

The effects of type and token frequency on semantic extension

Methods

Following Harmon & 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), and $-nem$ and $-shoon$ also overlap in meaning (they both occur in diminutive contexts). Since all four suffixes are possible candidates for the diminutive plural meaning, we will examine how properties of the language affect which suffixes are extended to express the diminutive plural meaning.

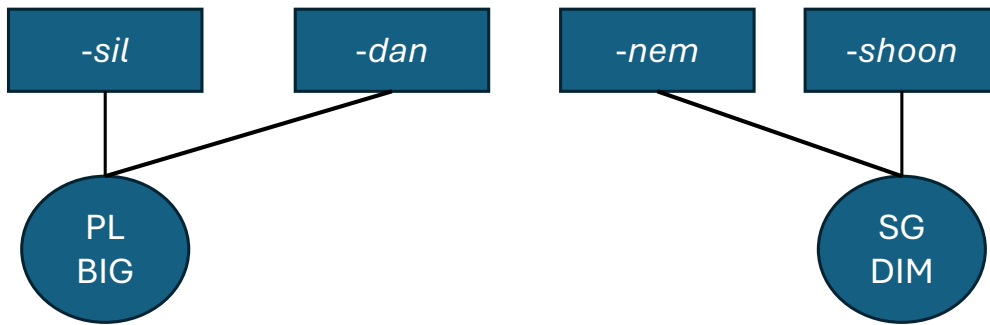


Figure 1: A description of the suffixes in our artificial languages. The thicker lines denote the more frequent form in each language: the plural $-dan_{PL}$ in the Dan language and the diminutive $-nem_{DIM}$ in the Nem language.

During the exposure phase, each suffix was 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 results 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.

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.

	Frequent.Token	Frequent.Type	Infrequent.Token	Infrequent.Type
Type Frequency	12	12	12	3
Token Frequency	12	3	3	3
Type-Token Frequency	12	12	3	3

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 token frequency was manipulated (Token Frequency), and one in which both were manipulated (Type-Token Frequency). In this context, a higher type-frequency corresponds to the suffix appearing with a larger number of different stems relative to the competing suffix, while a larger token-frequency corresponds to simply appearing a greater number of times relative to the competing suffix, regardless of the number of different stems it occurs with.

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 were tested using a production task, a form choice task, and a comprehension task. We describe each of these below.¹

Exposure Phase

Following Harmon & 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 presentation of that label. Specifically, the image first appeared on the screen and

¹Additionally, a demo of the experiment can be found at the following link: https://run.pavlovia.org/znhoughton/generalizability_demo.

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 to respond after the presentation of the name of the creature and were given feedback as to whether they were correct or not.

Participants saw each trial within the exposure phase 5 times, each in a randomized order.

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. In the production task, some of the stems were familiar (i.e., seen in training) and others were novel (i.e., not seen in training).

Participants saw each trial within the production task 4 times, each in a randomized order.

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. In the form choice task, some of the stems were familiar (i.e., seen in training) and others

were novel (i.e., not seen in training). 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.

Participants saw each trial within the form choice task 2 times, each in a randomized order.

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 click one of the images on the screen that corresponded to the label. Similar to the before-mentioned tasks, in the comprehension task some of the stems were familiar (i.e., seen in training) and others were novel (i.e., not seen in training).

Participants saw each trial within the comprehension task 2 times, each in a randomized order.

Analyses and Results

A plot of our model estimates for each task is presented in Figure 2. The points represent individual subject estimates, the bars and credible intervals correspond to estimates from a Bayesian linear mixed-effects regression model.²

Production Task

In order to examine the effects of type and token-frequency on participants choice of suffix, we examined the effects of type and token-frequency on the original meaning as well as the novel

²All data and code for the analyses can be found here: <https://github.com/znhoughton/Generalizability-Type-Token>.

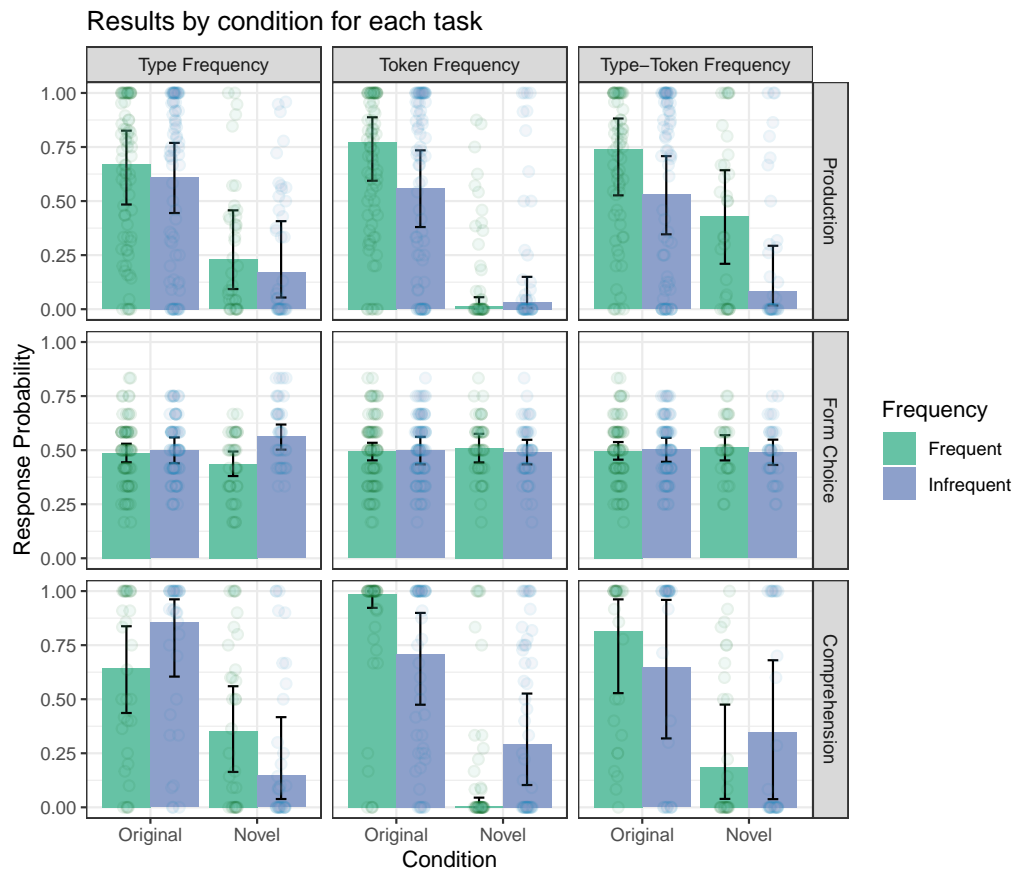


Figure 2: Plot of our results. Points indicate individual subject values.

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

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept (Type-Token Frequency)	1.01	0.56	-0.05	2.13	96.87
Type Frequency	-0.75	0.71	-2.22	0.58	13.79
Token Frequency	-0.20	0.68	-1.57	1.13	38.40
Novel Meaning	0.10	0.32	-0.54	0.74	62.06
Type Frequency:Novel	0.25	0.42	-0.57	1.08	72.89
Token Frequency:Novel	-0.59	0.44	-1.47	0.26	8.69

meaning (diminutive plural). Specifically, we ran a Bayesian linear mixed-effects model on the production data. The dependent variable was whether the participant produced the frequent suffix (*dan* in the *dan* language, *nem* in the *nem* language). A larger estimate indicates that the frequent suffix was more likely to be produced than the infrequent suffix. The independent variables were condition, meaning, and stem (either familiar stem or novel stem). Condition was treatment coded such that the intercept was the high type and high token frequency condition. Meaning was sum coded. We also included a random intercept for participant and a random slope for meaning by participant. The equation for our model is included below in Equation 1.

$$\text{frequent_suffix} \sim \text{condition} * \text{meaning} + (1 + \text{meaning} | \text{participant}) \quad (1)$$

The results are below in Table 2. The results suggest that in general the frequent suffix is used more than the infrequent suffix for all conditions. Notably however, we find no meaningful interaction effect for type frequency and novel meaning, suggesting that the effect of meaning is similar for both conditions. We do, however, find an interaction effect between token frequency and novel meaning, suggesting that the frequent suffix is used less for the novel meaning in the token frequency condition. This suggests that there is semantic extension when there is high type and high token frequency as well as the case where there is high type-frequency and low token-frequency, but in the case where there is high token-frequency with a low type-frequency, we see entrenchment in production.

In order to follow up on this, we ran an additional model that included whether the stem was

Table 3: Results of the statistical models for the production task.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept (Type-Token Frequency)	0.70	0.44	-0.14	1.61	94.74
Type Frequency	-0.46	0.53	-1.66	0.45	18.21
Token Frequency	0.00	0.47	-0.94	0.94	50.16
Novel Meaning	0.00	0.26	-0.50	0.50	49.99
Novel Stem	0.26	0.11	0.05	0.48	82.79
Type Frequency:Novel Meaning	0.30	0.32	-0.32	0.96	99.33
Token Frequency:Novel Meaning	-0.45	0.35	-1.17	0.20	99.16
Type Frequency:Novel Stem	0.34	0.14	0.06	0.62	9.48
Token Frequency:Novel Stem	-0.27	0.17	-0.61	0.07	41.12
Novel Meaning:Novel Stem	-0.01	0.11	-0.22	0.21	5.90
Type Frequency:Novel Meaning:Novel Stem	0.35	0.14	0.07	0.64	46.59
Token Frequency:Novel Meaning:Novel Stem	-0.04	0.17	-0.38	0.30	99.21

familiar (occurred in training) or novel (did not occur in training). The results are reported in Table 3 and visualized in Figure 3.

Our follow-up analysis suggests that type-frequency leads to entrenchment with stems that learners have encountered before. However, for novel stems, participants are more likely to use the frequent suffix to convey a novel meaning.

Form Choice Task

In order to examine the effects of form-choice, we similarly ran a model analogous to the first model 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.

The results of the form-choice task suggest that when participants are presented with both possible options, for both token-frequency and type-token frequency, participants are equally likely to choose the frequent suffix as the infrequent suffix. Interestingly, for the type-frequency condition, participants were less likely to choose the novel meaning than the original meaning, though the effect

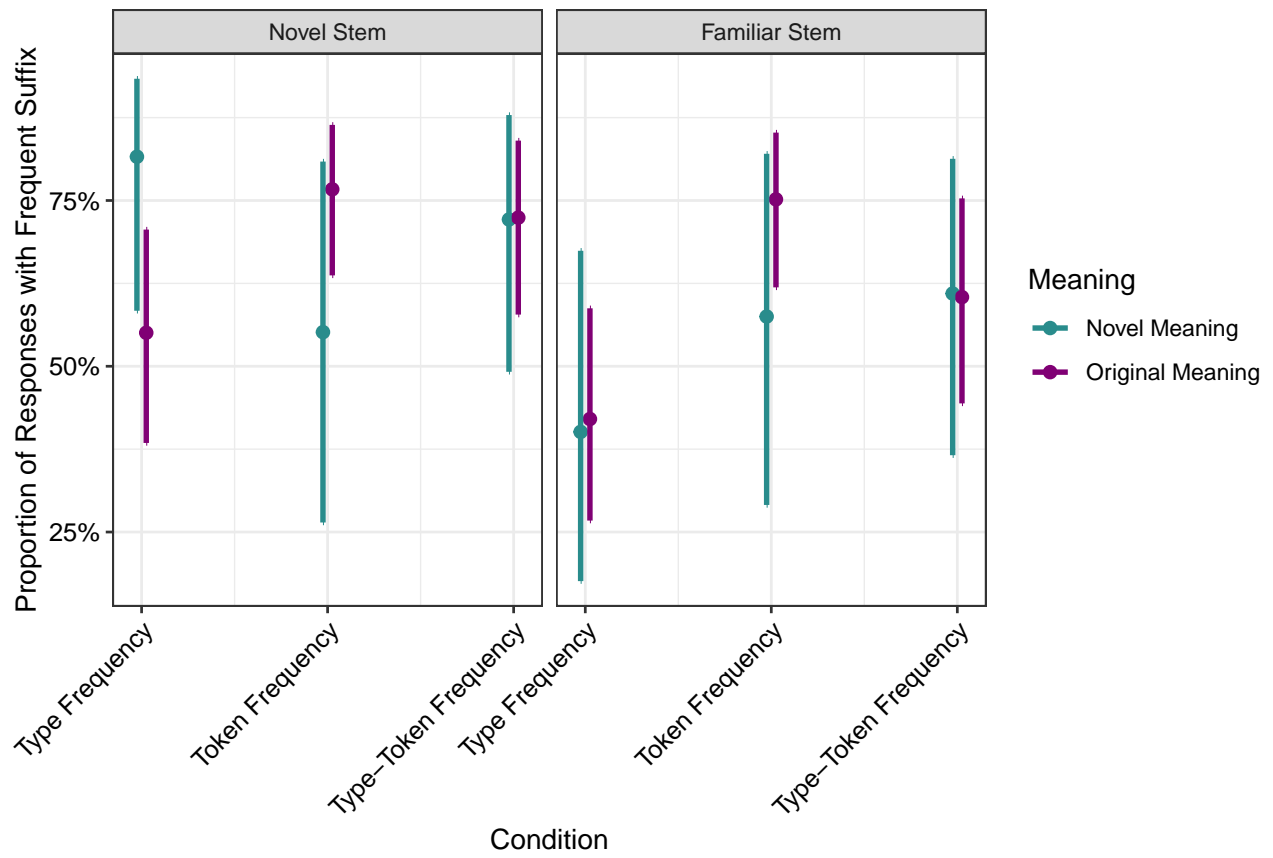


Figure 3: Plot of the statistical model estimates for the production task.

Table 4: Results of the statistical models for the 2afc task.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept (Type-Token Frequency)	-0.01	0.07	-0.15	0.12	40.99
Type Frequency	-0.12	0.10	-0.31	0.07	10.31
Token Frequency	0.05	0.09	-0.13	0.22	70.61
Novel Meaning	0.06	0.07	-0.08	0.19	79.22
Type Frequency:Novel	-0.18	0.09	-0.36	0.00	2.50
Token Frequency:Novel	-0.05	0.09	-0.22	0.12	30.24

size is quite small and their selections were still close to chance.³

Comprehension Task

Our results for the comprehension task are included below in Table 5. 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.

For comprehension, our independent variables were different from the other tasks. Specifically, to maximize our power, we divided the data up based on whether the token-to-type ratio was 12-12, 12-3, or 3-3. For example, in the Type Frequency condition, while the frequent suffix occurs 12 times with 12 different types, the infrequent occurs 12 times in only 3 different types. By comparing the 12-3 condition to the 12-12 and 3-3 conditions, we can see how an increase in type-frequency and a decrease in token-frequency affect participants' choice of the novel or original meaning.

We treatment coded condition such that the intercept was the 12-3 condition. As such, for each condition, a positive coefficient indicates that participants were more likely to choose a novel meaning (relative to the 12-3 condition/intercept). The model equation is included below in Equation 2.⁴

³We ran an additional model with stem condition as a fixed-effect (along with its interaction with condition and meaning) and found no effect of stem condition.

⁴Similar to form choice, we also ran a model with stem condition as a fixed-effect and found no effect.

Table 5: Results of the statistical models for the comprehension task.

	Estimate	Est.Error	Q2.5	Q97.5	% Samples > 0
Intercept (12-3)	-2.80	0.27	-3.35	-2.28	0
12-12	1.46	0.19	1.09	1.84	100
3-3	1.98	0.18	1.62	2.34	100

$$\text{meaning} \sim \text{condition_numeric} + (1|\text{participant}) \quad (2)$$

First, we find a meaningful effect for the intercept (12-3 condition), suggesting that participants are more likely to select the original meaning when there is a high token-to-type ratio. Additionally, we find positive coefficient estimates for both 12-12 and 3-3 suggesting that when there is an equal number of tokens as types, participants are more likely to select the novel meaning. In other words, entrenchment seems to be driven not simply by the raw value of token or type frequency, but rather the proportion of tokens to types. That is, it is the relationship between type-frequency and token-frequency that is relevant for semantic entrenchment in comprehension. This mirrors the results for our production task where a high type and high token frequency leads to semantic extension, but having a high token frequency with low type frequency leads to entrenchment.

Discussion

Our results suggest that in production, having a high type and high token frequency (relative to the competitor), along with simply having a high type but equal token frequency (relative to the competitor), leads to semantic extension, while having a high type frequency but low token frequency leads to entrenchment (i.e., using the suffix more with the original meaning than novel meanings). Further, this mirrors what we see in comprehension where participants are more likely to choose the novel meaning when the type and token frequencies are matched, but less likely when the suffix has high token frequency but low type frequency. Further, when both forms are made accessible as in the form choice task, this effect is reduced for all conditions, though less-so for the type-frequency

condition.

Harmon, Z., & Kapatsinski, V. (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>