A morphosyntactic inductive bias in artificial language learning*

Itamar Kastner & Tal Linzen

Humboldt-Universität zu Berlin, Johns Hopkins University

1. Introduction

A learner's job is to look at the data and find patterns. But in any given dataset, a large number of patterns can be found if the hypothesis space is not constrained. This is particularly true for language acquisition, where the learner has in principle infinitely many hypotheses to pursue, all of which match the data. Polar questions in English are formed by fronting the auxiliary before the subject: ***Is*** *the man who is tall __ in the other room?* When acquiring language, children seem to gradually acquire this rule, but they do not adopt other hypotheses (Crain & Nakayama 1987). For instance, they do not conclude that the sixth word of the sentence should be fronted, even though the auxiliary *is* starts off as the sixth word. The hypothesis space explored by the learner appears to be constrained in many parts of the grammar. It is therefore important to study which biases a human learner might bring with them when making generalizations over the data; not all patterns in the input are necessarily learned as such.

This conclusion has been supported by studies of phonological systems in which spurious patterns are dispreferred or simply not learned. For example, Becker et al. (2011) showed that unnatural constraints are not learned as productive patterns. But probing deeper, it can be shown that while unnatural constraints can be learned in principle, natural ones are preferred by learners (Hayes et al. 2009, Hayes & White 2013).

These studies all examined competing generalizations within the same domain (syntax or phonology). Yet generalizations from different domains can also compete with one another. Recent work has contrasted semantic with phonological cues in the acquisition of noun classes, aiming to tease apart which cues learners are more sensitive to. The Caucasian language Tsez has four noun classes which largely fall along semantic criteria—masculine human, feminine human, other animate and inanimate—although these criteria are not watertight. In their work on the language, Gagliardi & Lidz (2014) found that an additional

*Thanks to Katya Pertsova for discussion and to our colleagues at LSCP (Paris) and RUESHeL (Berlin). This study was supported by DFG grant AL 554/8-1 (DFG Gottfried Wilhelm Leibniz Preis 2014) to Artemis Alexiadou (I.K.).

phonological criterion exists. For inanimate nouns, the first segment functions as a predictor of noun class with fairly high accuracy. Gagliardi & Lidz (2014) asked adult speakers and children to classify nonce nouns in the language, finding that speakers classify nouns based on phonological information rather than semantic information.

While this finding is relevant for the question of what information learners are attuned to, it is important to note that the semantic information matched with Tsez nouns is not represented as morphosyntactic features. That is to say, there is no syntactic mechanism in the language sensitive to the presumable features [MASCULINE HUMAN], [FEMININE HUMAN], [NON-HUMAN ANIMATE] and [INANIMATE]. The Tsez case might be more similar to gender distinctions in European languages, where phonological and semantic cues both influence the assignment of gender to a given noun. Will there be a similar bias for the phonological generalization if the competing generalization is based on a morphosyntactic feature?

Artificial language paradigms can shed additional light on the question of what cues learners rely on. In a series of experiments, Culbertson et al. (2017) designed an artificial language in which noun class was assigned based on phonological and semantic information. Nouns were assigned to semantically-informed classes based on whether they denoted animate/inanimate, narrow/flat, or rigid/flexible objects. Phonological cues took the shape of a prefix and a suffix. The experimenters further varied whether the learners were exposed to one source of information (semantic or phonological) before the other. The results indicated that when cues conflicted, learners preferred the type of information they were exposed to first, and otherwise preferred more salient cues (e.g. animacy) to less salient ones (e.g. rigidity).

Even though such studies of noun classes are instructive, they do not target morphosyntactic features that are active in the language as a whole. Nouns, almost by definition, have semantic content. But this content is not necessarily part of the grammar, as in the case of Tsez, where the semantic distinctions are not directly related to the rest of the grammatical system. This semantic information is different from morphosyntactic features in that the latter can be seen as functions on an argument: a plural feature [PL] maps the noun onto a plural representation. A noun class has no deterministic semantic interpretation.

These previous studies have shown that under certain conditions, phonological information is preferred to semantic information about the noun. The current study tests what generalizations learners make when presented with competing generalizations. The task is to learn alternations in the morphology of an artificial language, where the input can form the basis for two distinct generalizations: one "contentful", morphosyntactic, and one surface-based, morphophonological. When the input is consistent with both a syntactic and a phonological generalization, which one do learners prefer? In our artificial language, learners were trained on word forms that underwent a phonological change which could be described by reference to either the syntactic or the phonological environment. They were then tested on held-out cases to see which of the two generalizations was learned (Wilson 2006).

The artificial language used in the current study contained input which is consistent with two distinct generalizations:

(1) **Syntactic generalization:** Lower the stem vowel before (all and only) plural suffixes.
Phonological generalization: Lower the stem vowel before (all and only) consonant-initial suffixes.

Experiment 1 found a preference for the syntactic generalization, and Experiment 2 additionally found that the order of conditions during testing influenced the learners' responses.

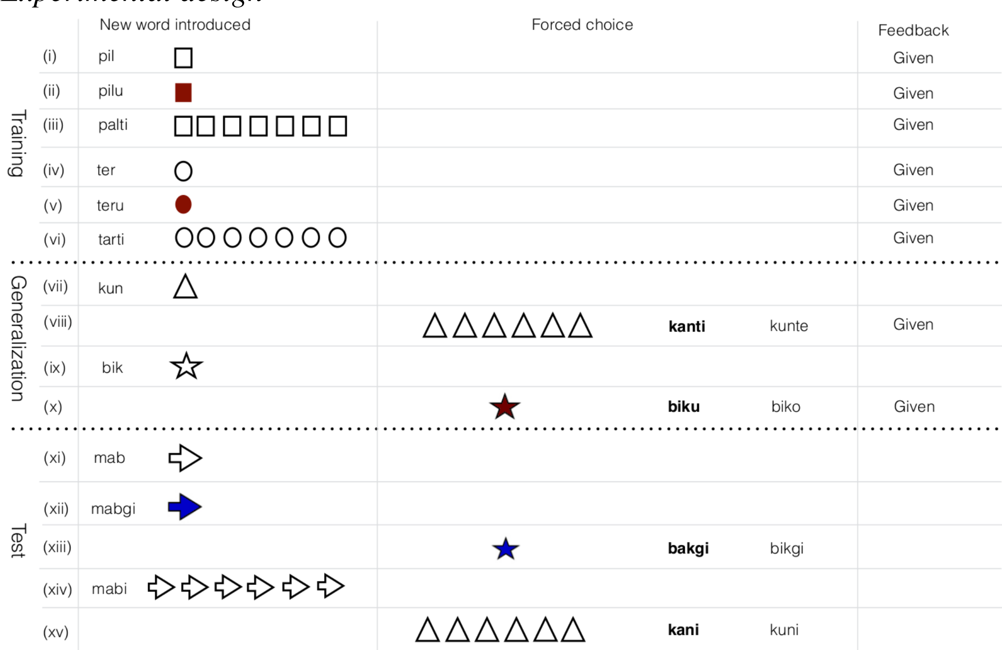
2. Experiment 1

2.1 Methods

2.1.1 Design

The study consisted of three stages, presented visually, as in (2), in two minimally different designs which we refer to as Experiment 1 and Experiment 2. Participants were presented with a language in which words were of the form CVC, or CVC-V and CVC-CV when suffixed. There were ten different words, assigned to ten different pictures, and four suffixes. All words had the prosodic form CVC (consonant-vowel-consonant). Words were assigned to shapes pseudorandomly in two different lists, so that the same word would not be assigned to the same shape in both lists. Two suffixes were used for training and two were introduced for testing, as explained below.

(2) *Experimental design*



Affixes varied across two dimensions: (morpho-)syntactic and (morpho-)phonological. Syntactically, the suffixes controlled either the plurality of the picture or its color. Phonologically, the suffixes began with either a consonant or a vowel. Two suffixes were introduced in the training stage and two in the test stage.

During **Training**, three words and two suffixes were introduced: V-initial (*-u*) for a red version of the shape and C-initial (*-ti*) for a plural version. The plural suffix triggered lowering of the stem vowel to *-a-* (words in this stage had only non-low stem vowels). Participants were taught what each word means, (i)–(vi) in (3). Plural stimuli consisted of the singular picture repeated between 5–8 times (randomly chosen for each plural trial).

In this stage participants heard a word and were shown a corresponding picture. They were then given two written options for the word and were instructed to click on the word they just heard, training them to match words to pictures in the new language. Feedback was given, indicating whether the choice was correct or not. Audio was synthesized using the Python package `gTTS`, an implementation of the Google Text-to-Speech API, with the voice `pl` (Durette 2017). Three stems were used in the training stage, chosen randomly from the list of eight stems. Each "paradigm" of singular, red and plural was shown between 2–3 times in sequence, and "paradigms" were randomly interleaved with each other.

During **Generalization**, five new words were introduced. Participants were asked to pick the correct "red" and "plural" forms and feedback was provided after each trial, (vii)–(x). This part ensured that the affixes were correctly generalized; participants who failed at least half the trials in this stage were excluded from analysis.

Participants were shown a novel word in orthography, its singular form picture, and a picture containing either the plural or red form. Five novel stems were introduced in this stage. The participants were then asked to choose between two written words corresponding to the new pictures: one correct, and one a distractor in which one of the phonemes was changed randomly (order of presentation, right vs left, was randomized). Feedback was given, indicating whether the choice was correct or not. The order of trials in Generalization was randomized.

Finally, at **Test**, we investigated which generalization was made. Two new suffixes were introduced: C-inital (*-gi*) for a blue version of the shape and V-initial (*-i*) for a plural version, (xi)–(xv). These suffixes were demonstrated using two new stems whose vowel was *-a-*. Since these stems already had a low vowel, they would be unchanged by either of the rules and would therefore be congruent with both generalizations. The novel stems were *mab* and *ram*: for example, participants saw the singular *mab*, the plural picture corresponding to it, and the word *mabi* as the plural form.

After the new suffixes were introduced, participants were tested on the five stems from the generalization stage but with the new suffixes. The singular form and its corresponding picture appeared, a plural picture appeared, and the participant was asked to choose between two written forms. Participants were asked to pick the correct forms from two choices, one faithful to the stem and another with lowering. The order of trials was randomized. In addition, each participant was randomly assigned to one of two word lists, each list matching the stems to different pictures.

The new suffix was put to use when the learner was presented with a plurality and asked to choose between e.g. *teri* and *tari* as plural forms. Choosing the latter is consistent with the syntactic generalization, while choosing the former is consistent with the phonological one. In the blue part of testing, the learner was presented with a blue form and asked to choose between e.g. *tergi* and *targi*. In this case, choosing the latter would be consistent with the phonological generalization.

2.1.2 Participants

256 participants were recruited. All participants self-identified as native speakers of English older than 18. Participants were recruited via Prolific Academic and were paid £1 ($1.25) for participation in the study. The experiment concluded with a written debriefing.

2.1.3 Analysis

Participants were excluded from analysis if they did not answer more than half of the Generalization trials correctly (despite receiving feedback after each trial). This criterion, together with the discarding of incomplete studies and similar exclusions, resulted in $N = 191$.

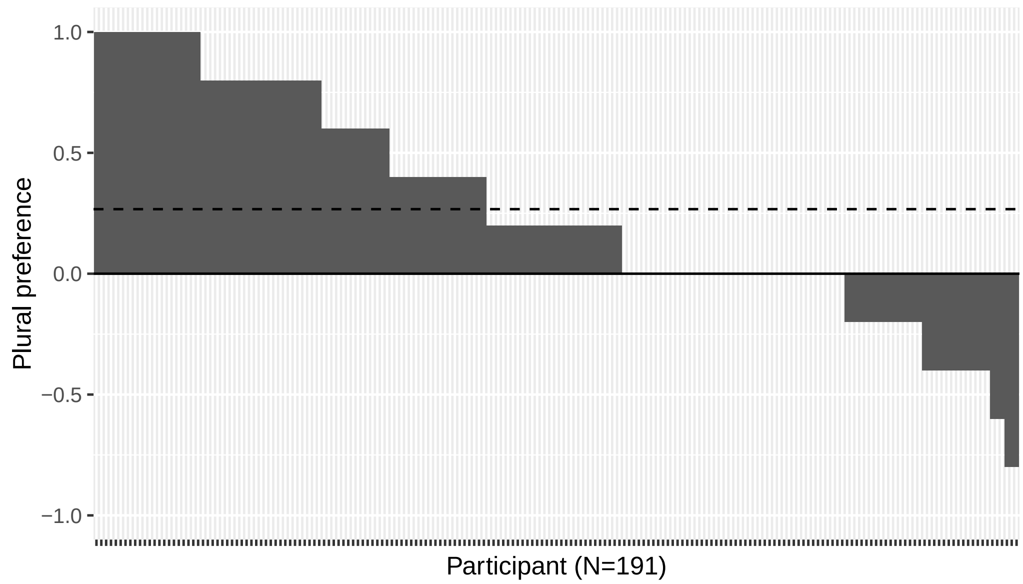
Logistic mixed-effects models (Baayen et al. 2008) were fitted to the participants' responses (i.e. whether they chose the lowered or non-lowered word) using version 1.1.12 of the `lme4` package in R (Bates et al. 2015). The random effect structure for all models included a by-subject slope for the effect of affix type, as well as by-subject and by-item intercepts. The statistical significance of affix type (plural or blue) was assessed by comparing the likelihood of the model to the likelihood of a baseline model that did not include affix type as a predictor, but did include the random effects.

2.2 Results

Generalization stage trials were answered correctly 75.4% of the time (69.6% for the plural/C-initial suffix and 81.2% for the red/V-initial suffix).

Results showed a main effect of **Condition**, with novel plural forms being lowered more than novel blue forms (54.6% vs 27.8% of the time), $\chi^2(1) = 53.77$, $p < 0.0001$. The figure in (3) plots the results by participant as the rate of novel plural lowerings minus the rate of novel blue lowerings; most participants lowered the novel plural forms more often than the novel blue forms. The dashed line gives the average rate of plural preference across participants, 26.8%.

(3) *Experiment 1 preference for lowering by participant*



There was no effect of Word List (between the two word lists) or Presentation Side (whether the word for a picture was presented on the left or right side of the screen).

Summarizing briefly, we found a general preference for the syntactic generalization over the phonological generalization.

3. Experiment 2

3.1 Methods

3.1.1 Design

Experiment 2 further tested whether the method of presentation further biases learners. Trials were randomized within blocks: presentation of Generalization trials was counterbalanced between first showing plural forms and first showing color forms, as was the case in Test.

Specifically, Generalization trials were blocked and counterbalanced such that for half of the participants, the "plural" trials preceded the "red" trials (and vice versa). In Test, for half of the participants, the novel "plural" trials preceded the novel "blue" trials (and vice versa). In total, Experiment 2 yielded four lists, counterbalancing order of blocks in Generalization and Test. The remainder of the design and procedure were identical.

3.1.2 Participants

256 participants were recruited according to the same criteria as in Experiment 1 (64 participants for each of the four lists).

3.1.3 Analysis

The same analysis method was followed, leading to $N = 237$ after exclusions.

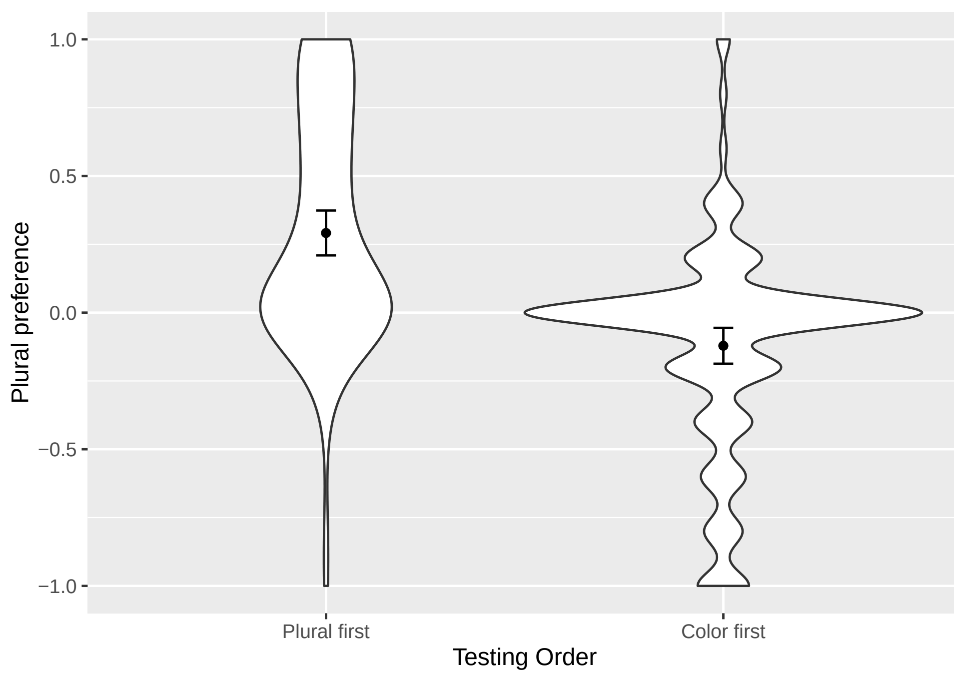
3.2 Results

Generalization stage trials were answered correctly 81.9% of the time (76.6% for the plural/C-initial suffix and 87.1% for the red/V-initial suffix).

Running the same analysis as in Experiment 1, **Condition** was still significant, exhibiting a preference for novel plural lowerings (31% of the time) over novel blue lowerings (25%), $\chi^2(1) = 15.913$, $p < 0.0001$.

When adding Order of Test as a predictor, a main effect of **Condition** persisted ($\chi^2(1) = 3.949$, $p = 0.0469$), accompanied by a main effect of **Order of Test** (novel plural firsts vs novel blues first), $\chi^2(1) = 17.21$, $p < 0.0001$, and an **interaction**, $\chi^2(1) = 51.896$, $p < 0.0001$. The violin plots in (5) bring out the interaction as preference for lowering novel plural forms by Order of Test, by participant.

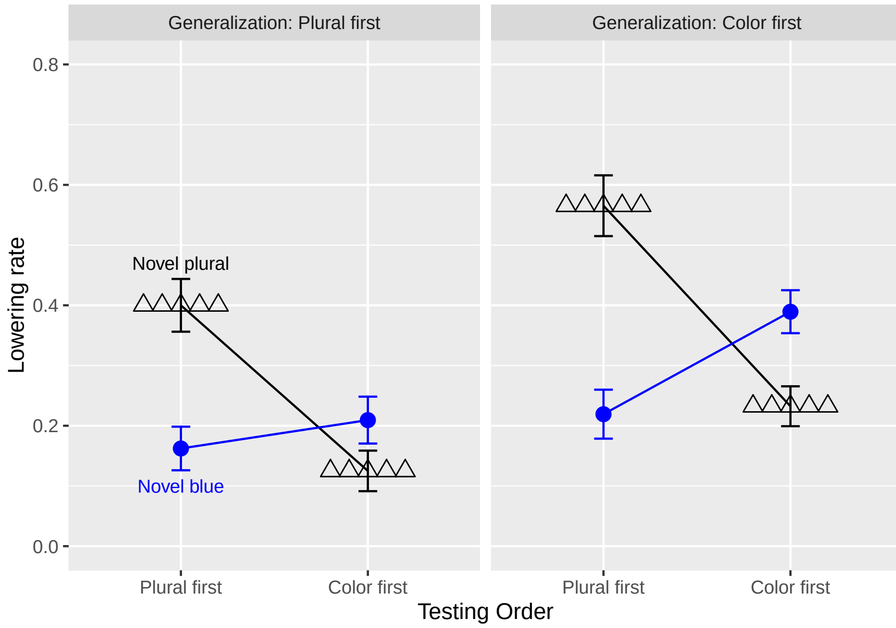
(4) *Experiment 2 preference for lowering by participant*



Finally, when **Order of Generalization** was added as predictor, it too showed significance ($\chi^2(1) = 11.665$, $p = 0.00064$) without interacting with any of the other factors. The figure in (5) plots the mean lowering rates per condition across participants. Error bars plot 95% CIs within subject.

There was no effect of Word List or Presentation Side.

(5) *Experiment 2 preference for lowering across participants*



4. Discussion

4.1 Learning biases

In our experiments, we asked whether a learner would show a bias towards morphosyntactic or morphophonological generalizations, when both can be invoked to explain patterns of allomorphy in an artificial language. The results show that learners prefer the syntactic generalization. Experiment 1 found that learners prefer the syntactic generalization, when order of presentation was randomized in the Generalization and Test stages.

Throughout the paper we have referred to the plural condition as “syntactic”. This label is meant to contrast with the experimental design of Culbertson et al. (2017), who presented pictures of concrete items differing in *saliency* along the axes of animacy, shape and flexibility. For example, in a high salience semantic condition, nouns were assigned to one of two classes based on whether they denoted animate or inanimate targets; this is a frequent pattern typologically. In a low salience condition, nouns were assigned to one of two classes based on whether the object they denoted was rigid or flexible. This distinction is not one commonly found crosslinguistically.

As noted in the introduction, these are semantic characteristics. It is not common for a language to formalize them as syntactic features. Similarly, the noun classes in Tsez exploited by Gagliardi & Lidz (2014) could be differentiated on semantic grounds which did not translate into syntactic features (and only to a limited extent).

The current study has attempted to pit phonological cues against syntactic ones. But there is no clear distinction between syntactic and semantic information; while we chose plurality because it is a syntactic feature in many languages (and red/blue is not), plurality is clearly semantically interpretable as well. In addition, the syntactic cue might arguably be more salient than the phonological one. In future work we hope to put the learning question to a more rigid test by using additional (morpho-)syntactic phenomena, beyond the conditioning of allomorphy.

4.2 Order of learning

Experiment 2 asked whether ordering the trials in blocks can bias either learning or performance on the task. While the overall preference for the syntactic generalization was still evident, it was qualified by the effect of presentation order. In particular, **Order of Test**, **Order of Generalization**, and the interaction of **Order of Test** with **Condition** all turned out to be significant factors for the task. It appears that where testing was concerned, participants showed a preference for whichever condition was tested on first (**Order of Test**), qualified by an overall preference for the plural condition.

We speculate that when the input is presented in blocks, learners first entertain both hypotheses, and then wait for a cue to adopt one. In Experiment 1 no such cue was forthcoming, so the general preference for the syntactic generalization took over. In Experiment 2, participants might have still been entertaining both hypotheses when they were presented with the novel blue forms as the first test block. They would have then applied that generalization to the novel items; by the time the second block appeared, they had already chosen to apply the lowering rule to one condition, and did not extend it to the other one.

This situation is reminiscent of recent work by Moreton & Pertsova (2017), who contrasted *implicit learning* with *explicit learning*. The former is similar to the current Experiment 1, where the order of trials is randomized across conditions. The latter is more similar to our Experiment 2, where conditions are blocked separately from each other, prompting learners to learn the conditions as distinct from one another.

4.3 Syntactic vs phonological generalizations in a natural language

The patterns contrasted in the artificial language were inspired by the real acquisition task of a natural language, namely Modern Hebrew. In this language, verbs appear in a "template" of consonants and vowels which instantiate a consonantal "root". The three verbs in (6) all share the root *g-d-l*. Of relevance here is the last vowel of the stem, before the agreement affix. It is different depending on whether the subject is 1st and 2nd person or 3rd person. Boldfaced vowels show the difference in agreement: /a/ above the line (1st and 2nd person), /e/ or /i/ below it (3rd person), depending on the template.

(6) *Alternating vowels in Hebrew allomorphy*

	hegdil 'enlarged'	
	SG	PL
1	hegd**a**l-ti	hegd**a**l-nu
2M	hegd**a**l-ta	hegd**a**l-tem
2F	hegd**a**l-t	hegd**a**l-tem
3M	hegd**i**l	hegd**i**l-u
3F	hegd**i**l-a	hegd**i**l-u

	gidel 'grew'	
	SG	PL
1	gid**a**l-ti	gid**a**l-nu
2M	gid**a**l-ta	gid**a**l-tem
2F	gid**a**l-t	gid**a**l-tem
3M	gid**e**l	gid**e**l-u
3F	gid**e**l-a	gid**e**l-u

The generalizations discussed above for the artificial language are the same ones that hold for Hebrew: the vowel changes when the meaning changes, and/or the vowel changes when the suffix is C-initial. While there are other theoretical reasons to believe that the grammatical generalization is correct (Kastner 2016, Ch. 3.4.2), it is possible that during acquisition, the child would first notice the phonological cue rather than the grammatical one. The experiment described above evaluated how such acquisition might progress.

References

Baayen, R. Harald, Doug J. Davidson, & D.M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59:390–412.

Bates, Douglas, Martin Mächler, Ben Bolker, & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67:1–48.

Becker, Michael, Nihan Ketrez, & Andrew Nevins. 2011. The surfeit of the stimulus: Analytic biases filter lexical statistics in Turkish devoicing neutralization. *Language* 87:84–125.

Crain, Stephen, & Mineharu Nakayama. 1987. Structure dependence in grammar formation. *Language* 63:522–543.

Culbertson, Jennifer, Annie Gagliardi, & Kenny Smith. 2017. Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language* 92:343–358.

Durette, Pierre-Nick. 2017. gTTS. Version 1.1.8 [software]. http://github.com/pndurette/gTTS.

Gagliardi, Annie, & Jeffrey Lidz. 2014. Statistical insensitivity in the acquisition of Tsez noun classes. *Language* 90:58–89.

Hayes, Bruce, & James White. 2013. Phonological naturalness and phonotactic learning. *Linguistic Inquiry* 44:45–75.

Hayes, Bruce, Kie Zuraw, Peter Siptar, & Zsuzsa Londe. 2009. Natural and unnatural constraints in Hungarian vowel harmony. *Language* 85:822–863.

Kastner, Itamar. 2016. Form and meaning in the Hebrew verb. Doctoral dissertation, New York University, New York, NY. Lingbuzz/003028.

Moreton, Elliott, & Katya Pertsova. 2017. Implicit and explicit processes in phonotactic learning. In *Proceedings of the 40th Boston University Conference on Language Development (BUCLD 40)*, ed. Jennifer Scott & Deb Waugtal, 277–290. Somerville, MA: Cascadilla Press.

Wilson, Colin. 2006. Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science* 30:945–982.

Itamar Kastner, Tal Linzen
itamar@itamarkast.net, tal.linzen@jhu.edu