

# generalizability-writeup

## Methods

Following Harmon and Kapatsinski (2017), two artificial languages were used: Dan and Nem. See 1. In each language, the same four suffixes were used:  $-sil_{PL}$ ,  $-dan_{PL}$ ,  $-nem_{DIM}$ , and  $-shoon_{DIM}$ . Notably, in our language  $-dan$  and  $-sil$  overlap in meaning (they both occur in plural contexts) as do  $-nem$  and  $-shoon$  (which both occur in diminutive contexts).

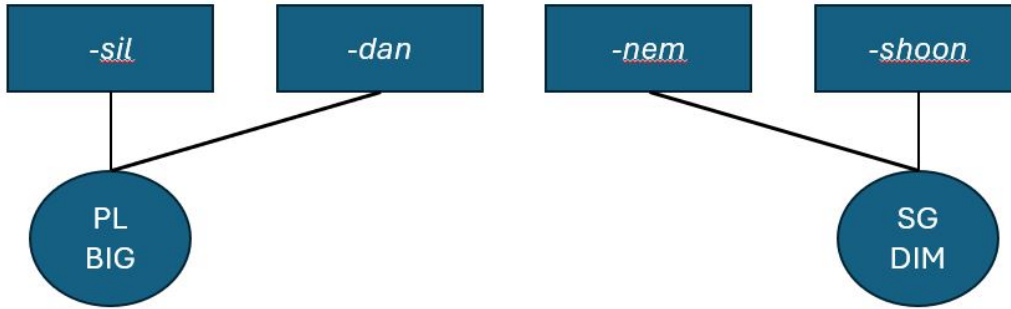


Figure 1: A description of the suffixes in our artificial languages. The thicker lines denote the more frequent form in each language. The Dan language is on the left, the Nem language is on the right.

During the exposure phase, the suffixes were paired with an image. The suffixes  $-sil_{PL}$  and  $-dan_{PL}$  were always paired with a picture of multiple large pictures. On the other hand, the suffixes  $-nem_{DIM}$  and  $-shoon_{DIM}$  were always paired with a picture of a single small creature. The design of the stimuli result in participants being able to learn that  $-sil$  and  $-dan$  are either simply plural or simply non-diminutive. Similarly,  $-nem$  and  $-shoon$  can be learned as either simply singular or simply diminutive. Thus the test phase tested this.

Our Experiment comprised of three different conditions (see Table 1), one in which the type frequency of the frequent language’s suffix was manipulated (Type Frequency), one in which the

Table 1: Description of each of our conditions. Note that in Condition 1, there are an equal number of tokens between the frequent and infrequent items, however there are a greater number of types in the frequent items. In Condition 2, the opposite is true: the frequent items occur more, but in the same number of types as the infrequent items. Finally, in Condition 3, the frequent items occur both a greater number of times and in a greater number of different contexts.

	<b>Frequent.Token</b>	<b>Frequent.Type</b>	<b>Infrequent.Token</b>	<b>Infrequent.Type</b>
<b>Type Frequency</b>	12	12	12	3
<b>Token Frequency</b>	12	3	3	3
<b>Type-Token Frequency</b>	12	12	3	3

token frequency was manipulated (Token Frequency), and one in which both were manipulated (Type-Token Frequency).

## Procedure

Each participant was randomly assigned one of the conditions. In each condition, participants were first presented with an exposure phase. After the exposure phase, participants took part in a production task, a form choice task, and a comprehension task. We describe each of these below.

## Exposure Phase

Following Harmon and Kapatsinski (2017), each exposure trial consisted of the presentation of a picture on the computer screen which was subsequently followed with a written label for the image as well as an audio version of that label. Specifically, the image first appeared on the screen and then 1.25 seconds later was followed by both the label of the creatures on the screen as well as the audio for that creature. Participants were instructed to type the name of the creature and press enter. Participants had 4 seconds after presentation of the name of the creature and were given feedback if they typed the wrong name in.

## Production Task

After the exposure phase, participants were presented with a production task. In this task, participants were presented with images and told to produce a label for the image. Specifically, initially the unaffixed form appeared on the screen along with the corresponding image. 2 seconds later, the four different possible images of that creature appeared (a singular big creature, multiple big creatures, a single small creature, and multiple small creatures). Three

of these images disappeared after 1.25 seconds, leaving a single image for the participant to produce a label for. Participants had 10 seconds to respond.

### Form Choice Task

After the production task, participants were presented with a form choice task. In this task, participants were presented first with the base, unaffixed form and the corresponding image. 2 seconds later four images flashed on the screen, remained on the screen for 1.25 seconds, and disappeared leaving a single image. Along with the single image, participants were also presented with two possible labels for that image, one label on the bottom right of the screen and one label on the bottom left of the screen. Participants were given four seconds to press either the left arrow or the right arrow to choose the corresponding label. The labels were counterbalanced with respect to which side of the screen they appeared on. The goal of this task was to assess whether type and/or token frequency influence the form choice when accessibility differences between frequent and rare forms have been attenuated.

### Comprehension Task

Finally, participants were presented with a comprehension task. In this task, participants were first given the label and corresponding audio for a given creature. After 0.25 seconds, four images appeared on the screen and participants had 4 seconds to choose one of the images.

## Analyses and Results

### Production Task

In order to examine the effect of type/token frequency on semantic extension, we examined instances where the image was diminutive plural.

In order to determine whether the effect of frequency on semantic extension differed between conditions, we ran two Bayesian linear mixed-effects models on the production data. For the first analysis, the dependent variable was whether participant’s produced the frequent suffix (*dan* in the *dan* language, *nem* in the *nem* language) for the original meanings. A larger estimate indicates that the frequent suffix was more likely to be produced than the infrequent suffix. The independent variables was whether the training consisted of high token frequency, high type-frequency, or both. We also included a random intercept for participant and a random slope for condition by response stem. Our second analysis was analogous except that it examined novel meanings. The equation for our models is included below in Equation 1:

$$frequent\_suffix \sim condition + (1|participant) + (condition|resp\_stem) \quad (1)$$

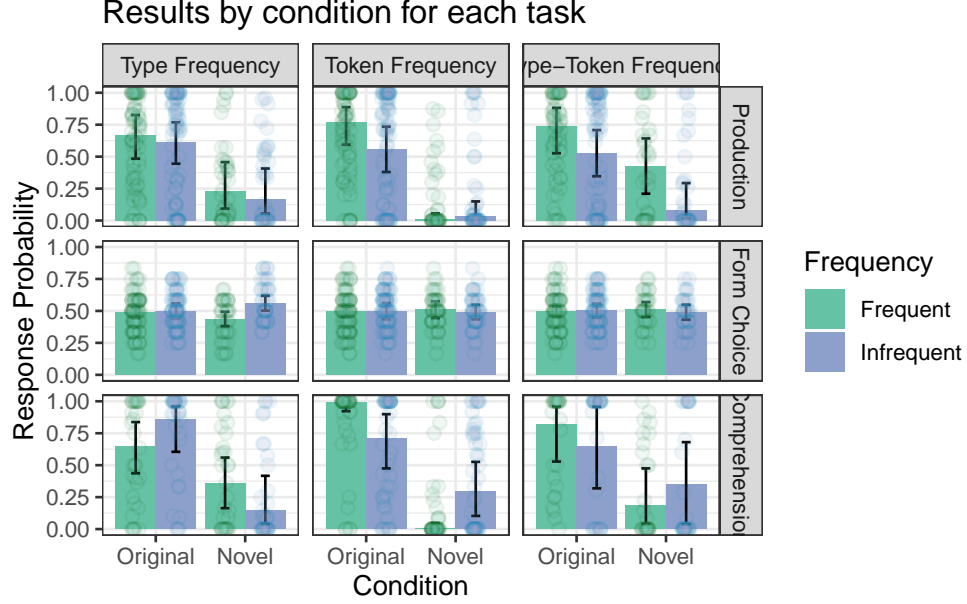


Figure 2: Plot of our results. Points indicate individual subject values.

The results are below in Table 2. The results suggest that

### Form Choice Task

In order to examine the effects of form-choice, we similarly ran two models analogous to those we ran for the production task (Equation 1). In the context of the form-choice task, the dependent variable reflects whether participants chose the option with the frequent suffix or the one with the infrequent suffix. Similarly, our first model examined participants' response for original meanings and our second examined responses for novel meanings.

Table 2: Results of the statistical models for the production task.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
<b>Original</b>					
Intercept	0.73	0.30	0.13	1.33	99.14
Type Frequency	-0.89	0.40	-1.67	-0.12	1.17
Token Frequency	0.47	0.40	-0.29	1.26	88.64
Type-Token Frequency	0.42	0.43	-0.42	1.26	16.42
<b>Novel</b>					
Intercept	0.70	0.66	-0.59	2.03	86.39
Type Frequency	-0.08	0.83	-1.73	1.57	46.57
Token Frequency	-1.15	0.85	-2.86	0.46	8.14
Type-Token Frequency	1.22	0.92	-0.58	3.06	8.86

Table 3: Results of the statistical models for the 2afc task.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
<b>Original</b>					
<b>Intercept</b>	-0.02	0.08	-0.17	0.14	39.78
<b>Type Frequency</b>	0.01	0.11	-0.20	0.22	55.03
<b>Token Frequency</b>	0.05	0.09	-0.14	0.23	70.60
<b>Type-Token Frequency</b>	-0.06	0.12	-0.29	0.18	71.09
<b>Novel</b>					
<b>Intercept</b>	-0.06	0.07	-0.19	0.07	17.61
<b>Type Frequency</b>	-0.20	0.09	-0.38	-0.01	2.02
<b>Token Frequency</b>	0.09	0.10	-0.11	0.29	84.08
<b>Type-Token Frequency</b>	0.11	0.12	-0.13	0.34	16.69

The results of the form-choice task suggest that when participants are presented with both possible options, there is only an effect of type frequency and only for novel items.

## Comprehension Task

Our results for the comprehension task are included below in Table 4. For the comprehension task, the dependent variable for our statistical analysis was whether the meaning that participants chose was novel or original. A positive estimate suggests that participants were more likely to choose the novel meaning, a negative estimate suggests that participants were more likely to choose the original meaning. The dependent variables were whether the training comprised of high token frequency, high type frequency, or both, and whether the suffix was frequent or infrequent (as well as their interactions). We also included a by-participant random intercept and a frequency by participant random slope. The model equation is included below in Equation 2.

$$meaning \sim condition * frequent + (1 + frequent|participant) \quad (2)$$

Our results suggest that the frequent suffix is used more for the original meaning when there is a high type-frequency.

Table 4: Results of the statistical models for the comprehension task.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
<b>Intercept</b>	-2.65	0.39	-3.46	-1.91	0.00
<b>Type Frequency</b>	0.41	0.44	-0.44	1.31	82.39
<b>Token Frequency</b>	-0.40	0.42	-1.24	0.41	3.36
<b>Type-Token Frequency</b>	-0.01	0.50	-0.98	0.97	51.13
<b>Frequent</b>	0.18	0.36	-0.52	0.90	16.91
<b>Type Frequency:Frequent</b>	-0.83	0.46	-1.75	0.06	98.06
<b>Token Frequency:Frequent</b>	0.91	0.45	0.05	1.80	69.85
<b>Type-Token Frequency:Frequent</b>	-0.08	0.52	-1.10	0.96	56.42

Harmon, Zara, and Vsevolod Kapatsinski. 2017. “Putting Old Tools to Novel Uses: The Role of Form Accessibility in Semantic Extension.” *Cognitive Psychology* 98: 22–44. <https://doi.org/10.1016/j.cogpsych.2017.08.002>.