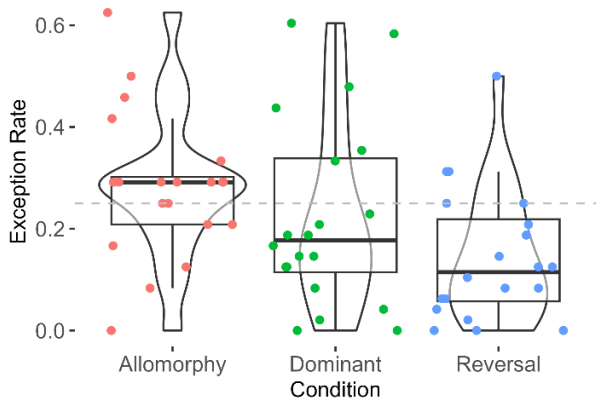

Figure 1. Each point shows the proportion of novel exceptional responses produced by a participant in that condition. Boxes show the median and 50% quantiles. The gray dashed line shows the proportion of exceptions in training.

Formula: Exception ~ Cond + (1 | Subject) + (1 + Cond | Stem) + (1 + Cond | Set)

	β	Stderr	z	Pr(>|z|)
(Intercept)	-1.51	0.16	-9.35	<2e-16 ***
Cond1	-0.33	0.43	-0.78	0.44
Cond2	-0.72	0.33	-2.20	0.028 *

Table 2. Logistic mixed effects regression with rate of exceptionality marking as the dependent variable. Condition is Reverse Helmert Coded (Cond2 compares REVERSAL to the other two). 'Set' is the stimuli set, counterbalanced for suffix form and stem-image associations.

References

[1]Hudson Kam, C. & Newport, E.(2005). Regularizing Unpredictable Variation *Language Learning and Development* [2] Hudson Kam, C.& Newport, E. (2009). Getting it right by getting it wrong. *Cognitive Psychology* [3] Schumacher, R. & Pierrehumbert, J. (2021). Familiarity, consistency, and systematizing in morphology. *Cognition.*