Predictability of meaning in grammatical encoding: Optional plural marking

Chigusa Kurumada[a,*], Scott Grimm[b]

[a]*Department of Brain and Cognitive Sciences, University of Rochester, United States*
[b]*Department of Linguistics, University of Rochester, United States*

Abstract

The *markedness* principle plays a central role in linguistic theory: marked grammatical categories (like plural) tend to receive more linguistic encoding (e.g., morphological marking), while unmarked categories (like singular) tend to receive less linguistic encoding. What precisely makes a grammatical category or meaning marked, however, remains unclear. One prominent proposal attributes markedness to the frequency or predictability of meanings: infrequent or less predictable meanings are more likely to receive extra linguistic encoding than frequent or more predictable meanings. Existing support for the predictability account is limited to correlational evidence, leaving open whether meaning predictability can *cause* markedness patterns. We present two miniature language learning experiments that directly assess effects of predictability on morphological plural marking. We find that learners preferentially produce plural marking on nouns that are less probable to occur with plural meaning—despite the fact that no such pattern was present in learners' input. This suggests that meaning predictability can cause the markedness patterns like those that are cross-linguistically observed.

Keywords: markedness, plural marking, optional number marking, predictability, miniature language learning, learning biases

*Corresponding author. Address: Department of Brain and Cognitive Sciences, University of Rochester, 304 Meliora Hall, Rochester NY, 14627, United States.
Email address: `ckuruma2@ur.rochester.edu` (Chigusa Kurumada)

1. Introduction

The mapping from linguistic forms (e.g., words forms) to meanings is largely arbitrary (De Saussure, 2011; Peirce et al., 1974). There is no principled reason why the word form *dog* rather than some other word form indicates canines. There are, however, certain more abstract cross-linguistic generalizations about the mapping of meaning to linguistic forms. One striking generalization is captured by the principle of "markedness" (Trubetzkoy, 1931; Jakobson, 1932). Across languages of the world, some grammatical categories (e.g., plural, past tense) are more likely to be realized with additional linguistic encoding, while other grammatical categories (e.g., singular, present) are realized without additional linguistic encoding. In English, for example, plural meaning is marked with the suffix -*s*, whereas the singular does not require additional morphology. Similar patterns of plural marking—where plural meanings receive more linguistic encoding than singular meanings—are observed in the vast majority of the world's languages (Dryer & Haspelmath, 2013). This and similar patterns for other grammatical categories are summarized in the influential generalization that marked meaning receives marked form (Jakobson, 1932; Horn, 1984; Levinson, 2000).[1]

Despite the seemingly straightforward nature of this generalization, its causal foundations remain controversial: What makes a meaning or grammatical category marked in the first place? That is, what causes some meanings to develop additional linguistic encoding across the world's languages with such typologically regularity? Two broad classes of proposals can be distinguished. One approach, appearing as early as in Jakobson (1932), attributes the markedness of a meaning to its relative semantic or conceptual *complexity*. For instance, in the case of grammatical number, the marked status for plural coding has been argued to follow from marked status of the meaning or concept of plurality relative to singularity. That is, the concept of MORE THAN ONE builds on and is complex than the concept of ONE. The second class of proposals grounds the notion of markedness in the frequency or *predictability* of the meaning. This proposal finds its roots in Greenberg (1966), who first articulated the correspondence between unmarked, default grammatical categories with higher frequency of use in comparison and the

[1] The notion of markedness has been influential in nearly all areas of of linguistic theorizing from phonetics (Chomsky & Halle, 1968; Kean, 1975) to pragmatics (Horn, 1984; Levinson, 2000). Here we focus solely on markedness in relation to grammatical categories, namely, plural and singular number.

marked, non-default categories. In the case of grammatical number, the category of singular is taken to be the default, as opposed to plural, since language users are overall more likely to talk about singular entities than plural entities.

Greenberg himself did not seem to see these two accounts as necessarily in conflict. The semantic complexity of a grammatical category and its frequency of use are typically not independent of each other. For instance, the grammatical number category *dual*, used in some languages to designate precisely two entities, is comparatively more complex or specific than the plural, since the meaning of a plural form (two or more) includes the meaning of a dual form (exactly two) (Greenberg, 1963). The specificity of the dual form naturally limits its usage frequency. For Greenberg, frequency of use, while a powerful motivator in linguistic evolution, is not the final cause of markedness: "frequency is itself but a symptom" (Greenberg, 1966, p. 65). Later proposals, however, clearly hypothesized frequency of use (Haspelmath, 2006; 2008; Haspelmath & Karjus 2017) or predictability (Hume, 2004; 2008) to be the driving force of the correlation between meaning and form. For example, Haspelmath and Karjus (2017, p.1218) write "(L)anguages generally use more coding for less predictable meanings". For now we group proposals of this type under the label "predictability hypothesis of markedness". In the discussion, we return to differences among versions of this class of proposals.

Despite the central role of markedness in linguistic research, direct tests of either of the two competing hypotheses—conceptual complexity or frequency/predictability of meaning—have been lacking. Addressing this gap is critical in providing an explanation for a fundamental organizational principle of human languages. The present study tests the hypothesis that the frequency/predictability of a grammatical category (e.g., plural meaning) can cause that meaning to receive less linguistic encoding. We approach this question through two miniature language learning experiments that investigate biases during language learning. Specifically, we ask whether there is a bias for learners to produce less linguistic encoding for more predictable meaning, even when such a correlation is not present in the language input.

Such a bias would be compatible with psycholinguistic research on the effects of predictability during language production. This line of work has found, for example, that the same word tends to be realized with shorter duration and less phonetic detail when it occurs in a predictive context, compared to when it occurs in a less predictive context (Aylett & Turk, 2006; Bell et al., 2009; Ernestus, 2014; Pluymaekers et al., 2005; Seyfarth, 2014). While the majority of this research has focused on the phonetic reduction of predictable words, there is also evidence that the predictability of *grammatical categories* can affect language production. Of direct relevance to the present study, language production experiments have found that speakers are more likely to omit optional affixes when the meaning of the affix is predictable in context (e.g., Kurumada & Jaeger, 2015; Norcliffe & Jaeger, 2016). For example, Japanese speakers are more likely to produce case-marking, which is largely optional in spoken Japanese, when the grammatical function of the argument is less predictable in context (Kurumada & Jaeger, 2015). There is also evidence that the predictability of plural meaning affects the pronunciation of the plural suffix *-s* in spoken New Zealand English: the more predictable plural meaning is given the noun, the shorter on average is the duration of the suffix (Rose, 2017).

Findings like these are expected under the hypothesis that predictability plays a causal role in the reduction of form (for a review and discussion, see Jaeger & Buz, 2017). However, the correlations between meaning predictability and linguistic encoding that is observed in these studies are already present in the language input native speakers receive during acquisition (for related discussion, see Bybee, 2002; Bürki et al., 2010; Jaeger, 2006, 2010; Seyfarth, 2014). This leaves open whether predictability can *cause* the type of markedness pattern that has been cross-linguistically observed.

1.1. The present study

In the current study, we exposed participants to miniature languages that did *not* already contain the hypothesized correlation between predictability of grammatical number category (singular vs. plural meaning) and its linguistic marking in the input. We manipulated the predictability of singular vs. plural meaning, and tested whether predictability could cause learners to induce the hypothesized correlation in their productions. Figure 1 illustrates the design of the two experiments we present. Both experiments employed two types of referents that differed in their predictability of plural meaning. During exposure, a class of nouns called 'individuals' occurred with singular meaning (singleton visual referents) 75% of the time and with plural meaning (multiple visual referents) 25% of the time. This contrasted with another class of nouns called 'collectives', for which the ratio was reversed. In the input during exposure, plural meaning was thus more predictable for collectives than for individuals.

Like in English, singular meanings were never marked with additional linguistic coding in our miniature language. Unlike in English, plural marking was not obligatory in our language, and occurred only on 66% of the nouns that referred to multiple referents. Optional plural marking is, in

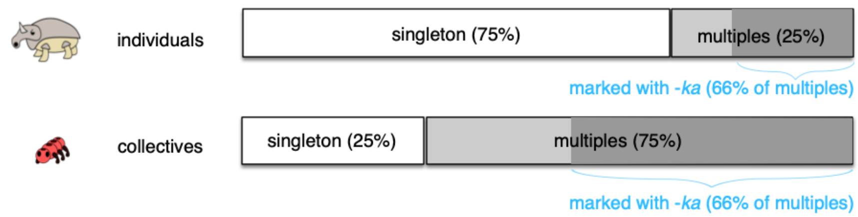

Figure 1: Illustration of design. Both Experiment 1 and 2 employ two types of referents that differed in the predictability of plural meaning. During exposure, 'individuals' occur predominantly singular meanings (singleton referents), whereas 'collectives' occur predominantly with plural meanings (multiples). In the input, the proportion of plural marking was held constant across individuals and collectives (66% of all references to multiples occur with the marker -*ka*; dark gray). The predictability hypothesis of markedness predicts that learners will exhibit a bias to increase the proportion of plural marking for individuals, and/or decrease the proportion of plural marking for collectives.

fact, cross-linguistically common: Approximately 34% of the 291 languages in Haspelmath (2013) are classified as manifesting optional plural marking. Critically, in the input for our miniature language the proportion of plural marking (-*ka*) was identical for the two classes of referents (66%, indicated in dark gray in Figure 1). After exposure, participants completed a test phase during which they described novel visual scenes in the miniature language. If learners were indeed more likely to produce plural marking for nouns for which plural meaning is less predictable, this should influence learners' productions during the test phase. Specifically, the predictability hypothesis of markedness predicted that learners would tend to increase the proportion of plural marking for individuals and/or decrease the proportion of plural marking for collectives.

Our experiments were inspired by two independent lines of work. One is a body of typological research on "markedness reversals" (Tiersma, 1982). Markedness reversal often arises in grammatical number systems when a language employs a mixed system to convey singular or plural meaning (e.g., in Murle, Arensen 1998; Dagaare, Grimm 2012a; Maltese, Mifsud 1996; and Welsh, Stolz 2001; see discussion in Grimm 2018; Haspelmath & Karjus 2017). Some nouns in these languages follow the pattern observed in English, where bare (unmarked) nouns designate singleton entities and plural designations require suffixation. For example, the Welsh noun *tad* "father" is bare and singular and plural meaning is produced through suffixation, namely *tadau* "fathers". Unlike English however, Welsh also exhibits the reverse pattern on some nouns. For these nouns, the bare form has a plural meaning by default, and additional linguistic encoding is used to convey singular meaning. For example, *psy* is bare but designates plural “peas” and an additional suffix *-en* is required to designate the singular meaning (the individual entity).

These cases are problematic for any account that reduces markedness to language-wide properties of the grammatical category, such as the complexity or *overall* frequency of plural meaning. But they can be accommodated if the number marking pattern is at least partly determined by the probability that the noun will occur with plural meaning (for discussion, see Grimm, 2012b; Haspelmath & Karjus, 2017). That is, the plural meaning is generally more predictable for *pea* than for *father* since speakers often talk about more than one pea whereas the opposite is true for *father*. Additional linguistic encoding (suffixation) thus may indicate grammatical number (singular or plural) that is less predictable given the referential meaning of the noun. Existing cross-linguistic evidence is compatible with this hypothesis.

For instance, Haspelmath & Karjus (2017) indirectly estimated the predictability of plural meaning for 18 Welsh nouns through the frequencies of meaning equivalents in five different languages. (A sufficiently large Welsh corpus was not available). Translation equivalents of Welsh nouns with markedness reversal (e.g., *pea*) were indeed more likely to occur in the plural form, than in the singular form. For example, the English word *pea* occurs 173 times in the British National Corpus (British National Corpus Consortium, 2007), where the word *peas* occurs 603 times. This contrasts with the majority of nouns in English, for which plural forms are less frequent than singular forms. For example, using the same British National Corpus, we find that the frequency of *father/fathers* is 22508/1181, respectively. Correlational evidence like this is in line with the predictability hypothesis of markedness, but constitute relatively weak evidence. Our experiments aim to directly test whether the predictability of plural meaning can cause asymmetries in plural marking. Miniature language learning thus offers a complement to typological work.

The second line of research immediately relevant to our goal are miniature language learning experiments on typological variations and language change (e.g., in phonology, Moreton & Pater 2012; White 2014, morphology Finley & Badecker 2009; Schumacher et al. 2014; morpho-syntax, Christiansen 2000; Culbertson & Adger 2014; Fedzechkina et al. 2016a; word order, Tily et al. 2011; Fedzechkina et al. 2017). We leverage the insight brought forward by this line of work, in particular that language learners induce structure into languages with unconditioned “free” variation—such as the optional plural marking in our experiments (e.g., Culbertson & Smolensky, 2013; Hudson Kam & Newport, 2005; Hudson Kam, 2015; Fedzechkina et al., 2016b; Singleton & Newport, 2005). For example, Smith & Wonnacott (2010) exposed participants to a miniature language with two plural markers. In the input, the relative frequency of one marker over the other was held constant across all nouns (similar to our miniature language, except varying the presence vs. absence of a plural marker, rather than the choice between two markers). After exposure to the language, participants had to describe unfamiliar scenes. Although such patterns were not present in the input, participants tended to condition the choice between plural markers on properties of the referent (e.g., using one plural marker for animal referents and another for human referents). As Smith and Wonnacut point out, this allows learners to reduce the overall complexity of the plural marking system. Findings like this suggest that learners are biased to simplify grammatical systems by removing or conditioning variation (for further evidence, see Fedzechkina et al., 2012; Kirby et al., 2008; Reali & Griffiths, 2009; Samara et al., 2017; Smith et al., 2017).

Building on these two lines of research, we set out to test the predictability hypothesis of markedness. In particular, we test (1) whether learners condition plural marking on plural meaning predictability; and (2) whether learners would indeed simplify the optional plural marking system in the direction compatible with the predictability hypothesis of markedness: more predictable meaning receives less linguistic marking.

Following the presentation of our experiments, we relate our findings to recent research on the causal link between word frequency and reduction (Kanwal et al., 2017; Sóskuthy & Hay, 2017). These studies differ from the present study in that they focus on frequency rather than predictability, and on the lexicon rather than grammatical encoding. But, as we highlight in our discussion, there is a common theme emerging between these studies and ours. In that context, we return to the notion of meaning predictability, and discuss *why* predictability affects production, a topic of ongoing and productive debates in psycholinguistics (see Arnold et al., 2012; Ernestus, 2014; Jaeger, 2013; MacDonald, 2013). We then highlight implications of the current work for future studies in language production and empirical investigations of markedness.

2. Experiment 1

In Experiment 1, we put into play three factors—manipulated jointly, rather than independently—combining to increase the plural predictability of collective referents, compared to the plural predictability of individuals.

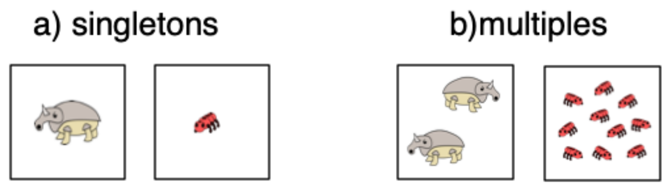

Figure 2: Examples of the visual referents presented during exposure in Experiment 1.

First, as described above, we manipulated the proportion of times that collectives and individuals occurred with multiple rather than singleton referents (Figure 1). Second, we designed the visual referents to resemble entities which participants would likely have prior experience with in their native language (see Figure 2). Specifically, we used fictitious large animals to depict individuals, and fictitious small insects to depict collectives. In general, large animals and humans are more likely to be individuated (i.e., conceptualized as individuals) compared to the insects, grains, and fruits, which are more likely to be conceptualized as collectives (see Grimm, 2018). This was done to encourage participants to transfer their semantic knowledge from prior native language experience to the experimental context, so they would assign low plural meaning predictability to individuals (animals), and high plural meaning predictability to collectives (insects). This was revisited in Experiment 2, which is designed to reduce the influence of native language experience.

Third and finally, the visual grouping for multiples of the two classes of referents was manipulated so as to further facilitate individuation of the large animal referents, and to do the opposite for insects (Figure 2). Specifically, during exposure, multiples of large animals were pairs of referents, whereas multiples of insects always involved many (10) referents. This also meant that an individual instance of animals looked larger compared to that of insects. In short, Experiment 1 manipulated three properties that would together increase the predictability of plural meanings for collectives, compared to individuals.

2.1. Methods

2.1.1. Participants

48 native speakers of American English at University of Rochester participated in this study. They received \$10 for their participation. They were all native speakers of American English without an advanced command of any other language. We a priori decided to exclude subjects who failed to

(1) achieve 65% accuracy in four-alternative forced choice trials and (2) produce 50% of sentences in the final production test. This led to the exclusion of 6 subjects (see Result section).

2.1.2. Materials

Visual stimuli. We created 12 visual referents for the fictitious animals and insects (six each). We used the software GIMP (https://www.gimp.org/) to draw 24 still images to be used in the training phase of the experiment. 50% of them depicted only one instance of an animal or an insect (singletons) and the rest of them more than one (multiples) (Figure 2). Based on these still images, we created 36 video (GIF) clips (12 singletons, 12 multiples for training, and 12 multiples for test) using the software Inkscape (https://inkscape.org/en/). In these clips, animals and insects showed a vertical jumping movement. As in the still images, the video clips of multiples were slightly altered between the two noun classes to present animals as more individuated and insects as gregarious (coming in greater numbers): Two instances of animals moved up and down in a staggered manner (e.g., one moving up and down while the other moving down first) while instances of the insects moved up and down together as a group. We used the Exbuilder software (a custom built experiment program created at the University of Rochester) to present the stimuli and record productions.

The language. The vocabulary of miniature language we use constitutes a subset of Fedzechkina et al. (2012). We selected 12 nonce nouns. All of the nouns were one to two syllables, obeying English phonotactics (e.g, *norg*, *velmick*, *zamper*). When characters were presented as multiples, the noun was optionally suffixed with the plural-marker (*–ka*) 2/3 of the time. To avoid potential confounds due to phonological or lexical factors, we constructed two lists by making two different mappings between words (e.g., "norg") and visual referents (e.g., an ant-like insect). Between-participants, each word was used equally often with an individual referent and a collective referent. Participants were randomly assigned to one of the two lists.

The language included only one verb (*glim*), meaning "moving up and down." Sentences in our language followed a simple subject verb order (e.g., *norg-ka glim*). The verb was never contained any plural marking regardless of the form of the noun.

Auditory stimuli (12 nouns, *–ka*, and one verb) were recorded by a male native speaker of American English. Plural marked forms were created for each noun by adding the audio recording for *–ka* using Praat (http://www.fon.hum.uva.nl/praat/). We then created sentences by joining the recording of nouns (with or without *–ka*) and the recording of the

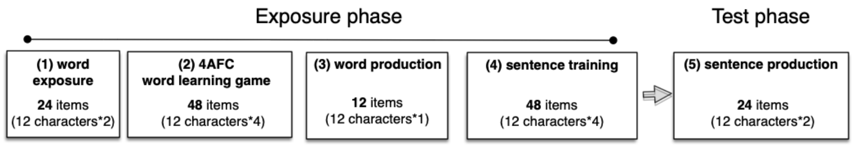

Figure 3: Experimental flow in the exposure and the test phases in Experiments 1 and 2.

verb (*glim*). This procedure was used to control the prosodic contour of phrases and sentence as well as to ensure that –*ka* and *glim* have identical phonological forms across items. If we had recorded natural production of sentences, the short morpheme –*ka* would have slightly different acoustic realizations depending on a preceding syllable, which would leave open whether participants would perceive it as one or more than one new plural marking morpheme.

2.1.3. Procedure

We employed an exposure-test paradigm (Figure 3), following previously conducted miniature language learning experiments (e.g., Fedzechkina et al., 2012). During exposure (phases 1-4), participants first learned the nouns and then the verb. Participants first completed phases (1)-(3) for six of the 12 noun types (three animals and three insects), and then repeated the same procedure to learn the other six words before continuing to phase (4). During test, they produced the description of novel scenes.

(1) Word exposure (12 characters * 2 = 24 trials total). Participants were presented with pictures of each of the characters and instructed to repeat the names of the characters aloud. All the characters were presented as singletons (and thus never with plural marking).

(2) Four alternative forced choice word learning game (12 characters * 4 = 48 trials total). The initial word presentation was followed by a word learning phase where participants were presented with four pictures (4 Alternative-Forced-Choice, 4AFC, task) and asked to choose the correct match for the noun provided (48 trials total). Feedback was provided after each trial. In this phase, individuals and collectives were presented as singletons and multiples at different rates. Individuals occurred 75% of the time as a singleton (i.e., one animal, Figure 2a), and 25% as multiples (Figure 2b). This ratio was applied to instances of characters used as non-target (distractor) objects on the screen. Collectives had the inverse distribution (25% singleton, 75% multiples). Both individual and collective nouns were followed by the plural-marker (-*ka*) 2/3 of the time when occurring as multiples. The plural marker -*ka* was never used for singletons.

(3) Word production (12 characters * 1 = 12 trials total). Participants were shown 12 characters (singleton) one by one and asked to name each of them. They received no feedback.

(4) Sentence training (12 characters * 4 = 48 trials total). Participants viewed short clips and heard their descriptions in the novel language. They were then asked to repeat the sentences out loud. This phase was included to introduce the verb (*glim*) as well as the word order information (Subject-Verb). As in the word learning phase, individuals and collectives occurred as singletons 75% and 25% of the time, respectively, and they were followed by the plural-marker (*-ka*) 2/3 of the time when occurring as a multiple. Consequently, participants heard the animal and insect nouns with *-ka* 10 times and 30 times, respectively, by the end of this phase (Figure 3).

(5) Sentence production (12 characters * 2 = 24 trials total). In the final test (sentence production) phase, participants saw silent videos of singletons and multiples and produced intransitive descriptions. Each character appeared twice, once as a singleton and once as multiples. In this phase, visual prompts (clips) for the multiples had three instances of the characters both for animals and insects. This was done to ensure that participants use *-ka* to signal plurality rather than the particular number of instances (two for animals and ten for insects) seen in the exposure input.

2.1.4. Scoring and exclusions

Comprehension accuracy in phase (2). In the 4AFC comprehension test, participants' responses were scored as "correct" if they matched the intended referent. Following the standard used in similar studies (e.g., Fedzechkina et al., 2012), we decided prior to any analyses to exclude participants who failed to achieve mean accuracy of 65% from all analyses. This excluded 0 participants. The average rate of correct response was 93.9% (animals, 93%; insects, 94%). By-participant accuracies are shown in Figure 4.

Word production accuracy in phase (3). A research assistant blind to the manipulation conditions listened to the word productions obtained in phase (3) and transcribed them. The mean accuracy was about 85%, with all participants performing better than the 65% cut-off. This suggests that the task was feasible and the lexicon was acquired reasonably well before participants performed the sentence production task during the test phase. No participant used *-ka* in this phase.

Sentence production accuracy during the test phase (5). The same research assistant listened to the productions obtained in phase (5),

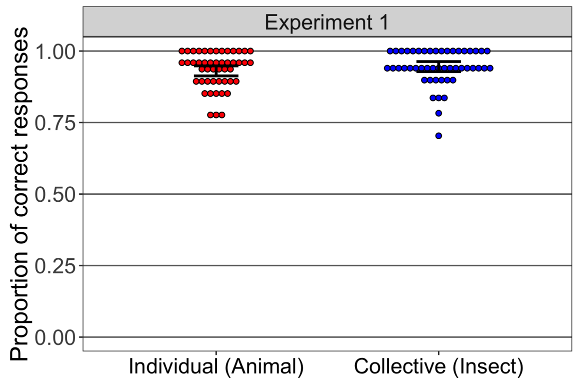

Figure 4: Proportions of correct responses in the 4AFC comprehension trials in Experiment 1. Dots present by-participant averages. Error-bars show 95% Confidence Intervals of the mean.

and annotated if participants produced a given noun correctly and if a noun was produced with *–ka* or not. Participants' responses were scored as "correct" if it matched the provided input, while subtle phonological variations (e.g., *velmick* pronounced as *belmick*) were ignored. Most of the "wrong" nouns were a silence (e.g., "....-ka glim") or a filler (e.g., "something -ka glim"), where the participant seemed to have failed to recall a word. Productions with incorrectly produced nouns were excluded from analysis (removing 116 cases, 13%). In the rare cases of restarting (e.g., "Velm... Oh, Peza-ka glim"), we accepted the second, and more fully realized, production.

Six participants (12.5%) failed to correctly produce 50% of the sentences. These participants were excluded from any further analysis, so as to ensure that only productions from participants who had successfully learned the miniature language inform our conclusions. This left data from 42 subjects and 773 sentences for analysis. All results reported below remain unchanged, if all participants are included in the analysis.

2.2. Results

The predictability hypothesis of markedness predicts that learners will be more likely to produce plural marking for multiple **individuals**, compared to multiple collectives, despite the fact that no such pattern was present in the input during exposure. We first tested this prediction. Then we conducted additional post-hoc analyses that related our findings to research on learners' bias to regularize or condition free variation.

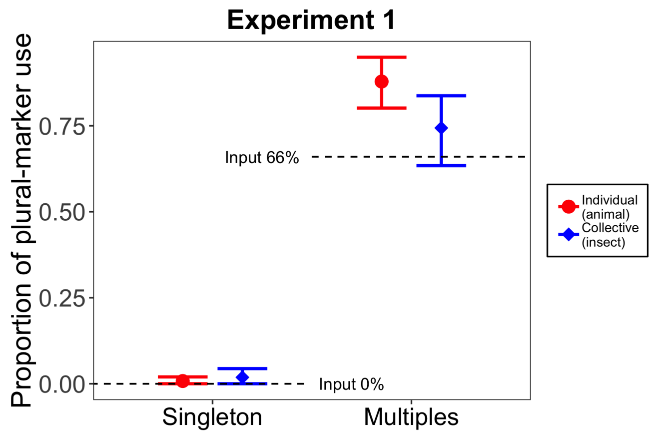

Figure 5: Plural marker use by conditions in Experiment 1. Dots present the mean of by-participant averages. Error-bars show bootstrapped 95% confidence intervals. Dashed black lines indicate the input ratio of the *-ka* marking: 0% for singletons and 66% for multiples.

2.2.1. Proportions of plural marking

Proportions of participants' plural marker use in Experiment 1 are summarized in Table 1 and Figure 5. Overall, participants predominantly used the optional plural marker *-ka* for multiples but not for singletons. This indicates that participants correctly learned that the *–ka* marker signals plurality.

Table 1: Mean and Standard Deviation of proportions of plural marker use by conditions in Experiment 1.

	Singleton	**Multiple**
individual (Animal)	.007 (.035)	.878 (.25)
collective (Insect)	.018 (.069)	.743 (.34)

To examine the effect of noun classes, we applied a mixed-effects logistic regression (Jaeger, 2008) using the `lme4` package in `R` (Bates et al., 2015), predicting the use of *-ka*. We included noun class (sum-coded: individuals = 1 vs. collectives = -1), visual prompt (sum-coded: multiples = 1 vs. singletons = -1), and their interaction as fixed effects and by-participant and by-item random effects. We employed the maximally converging random effect structure justified by the design (Barr et al., 2013). The model reported here includes by-subject and by-item intercepts.[2]

The result summary is provided in Table 2. There was an expected significant main effect of visual prompt such that participants were more likely to produce the optional plural marker *-ka* for multiples than for singletons ($\hat{\beta} = 3.6.$, $z = -2.09.$, $p < .0001$). Critically, the interaction between the noun class and the visual prompt was also significant ($\hat{\beta} = .52$, $z = 2.19$, $p < .03$): learners were significantly less likely to use *-ka* for collectives when they appeared as multiples. They did so despite the fact that they were exposed to three times as many instances of *-ka* with the collectives (insects) compared to the individuals (animals) in the input.

Table 2: Summary of effects in the mixed-effects logistic regression for Experiment 1

Fixed effects	Estimate	Std. Error	z value	p-value
(intercept)	-1.442	.346	-4.172	$3.02e-05$
Noun class (individual)	.24	.238	.716	n.s.
Visual prompt (Multiples)	3.6	.298	-2.09	$.2e-16$***
Noun class × Visual prompt	.518	.237	2.186	$< .029$*
Number of observations 773, Participants 42, Items 12				

Figure 5 shows that participants overall produced more plural marking than in the input. This suggests that participants to some extent regularized plural marking, similar to what has been observed in other miniature language learning studies (e.g., Hudson Kam & Newport, 2005; Hudson Kam & Newport, 2009). The confidence intervals in Figure 5 further suggest that this tendency was stronger for individuals compared to collectives.

[2]For the purpose of this analysis, we defined items in terms of the phonological forms of words (nouns), rather than visual referents. The same results were obtained when the analysis instead included random intercepts by visual referent. (Models with random intercepts by both nouns and visual referents did not converge.) The same holds for Experiment 2.

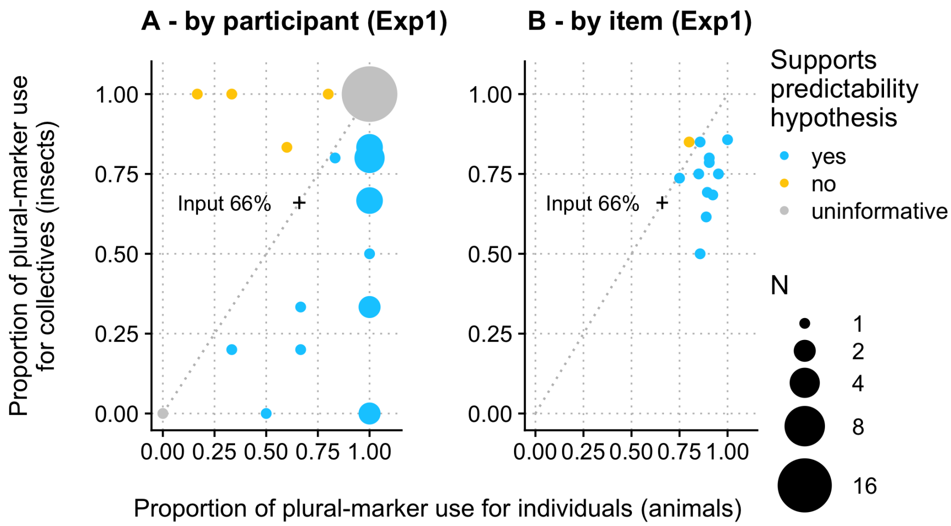

Figure 6: Plural marker use by participant (Panel A) and by item (Panel B) in Experiment 1. Points below the diagonal dotted line follow the predictability hypothesis of markedness, with a tendency to use *–ka* more often when plural meaning is less predictable.

2.2.2. Regularization and conditioning of free variation

To better understand the patterns regularization we observed in the main analysis, we explored the extent to which plural marking differed across participants and items. Figure 6 shows participants' production of the plural marker *-ka* in response to multiple referents. Panel A and Panel B illustrate distributions of the same data by participants and items, respectively. Here we make four observations.

The first observation is that there were considerable individual differences between participants (Figure 6, Panel A), in line with previous work on miniature language learning). On the other hand, there was comparatively little variation across items (Figure 6, Panel B). This observation was corroborated by the variances of the by-participant and by-items random effects in our main analysis reported above. The variance of the by-participants intercepts ($\hat{\sigma}^2 = 2.52$) was two orders of magnitude larger than the variance of the by-item intercepts ($\hat{\sigma}^2 = .04$). [3] This suggests that learners did not

[3]Recall that we defined items in terms of nouns for the purpose of the main analysis. To explore if participants conditioned their use of *-ka* on visual representation of referents rather than nouns, we repeated the main analysis but replaced the random by-noun intercepts with random intercept by visual referent. The variance of the random intercepts by

condition plural-marking on nouns. Rather, the relation between meaning predictability and plural marking seems to be more or less constant across nouns. All but one out of the 12 items follow the pattern expected under the predictability hypothesis.

Second, we observed a strong tendency for participants to regularize their plural marker use. Seventeen out of the 42 participants always produced *–ka* when referring to multiples, both for animals and the insects (see the large gray dot the top-right of Panel A). One participant never used *–ka* (the gray dot on the bottom left corner). This replicates previous findings that learners are generally biased to regularize or condition free variation in the input (e.g., Hudson Kam & Newport, 2009; Kirby et al., 2008; Smith & Wonnacott, 2010). The productions of these participants were uninformative in our regression models with regard to the predictability hypothesis: for participants who regularized the plural marker to 0 or 100%, it was impossible to tell whether they conditioned plural marking on plural predictability.

The third observation is that all but four of the remaining 24 participants (83%) exhibited plural marking in line with the predictability hypothesis of markedness. Fourth and finally, we see that even among these participants there was a strong tendency to condition variation. Eighteen of the 24 participants used *-ka* 100% for either individuals or collectives. Of these, 15 participants (83%) used *-ka* consistently for individuals but only probabilistically for collectives. This means that even among participants who conditionally regularized plural marking, the clear majority behaved in a manner expected under the predictability hypothesis of markedness.

2.3. Discussion

The results of Experiment 1 suggest that learners of a new, miniature, language preferred to produce nouns with plural marking when the intended plural meaning was *less* predictable given the noun (e.g., individuals as opposed to collectives).

In line with previous work, we also found a strong tendency towards regularization. Specifically, most participants produced plural marking *more* at a higher rate than what was demonstrated in the input and made it almost obligatory for individuals when they appeared as multiples. Two factors likely contributed to this tendency to over- rather than under-generalizing plural marking. First, the presence of a plural marker for plural meaning

---

visual representations ($\hat{\sigma}^2 = .14$) was one order of magnitude smaller than the variance of the by-participant intercepts. The same holds for Experiment 2.

was the dominant pattern in the input (i.e., 66% of multiples were accompanied by *-ka*). Second, plural marking is obligatory in participants' native language (English), and it is well-documented that transfer from the native language affect language learning (for review, see Pajak et al., 2016, pp. 920-928), including cases of miniature language learning (for review, see Fedzechkina et al., 2016b, pp. 222-224). With both of these factors present, related previous work has also found over-generalization (e.g., Hudson Kam & Newport, 2005; Hudson Kam & Newport, 2009).

This raises questions about effects of other types of transfer on the results of Experiment 1. It is possible that participants also transferred semantic knowledge about the visual referents from their native language, which could explain our results. Recall that the visual referents in Experiment 1 were intentionally designed to resemble existing animals and insects. This was done so as to facilitate transfer of implicit knowledge about the relative plural predictability: references to insects are more likely to occur with plural meaning, compared to references to large animals (Grimm, 2018; Haspelmath & Karjus, 2017). If participants drew on implicit knowledge of this type (e.g., "this looks like an ant, and references to ants tend to have plural meaning"), this would explain the results of Experiment 1 [4]. We thus conducted a second experiment to see whether we can replicate the results of Experiment 1 while minimizing the transfer of semantic knowledge from the native language.

3. Experiment 2

Experiment 2 examined whether the predictability of plural meaning *as learned during exposure* can cause the type of asymmetrical markedness pattern observed in Experiment 1. To this end, we used the same nouns from Experiment 1 but associated them with novel geometrical shapes, rather than animals and insects. If participants could learn the relative predictability of plural meanings during exposure, the predictability hypothesis of markedness would predict that participants should again be more likely to produce plural marking for individuals, compared to collectives. If we observe the same tendency as in Experiment 1, but to a lesser extent, this would suggest that both transfer of semantic knowledge *and* the learned

[4] It is also possible that the relative size differences between individuals and collectives and other visual features facilitated the transfer of learners' prior knowledge about relative meaning predictability. To address this possibility, we have conducted a follow-up experiment, results of which are reported in Appendix I

Figure 7: Examples of the visual referents presented during exposure in Experiment 2

probability of plural meaning during exposure contributed to the results of Experiment 1. Finally, if Experiment 2 does not replicate the results of Experiment 1, this would suggest that the effects of Experiment 1 were solely due to transfer of semantic knowledge from participants' native language.

3.1. Method

3.1.1. Participants

A new group of 52 native speakers of American English at University of Rochester participated in this study. They received \$10 for their participation. As in Experiment 1, they were all native speakers of American English without an advanced command of any other language. Using the same exclusion criteria as in Experiment 1, 10 subjects were excluded from analysis.

3.1.2. Materials

All materials and the grammar of the miniature language were identical to those in Experiment 1, except that the visual referents consisted of 12 geometric shapes with no commonly known names (Figure 7). To equate the visual features of the referents (e.g., size, spacial distributions, complexity of visual scenes) between Experiments 1 and 2, we created two classes of referents. Individuals consisted of six relatively large geometric shapes spatially distributed in a manner similar to how the animals were presented in Experiment 1. Collectives consisted of six smaller shapes. Individuals and collectives were associated with distinct colors (red/orange and blue/green, respectively) to facilitate the classification during learning.

3.1.3. Procedure

Same as Experiment 1.

3.1.4. Scoring and exclusions

Comprehension accuracy in phase (2). The mean accuracy in the 4AFC task was 86% (individuals, 89%; collectives, 83%). This was some-

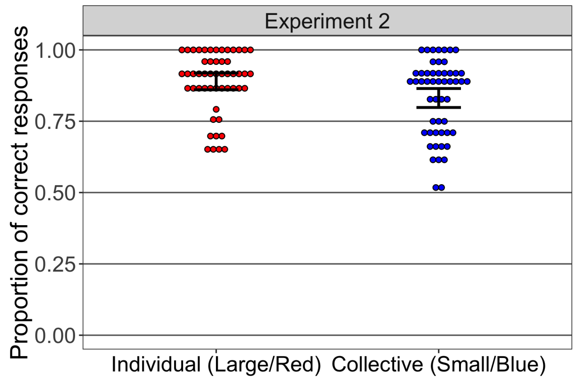

Figure 8: Proportions of correct responses in the 4AFC comprehension trials in Experiment 2. Dots present by-participant averages. Error-bars show 95% Confidence Intervals of the mean.

what lower than in Experiment 1 (93.9%), suggesting that the word learning was more difficult in Experiment 2 (Figure 8) compared to Experiment 1 (Figure 4). This was presumably due to the overall unfamiliarity with the geometric shapes as opposed to animals and insects that were visually more distinct and easier to memorize. One subject could not achieve the cut off rate of 65% and was removed from the analysis.

Word production accuracy in phase (3). The mean accuracy was about 80%, with all participants performing better than the 65% cut-off. As in Experiment 1, No participant used *-ka* in this phase.

Sentence production accuracy during the test phase (5). We excluded 9 (17.6%) of the remaining participants who failed to produce 50% of the sentences in the final sentence production phase. We then further removed 151 (15.5%) sentences that included wrong or no nouns. This left data from 42 subjects and 759 sentences for analysis. All results reported below remain unchanged, when all participants were included in the analysis.

3.2. Results

We again first analyzed whether, in response to multiple referents, learners were more likely to produce plural marking for individuals than for collectives, as expected under the predictability hypothesis. Then we compared the results of Experiment 2 to those of Experiment 1. Finally, we examined the patterns of regularization of *-ka* use in Experiment 2 by participant and item, as we did for Experiment 1.

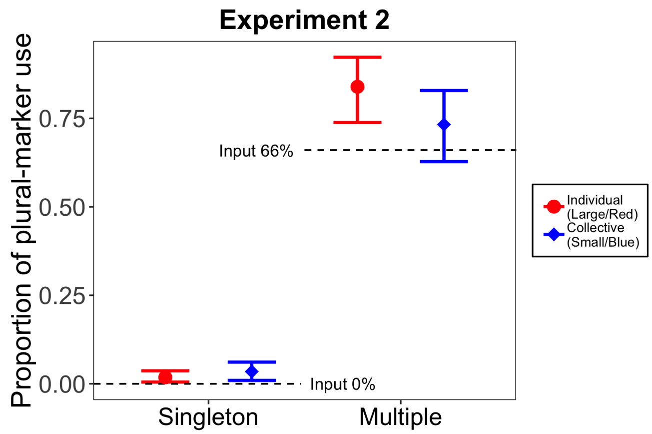

Figure 9: Proportions of plural marker use by conditions in Experiment 2. Dots present by-participant averages. Error-bars show bootstrapped 95% Confidence Intervals. Black dashed lines indicate the input ratio of the *-ka* marking.

3.2.1. Proportion of plural marking

Proportions of participants' plural marker use in Experiment 2 are summarized in Table 3 and Figure 9. Ten out of the 42 participants never used the optional plural marker *-ka* while none of them used it in all the trials.

Table 3: Mean and Standard Deviation of proportions of plural marker use by conditions in Experiment 2.

	Singleton	**Multiple**
individual (Animal)	.018 (.057)	.839 (.32)
collective (Insect)	.034 (.089)	.732 (.34)

We conducted the same mixed-effect logistic regression as in Experiment 1 (with the same coding and random effects). A model summary is provided in Table 4. As in Experiment 1, we found a significant main effect of visual prompts, with more *-ka* use for multiples ($\hat{\beta} = 3.47$, $z = 11.71$, $p < .0001$), and an interaction between the noun class and the visual prompts ($\hat{\beta} = .47$, $z = 2.44$, $p < .02$). This replicates the inverse relation between the proportion of *–ka* use and the predictability of plural meaning observed in Experiment 1.

3.2.2. Comparison to Experiment 1

In order to test whether Experiment 2 fully replicates Experiment 1, we combined the data from Experiments 1 and 2. We then analyzed the

Table 4: Summary of effects in the mixed-effects logistic regression for Experiment 2

Fixed effects	Estimate	Std. Error	z value	p-value
(intercept)	-1.44	.38	-3.75	$< .0001$
Noun class (individual)	.123	.193	.637	n.s.
Visual prompt (Multiples)	3.47	.297	11.713	$.2e-16$***
Noun class $\times$ Visual prompt	.47	.192	2.443	$< .015$*
Number of observations 759, Participants 42, Items 12				

combined data as a function of noun class (sum-coded as before), visual prompt (sum-coded as before), experiment (sum-coded: Experiment 2 = 1 vs. Experiment 1 = -1), and all of their interactions. We again included by-participant and by-item intercepts as random effects.

A model summary is provided in Table 5. This analysis replicates the main effect of visual prompts ($\hat{\beta} = 3.5$, $z = 17.02.$, $p < .0001$), and the interaction between the noun class and visual prompts ($\hat{\beta} = .5.$, $z = 3.24.$, $p < .002$). Crucially, there was no significant effect of the experiments, including no interactions of experiment with noun class or visual prompts (all ps$> .8$). This suggests that the effect of plural predictability in Experiments 1 and 2 is not limited to transfer of semantic knowledge from participants' native language.

3.2.3. Regularization and conditioning of free variation

The patterns we saw in Experiment 1 with regard to the conditioning of variation were observed in Experiment 2 as well. As shown in Figure 9, participants seemed to over-generalize plural marking (presumably for the same reasons as in Experiment 1). As shown in Figure 10, this again resulted in a large number of participants who always used plural marking for references to multiples (19 out of 42).

Critically, the majority of participants who did not regularize followed the asymmetric use patterns of *-ka* expected under the predictability hypothesis (16 out of 20, 80%, compared to 83% in Experiment 1). Also following Experiment 1, this was in large parts driven by conditional regularization: 12 out of the 16 participants (75%) always used plural marking

Table 5: Summary of the fixed effects in generalized mixed effect model for Experiments 1 and 2

Fixed effects	Estimate	Std. Error	z value	p-value
(intercept)	-1.455	.258	-5.648	1.62e-08
Noun class (individual)	.14	.154	.907	n.s.
Visual prompt (Multiples)	3.514	.206	17.018	$< 2e - 16$ ***
Experiment (Experiment 2)	.026	.245	.107	n.s.
Noun class × Visual prompt	.499	.154	3.247	.002 **
Noun class × Experiment	-.02	.154	-.128	n.s.
Visual prompt × Experiment	-.171	.177	-.966	n.s.
Noun class × Visual prompt × Experiment	-.056	.153	-.368	n.s.
Number of observations 1532, Participants 84, Items 12				

for individuals, but did not always use plural marking for collectives. Finally, there was little variability across items: plural marking on all 12 nouns followed the pattern expected based on the predictability hypothesis. This was again corroborated by the much larger variance of the random intercepts by participants than by items in mixed-effects analyses reported in Table 2 ($\hat{\sigma}^2_{participants} = 3.84$ vs. $\hat{\sigma}^2_{items} = .15$).

3.3. Discussion

The results of Experiment 2 largely patterned with those of Experiment 1, except that the word learning seemed to be slightly more challenging with the less nameable geometrical shapes. The results suggest that the asymmetrical pattern observed between the two noun classes cannot be solely attributable to participants' prior knowledge about existing animals and in-

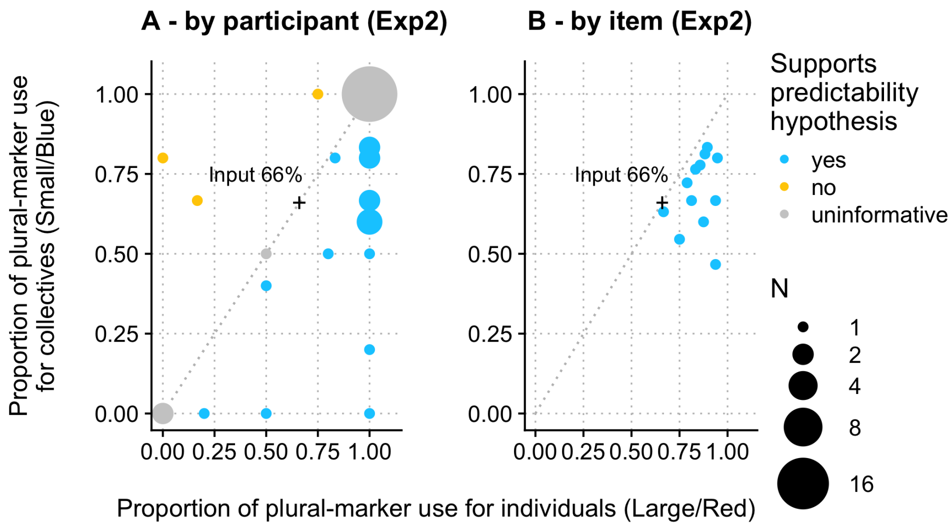

Figure 10: Proportions of plural marker use by participant (Panel A) and by-item (Panel B) in response to visual prompt with multiple referents in Experiment 2.

sects and/or linguistic encoding of these entities. We take this as evidence that participants can learn predictabilities of plural meaning for two classes of items and prioritize their use of the optional plural marker for those for which plural meaning is less predictable.

4. General Discussion

A central puzzle for language scientists over decades is the origin of the tendency for the world's languages to overtly mark some meanings (i.e., plural, past) while leaving others (e.g., singular, present) unmarked. This tendency has been hypothesized to relate to the predictability of meaning: Frequently expressed meanings are more predictable and can therefore be expressed by shorter forms. However, it has been difficult to empirically address this causal explanation using natural language data because predictability of meaning cannot readily be estimated independent of token frequency of associated forms. To overcome this difficulty, in the current study, we adopted a miniature language learning paradigm to study adult language learners' usage of an optional plural marker (-*ka*) based on the inconsistent (probabilistic) input. We controlled predictabilities with which two classes of referents appear as singletons vs. multiples, as proxies of their plural meaning predictabilities. We predicted that learners would be more likely to produce the optional plural marker for nouns whose referents were overall more likely to be singletons.

The results of two experiments provided support for this hypothesis: Native speakers of American English were more likely to produce nouns *without* plural marking when the plural meaning is more predictable given the noun class (e.g., animals vs. insects, Experiment 1). The effect was present even with nonce noun classes, when their within-experiment statistics, as well as visual features of referents (size, spacial arrangements, and movement patterns) support differential plural predictability (Experiment 2). This asymmetrical uses (and non-uses) of *–ka* cannot readily be accounted for by language learners' tendencies to veridically match their production to the statistics in the input (i.e., "probability matching" Hudson Kam & Newport (2005)). If participants had been simply tracking joint statistics of nouns and the optional marker (66% both for individuals and collectives) and matching their output to the statistics, the *-ka* marking would have been equally likely for individuals and collectives. However, our data suggest that this was not the case.

Likewise, it is unlikely that participants simply registered the absolute token frequencies of *–ka* use for the two noun classes and mirrored the patterns in their production. If that had been the case, participants should have produced *–ka* more often for collectives than for individuals because more instances of *-ka* were observed for collectives (30 times) compared to individuals (10 times). In fact, the results showed the opposite trend. To capture the existence and the direction of the asymmetry of *-ka* marking, we need an account that can go beyond faithfully representing patterns of linguistic forms observed in the input.

Critically, English does not have the optional plural marking system. Still, when native speakers of English are exposed to an optional number marking language with no bias to mark plurality for low-predictability items, they end up producing more plural marking for less predictable items. This constitutes novel evidence that predictability of meaning can cause markedness patterns resembling those observed cross-linguistically.[5] In the remain-

[5]No data as of yet demonstrates a historical link whereby an optional plural marking system develops into an obligatory plural marking system. There is, however, evidence of exactly such a shift for other types of optional grammatical marking (e.g., the shift to differential object marking in Spanish, see Von Heusinger & Kaiser 2005). Here we proceed under a common assumption that functional pressures have the potential to over generations shape optional systems into an obligatory system (for related discuss, see Fedzechkina et al., 2016b). Future studies within the iterative language learning paradigm

der of this discussion, we consider how predictability of grammatical meaning can lead to less linguistic encoding.

4.1. Why does the predictability of plural meaning affect the production of optional plural marking?

The results of Experiments 1 and 2 add to recent work on the causal link between frequency and the reduction of lexical, rather than grammatical, forms. Sóskuthy & Hay (2017), for example, investigate how changes in the frequency of word usage affect changes in word durations over 130 years of recorded New Zealand English (from 1851 to 1987). Consistent with the hypothesis that predictability can lead to form reduction, they found that increases in a word's usage frequency from one historical time to another tend to correlate with shortening of the average word duration. A concurrently published study yields a converging finding within an experimental, miniature-language learning paradigm. Kanwal et al. (2017) examined (among other things) the effect of frequency on the choice between a shorter (e.g., *zop*) and a longer word form (*zopudon*). They employed a communicative game in which participants had to communicate to a partner which of two objects they saw. One of the objects was presented much more frequently than the other. Importantly, both objects could be referred to with either a short and a long word form. The two short forms were identical (both *zop*), whereas the two longer forms differed (*zopudon* and *zopekil*). Kanwal and colleagues found that participants preferred to use the shorter label for the more frequent object.

Though different in important details, both the historical corpus study by Sóskuthy and Hay and the communicative game by Kanwal and colleagues provide strong evidence for a causal role of usage frequency and the shortening of lexical forms. This critically extends typological or psycholinguistic work that had established synchronic correlations between word frequency and word length (e.g., Bybee, 2003; Fenk-Oczlon, 2001; Ferrer i Cancho & Solé, 2003; Piantadosi et al., 2011; Zipf, 1949). The present study, for the first time, further expands the scope of such causal evidence from frequency effects in the lexicon to include effects of meaning predictability on grammatical encoding.

Where, then, does the pressure to associate frequently expressed meaning and relatively, or completely, reduced forms come from? A key to answering this question, we argue, can be found in the theory of Principle of Least

(Kirby et al., 2008) are required to test this assumption.

Effort (Zipf, 1949). It postulates that human behaviors — including linguistic communication—often converge on a form that minimizes the expected overall work expenditure to achieve a given goal.[6] In language use, if there are more than one way to encode the same meaning, language users are likely to adopt the one that overall incurs the least amount of production efforts. This insight has been incorporated into more recent, information-theoretic, approaches to language production and comprehension as an inference problem under uncertainty (Aylett & Turk, 2004; Frank & Jaeger, 2008; Genzel & Charniak, 2002; Gibson et al., 2013; Jaeger, 2010; Levy & Jaeger, 2007; Jaeger, 2013). These approaches posit that a guiding principle of language production is to send the signal that allows a listener to inferentially recover an intended meaning. The signal is, however, bound to be degraded due to multiple noise sources (e.g., articulatory and perceptual disturbances, ambient noise). Assigning more signal (form) to less frequent/predictable meaning is therefore considered as a key to "efficient" information transmission because it reduces the efforts for the speaker while increasing the chance of successful inference in the listener.

The bias to communicate an intended meaning with least effort can correctly predict the learners' conditional use of the optional plural marking (i.e., extra expenditure of effort) on the less predictable grammatical category found in the current study. (Similar findings were reported with respect to phonetic encoding of lexical items (Buz et al., 2016), case marker use in artificial language (Fedzechkina et al., 2012, 2016a, 2017) and in constituent ordering (Lockridge & Brennan, 2002).) Such results are not readily be accounted for by other influential theories of optionality in language production. Here, we consider two most prominent accounts that could be applied to explain our data.

The current results are not predictable based on the idea that optionality is employed primarily to maximize the talker's production ease (Arnold et al., 2012; Ferreira & Dell, 2000; MacDonald, 2013). For instance, Fer-

[6] *Communication* here is defined in the information-theoretic sense: bits of information being sent from a transmitter to a recipient, which can be humans or computers, to get a meaning across through a noisy channel. It therefore includes, but is not limited to, multi-party human interaction as often construed in a general sense of the word. Previous studies have examined efficiency principles of communication in laboratory experiments (Fedzechkina et al., 2012, 2016a; Kurumada & Jaeger, 2015) as well as in a variety of interactive settings e.g., Experimental study with a simulated interlocutor: Buz et al. (2016); Corpus-based studies using naturalistic, conversational, data: Levy & Jaeger (2007); Frank & Jaeger (2008); Jaeger (2010, 2013); Pate & Goldwater (2015)).

reira & Firato (2002) argued that speakers of American English are more likely to produce an omissible complementizer *that* when an upcoming noun (the subject of the complement clause) is difficult to retrieve. This is considered to facilitate fluency because production of the optional *that* buys speakers time for the planning of upcoming material. In the current miniature artificial language, however, there was only one type of verb following the grammatical subject, keeping constant the availability of the material following the noun (none of the participants failed to learn the verb).

Our results are also unexpected given certain automatization accounts of frequency and predictability effects. Some researchers have attributed reduction of forms to routinization (Bybee, 2007; Haspelmath, 2006, 2008), wherein speakers shorten or omit linguistic elements that they have ample experiences of producing. Indeed, in the sentence training phase of our experiment, participants had more opportunities to repeat plural-marked collective nouns (12 times), as compared to individual nouns (4 times). The current data are therefore compatible with an explanation that frequent production of a linguistic form predicts its reduction in subsequent uses. The routinization-based account, however, is usually given to historical changes of linguistic forms over generations of users (Bybee & Hopper, 2001; Bybee, 2003). It is rather inconceivable that the same type of changes arose within 12 instances of production in the current task.

This last line of reasoning, although unlikely to account for the current data, raises an important question: What constitutes frequency (predictability) of meaning, as relevant to grammatical encoding, independent of frequency of form use? We briefly discuss this issue before turning to a topic of future research.

4.2. What is meaning predictability?

How do language users estimate predictability of meaning in real languages? As discussed in the introduction, predictability of meaning is generally highly correlated with predictability of associated form. Previous studies on markedness in number marking, therefore, extrapolated predictability of singular vs. plural meaning from textual frequencies of singular vs. plural forms (e.g., "dog" vs. "dogs") (Haspelmath & Karjus, 2017). In the current study, we instead defined the notion as the average conditional probability of singular vs. plural meaning in the context of a particular noun class (e.g., How likely is singular vs. plural meaning expressed when the referent is an animal vs. insect?). A significant question remains, however, as to what constitutes predictability of number meaning and why a certain category (e.g., singular vs. plural meaning) is expressed more often for a given noun class than for others.

It has been convincingly argued that predictability of singular vs. plural meaning is not directly determined by frequency of relevant entities and events in the world *per se* (Haspelmath, 2006). Certainly, it is highly likely that objective numbers of, for example, elephants vs. ants that exist in the world contribute to how often language users talk about them as singletons or multiples. However, as Haspelmath (2006, p.45) points out, objective counts of entities do not straightforwardly predict how often they are linguistically mentioned. For instance, textual frequency of the word *dog* is much larger than that of *beetle*, even though the world has many more beetles than dogs. The more important metric, as relevant to grammatical number marking, likely pertains to whether, and how often, speakers conceptualize a particular referent as individuated (and as more compatible with singular meaning) or a collective (and as more compatible with plural meaning) when they communicate *linguistically*.

Further examination of the nature and roles of meaning predictability paves a way for a deeper understanding of perceptual and conceptual basis of grammatical number (Wierzbicka, 1988; Prasada et al., 2002; Middleton et al., 2004; Grimm, 2018). It has been observed that, in languages that allow markedness reversals (Section 1.1), the types of noun stems that have plural reference by default and require additional morphology to designate singular meaning correspond to entities that typically co-occur, either perceptually, such as seeds or swarming insects, or functionally, such as stairs or shoes (Tiersma, 1982; Arensen, 1998; Stolz, 2001; Grimm, 2012a). An intriguing hypothesis that emerges in light of the current study is that the perceptual and functional properties of these entities shape grammatical number marking through relative predictability of plural meaning. Plural meaning is made highly predictable in the context of nouns that refer to canonically co-occurring entities, which may call for extra linguistic encoding (number marking morphology) to signal the relatively unexpected meaning (e.g., singular).

Using an artificial language learning paradigm that we used in the current study, we can directly test this hypothesis by manipulating non-linguistic features of novel entities, in particular, likelihoods with which they appear and function in pairs or groups. For example, two physically independent objects can be introduced as functionally independent or coordinated (e.g., two sticks used individually like skewers or used together like chopsticks). And in the linguistic input they are optionally number-marked at a constant rate, just as in the current experiments. We predict that learners of this language would preferentially plural-mark the class of referents that are *not* perceptually or functionally grouped because the meaning is less predictable given their identities. Thus our paradigm provides a productive testing ground for a perennial question about whether, and if so how, factors stemming from language use – frequency, predictability, and humans' canonical patterns of interaction with entities in the world – can play a guiding role in shaping grammatical systems in world's languages.

4.3. Limitations and topics for future research

The present study adopted a paradigm that is far simpler and constrained than natural languages and everyday language use. Perhaps most notably, our paradigm did not involve an conversation partner.[7] If the predictability effects we observed are indeed mediated through consideration of successful information transmission, the effects should be even larger in more interactive communication. Recent advances have demonstrated that it is possible to incorporate interactive communication with real or simulated interlocutors into miniature language learning paradigms (e.g., Kirby et al., 2014; Smith et al., 2017). Working within such a paradigm, Kanwal et al. (2017)'s finding summarized above emphasized the importance of communicative context. In their study, the inverse relation between meaning frequency and shorter word forms manifested itself primarily when language users were under pressures for communicating information to interlocutors both accurately and efficiently. This makes interactive miniature language learning paradigms a promising venue for future research on the mechanisms underlying predictability effects like those observed here. Additionally, interactive paradigms allow researchers to evaluate how population-level, cultural, factors affect the acquisition of optional grammatical devices (Mesoudi & Whiten, 2008; Kirby et al., 2007; Smith et al., 2017; Samara et al., 2017).

In conclusion, the current results constitute an important step towards understanding the critical role of language use as a driving force of form-meaning mappings in human language. The miniature language paradigm enabled us to test long-standing proposals that are found to be challenging to evaluate on typological data. Further investigation in this domain will shed light on the role of distributional information of linguistic forms and

[7]But see Clark & Carlson (2004) on the notion of over-hearers. As participants were aware that they were being recorded, there is an implicit interlocutor and over-hearer in our experiments. In that sense, participants were engaging in communicative acts (see also discussion in Kurumada & Jaeger, 2015, p.169). Still, there is no question that our paradigm presents a constrained form of communication.

meanings, shaping our behaviors in linguistic communication and learning as well as typological variations across languages.

5. Acknowledgments

Thanks to Jennifer Andrews, Sahed Martinez, Rachel Myers, Wesley Orth and Joseph Plvan-Franke for assistance with stimuli construction, data collection, and annotation. We would also like to acknowledge T. Florian Jaeger, Masha Fedzechkina, Martin Haspelmath, the Kurumada-Tanenhaus Lab, and the Experimental Semantics/Pragmatics group at the University of Rochester for valuable discussions. Earlier versions of this paper were presented at the Cognitive Science Society meeting in July 2017 as well as at the Meaning in Flux Workshop at Yale University in October 2017.

6. Appendix I. A follow-up Experiment (Experiment 3) on the effect of visual representations

6.1. Rationale

We conducted a follow-up experiment to examine possible effects of the visualization of the noun classes (individual vs. collective). In this experiment, we replicated Experiment 1 except that we reversed the associations between the noun classes (animals vs. insects) and visual features of the referents (e.g., size, spacial distributions, movement patterns). That is, in Experiment 3, animals were depicted smaller (the same size as the insects in Experiment 1) and moved in a group when they appeared as multiples. On the other hand, insects were depicted larger and moved independently. The statistics of singletons vs. multiples, however, remained identical to Experiment 1: animals were more likely to appear as singletons whereas insects were more likely to appear as multiples.

The goal of Experiment 3 was to dissociate two factors that were likely at work in Experiment 1, and thereby further evaluating the validity of the predictability hypothesis. Specifically, we addressed if the effect observed in Experiment 1 was solely dependent on the visual features of the referents. Table 6 summarizes the factors manipulated across the experiments. Visual features refer to the relative sizes of referents, their spatial distributions, and movement patterns. In Experiments 1 and 2, an expected effect of visual features of referents pointed in the same direction as the effects of input statistics (i.e., Animals/Large objects occurred as singletons more often than insects/small objects). As discussed in the Introduction and General discussion, plural marking is likely influenced by the conceptual associations

Table 6: An overview of the expected effects of the manipulations in Experiments 1-3. Input statistics (the left most column) refer to the relative proportions of times in which referents were presented as individuals and collectives. Semantic knowledge (the center column) refers to participants' prior linguistic knowledge as to how often (existing) animals and insects are described as individuals and collectives. Visual features (the right most column) refer to the visualizations of the stimuli in the current experiments. Values in each cell indicates a bias a given factor yields with respect to plural marking. For instance, we expected based on our meaning predictability hypothesis that input statistics endorsed plural marking for animals, as opposed to insects, in Exp.1.

	Input statistics	Prior semantic knowledge regarding animals vs. insects	Visual features (size/movement) of referents
Exp.1	Animals	Animals	Animals
Exp.2	Large objects	N/A	Large objects
Exp.3	Animals	Animals	Insects

between visual features of referents and how they are often talked about, e.g., larger objects are often individuated and frequently talked about as singletons whereas smaller objects are often aggregated and frequently talked about as multiples. (Note that this is correlated, but meaningfully separable from, the potential effect of semantic knowledge about animals and insects we attempted to control for in Exp.2). It is logically possible that the effect we observed in Experiments 1 and 2 was driven primarily by the visual features rather than the input statistics.

There were three possible outcomes for Experiment 3.

1. If the statistics were the sole source of information that guides plural marking, **animals** should receive more plural marking when they appear as multiples.
2. If the visual features were the sole source of information that guides plural marking, **insects** should receive more plural marking when they appear as multiples.
3. If both statistics and visual features affect the probability of speakers use of plural marking (i.e., if these effects somehow cancel each other out), **animals and insects should receive plural marking at a more or less constant rate**.

6.2. Participants

A new group of 47 native speakers of American English at University of Rochester participated in this study. They received $10 for their participation. As in Experiments 1 and 2, they were all native speakers of American

English without an advanced command of any other language. We used the same exclusion criteria as in Experiments 1 and 2.

6.3. Stimuli

The stimuli were identical to those in Experiment 1 except that the visual features of the noun classes (e.g., size, spatial distributions, complexity of visual scenes) were reversed. Animals were shrunk to the size of the insects in Experiment 1, and appeared as a group of ten when presented as multiples. In trials where the referents were to be described with a verb, they showed a group motion (i.e., all the instances moved up and down synchronously). Insects, on the other hand, were enlared to the size of the animals in Experiment 1, and appeared as a pair when presented as multiples. They showed a staggered motion, as individuals did in Experiments 1 and 2.

We created a new set of still pictures for Phase 1-3 and a new set of video clips for Phases 5-6. The audio stimuli, on the other hand, remained identical to those in Experiments 1 and 2.

6.4. The language

The lexicon and its occurrence statistics were identical to those in Experiment 1. That is, animals were more likely to appear as singletons whereas insects are more likely to appear as multiples (i.e., in pairs). Nouns used for these referents were probabilistically marked with *-ka* 2/3 of the time regardless of the class affiliations.

6.5. Procedure

6.5.1. Scoring and exclusions

Comprehension accuracy in phase (2). The mean accuracy in the 4AFC task was 90.5% (individuals, 87%; collectives, 94%). This was somewhat lower than in Experiment 1 (93.9%), suggesting that the word learning was more difficult in Experiment 3, especially for the animal referents. Four subjects could not achieve the cut off rate of 65% and were removed from the subsequent analyses.

Word production accuracy in phase (3). The mean accuracy was about 82%, with all participants performing better than the 65% cut-off. As in Experiment 1, no participant used *-ka* in this phase.

Sentence production accuracy during the test phase (5). We removed 92 (9.8%) sentences that included wrong or no nouns. We then excluded one (2.3%) participant, who failed to produce 50% of the sentences in the final sentence production phase. This left data from 42 subjects and 766 sentences for analysis. All results reported below remain unchanged, when all participants were included in the analysis.

6.6. Results

As in Experiments 1 and 2, we first analyzed whether, in response to multiple referents, learners were more likely to produce plural marking for individuals than for collectives, as expected under the predictability hypothesis. We then examined the patterns of regulartization of *-ka* use in Experiment 3 by participant and item, as we did for Experiments 1 and 2.

6.6.1. Proportion of plural marking

Table 7: Mean and Standard Deviation of proportions of plural marker use by conditions in Experiment 3.

	Singleton	**Multiple**
individual (Animal)	.004 (.025)	.702 (.36)
collective (Insect)	.021 (.087)	.670 (.37)

The result summary is provided in Table 7 and Figure 11. We conducted the same mixed-effect logistic regression as in Experiments 1 and 2 (with the same coding and random effects). A model summary is provided in Table 8. There was an expected significant main effect of visual prompt such that participants were more likely to produce the optional plural marker *-ka* for multiples than for singletons ($\hat{\beta}$ = 3.514., z = 10.285., $p < .0001$). Critically, the interaction between the noun class and the visual prompt was marginally significant ($\hat{\beta}$ = .515, z = 1.79, $p < .08$). This suggests that learners were relatively more likely to use *-ka* for individuals (animals), compared to collectives (insects), when they appeared as multiples. This was predicted by the hypothesis that meaning predictability as supported by the statistics in the input conditioned learners' uses of *-ka*. The effect of the interaction term was, however, weaker than in Experiments 1 and 2, only approaching the significance level of .05.

6.6.2. Regularization and conditioning of free variation

The patterns of regularization according to participants and items are plotted in Figure 12. As compared to Experiments 1 and 2, in which the input statistics of singleton vs. multiple referents and their visual properties combined to support meaning predictability, a wider variation was observed

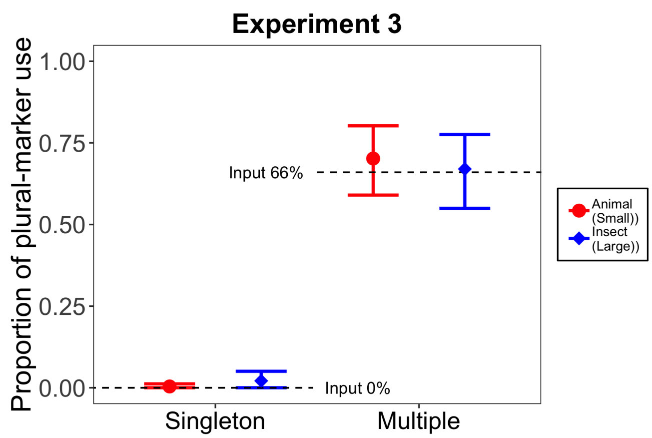

Figure 11: Proportions of plural marker use by conditions in Experiment 3. Dots present by-participant averages (White = singleton visual prompt; Black = multiple visual prompt). Error-bars show 95% Confidence Intervals. Error-bars show 95% Confidence Intervals. Dashed lines indicate the input ratio of the *-ka* marking: 0% for singletons and 66% for multiples.

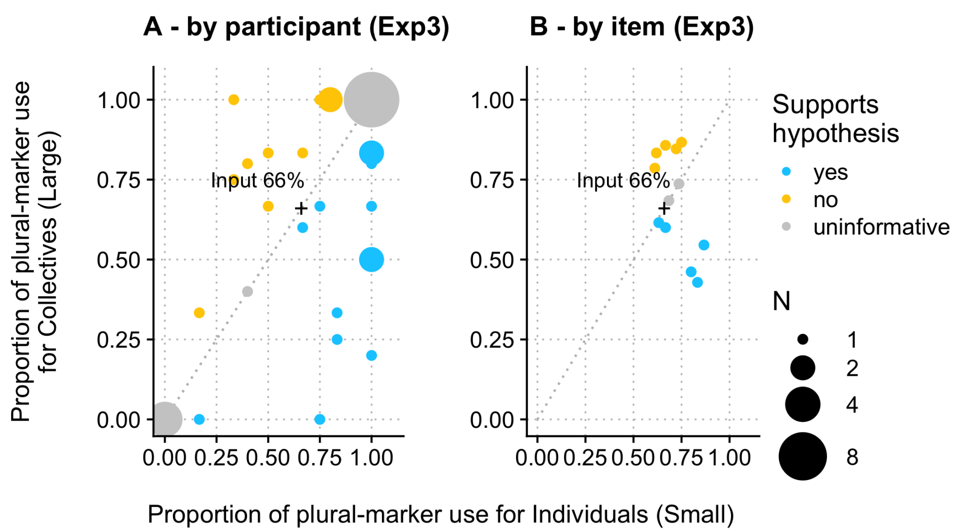

Figure 12: Proportions of plural marker use by participant (Panel A) and by-item (Panel B) in response to visual prompt with multiple referents in Experiment 3.

Table 8: Summary of effects in the mixed-effects logistic regression for Experiment 3

Fixed effects	Estimate	Std. Error	z value	p-value
(intercept)	-2.299	.445	-5.163	$2.43e-07$
Noun class (individual)	-.385	.29	.183	n.s.
Visual prompt (Multiples)	3.514	.342	10.285	$2e-16$***
Noun class × Visual prompt	.515	.288	1.79	$< .073$*
Number of observations 766, Participants 42, Items 12				

for the regularization patterns. As is compatible with the results of the regression analysis reported above, a relatively smaller number of learners and items, compared to those in Experiments 1 and 2, exhibited the pattern straightforwardly expected under the predictability hypothesis. It appears that some learners conditioned their *-ka* use on the input statistics while others relied more on the visual features (e.g., size/motion). Similarly, items seemed to have been treated in a non-uniform manner, suggesting that items were deferentially susceptible to the two conflicting sources of meaning predictability estimates.

6.6.3. Discussion

The central implication we can draw from Experiment 3 is that learners did pay attention to the visual characteristics of the stimuli as they learned and produced the novel plural marker. This, however, does not contradict the main argument of the current study, i.e., use patterns of the optional marker were conditioned on the predictability of the plural meaning represented by the input statistics of singletons and multiples. If that was *not* the case, and if the visual features were the primary source of the meaning predictability, collectives (i.e., insects) would have received plural marking at a higher rate than individuals (per Possibility 2 in Section 6.1). This, however, was not the case in Experiment 3.

In fact, the observed pattern of results accords with the general view we endorsed in the general discussion (Section 4.2): learners are inferring meaning predictability based on information from multiple sources (e.g., statistical and conceptual properties). In Experiment 3, the input statistics and the visual features were supporting the distinct, conflicting, biases (Table 6) and therefore likely canceling each other out. Relative weighting of these different biases can vary across learners and items while exact mechanisms that yielded such divergence cannot be determined based on the current data. Future studies should delineate linguistic, statistical, and conceptual factors that serve as ingredients of meaning predictability. Such examinations will lead to better understanding of how these factors, individually as well as in combination with other factors, support form-meaning mappings in natural language.

7. Data references

Chigusa Kurumada & Scott Grimm, Artificial language learning response data for "Predictability of meaning in grammatical encoding", 2019. The data file can be downloaded from here:
https://osf.io/f26ah/?view_only=54389f2552044c5b8eb3e7b1bd72ed85

8. References

Arensen, J. E. (1998). Murle categorization. In G. J. Dimmendaal, & M. Last (Eds.), *Surmic Languages and Cultures* (pp. 181–218). Köln: Köppe Verlag.

Arnold, J. E., Kahn, J. M., & Pancani, G. C. (2012). Audience design affects acoustic reduction via production facilitation. *Psychonomic Bulletin & Review*, *19*, 505–512.

Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis. *Language and Speech*, *47*, 31–56.

Aylett, M. P., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *The Journal of the Acoustical Society of America*, *119*, 30–48.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*, 92–111.

British National Corpus Consortium (2007). British National Corpus version 3 (BNC XML edition). Distributed by Oxford University Computing Services on behalf of the BNC Consortium (http://www.natcorp.ox.ac.uk/).

Bürki, A., Ernestus, M., & Frauenfelder, U. H. (2010). Is there only one fenêtre in the production lexicon? On-line evidence on the nature of phonological representations of pronunciation variants for French schwa words. *Journal of Memory and Language*, *62*, 421–437.

Buz, E., Tanenhaus, M. K., & Jaeger, T. F. (2016). Dynamically adapted context-specific hyper-articulation: Feedback from interlocutors affects speakers' subsequent pronunciations. *Journal of Memory and Language*, *89*, 68–86.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*, 261–290.

Bybee, J. (2003). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2007). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711–733.

Bybee, J., & Hopper, P. (2001). Introduction. In J. L. Bybee, & P. J. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 229–254). Amsterdam: John Benjamins.

Ferrer i Cancho, R., & Solé, R. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, *100*, 788–791.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English*. Studies in language. MIT Press.

Christiansen, M. H. (2000). Using artificial language learning to study language evolution: Exploring the emergence of word order universals. In L. Dessalle, & L. Ghadakpour (Eds.), *The Evolution of Language: 3rd International Conference* (pp. 45–48). Paris, France: Ecole Nationale Superieure des Telecommunications.

Clark, H. H., & Carlson, T. B. (2004). Hearers and Speech Acts. *Language*, *58*, 332–373.

Culbertson, J., & Adger, D. (2014). Language learners privilege structured meaning over surface frequency. *Proceedings of the National Academy of Sciences*, *111*, 5842 – 5847.

Culbertson, J., & Smolensky, P. (2013). Testing Greenberg's universal 18 using an artificial language learning paradigm. In *Proceedings of the 40th Annual Meeting of the North East Linguistic Society* (pp. 133–146). volume 1.

De Saussure, F. (2011). *Course in general linguistics*. Columbia University Press.

Dryer, M. S., & Haspelmath, M. (Eds.) (2013). *WALS Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Ernestus, M. (2014). *Acoustic reduction and the roles of abstractions and exemplars in speech processing* volume 142.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *PNAS*, *44*, 17897–17902.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2016a). Balancing effort and information transmission during language acquisition : Evidence from word order and case marking. *Cognitive Science*, *41*, 416–446.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2016b). Miniature artificial language learning as a complement to typological data. In *The usage-based study of language learning and multilingualism*. Washington, DC: Georgetown University Press.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2017). Human information processing shapes language change. *Psychological Science*, *eprint https://doi.org/10.1177/0956797617728726*.

Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. In J. L. Bybee, & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 431–448). John Benjamins Publishing.

Ferreira, V. S., & Dell, G. S. (2000). The effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*, 296–340.

Ferreira, V. S., & Firato, C. E. (2002). Proactive interference effects on sentence production. *Psychonomic Bulletin & Review*, *9*, 795–800.

Finley, S., & Badecker, W. (2009). Artificial language learning and feature-based generalization. *Journal of Memory and Language*, *61*, 423–437.

Frank, A., & Jaeger, T. F. (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*.

Genzel, D., & Charniak, E. (2002). Entropy rate constancy in text. In *Proceedings of the Association of Computational Linguistics* (pp. 199–206). Philadelphia, PA.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*, 8051–8056.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (Ed.), *Universals of Human Language* (pp. 73–113). Cambridge, Mass: MIT Press.

Greenberg, J. H. (1966). *Universals of language*. (2nd ed.). Cambridge: MIT Press.

Grimm, S. (2012a). Individuation and inverse number marking in Dagaare. In D. Massam (Ed.), *Count and Mass Across Languages*. Oxford: Oxford University Press.

Grimm, S. (2012b). *Number and Individuation*. Ph.D. thesis Stanford University.

Grimm, S. (2018). Grammatical Number and the Scale of Individuation. *Language*, *94*, 527–574.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, *42*, 25–70.

Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics*, *19*, 1–33.

Haspelmath, M. (2013). Occurrence of nominal plurality. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Haspelmath, M., & Karjus, A. (2017). Explaining asymmetries in number marking: Singulatives, pluratives, and usage frequency. *Linguistics*, *55*, 1213–1235.

Horn, L. R. (1984). Toward a new taxonomy for pragmatic inference: Q-based and R-based implicatures. In D. Schiffrin (Ed.), *Meaning, Form, and Use in Context* (pp. 11–42). Washington D.C.: Georgetown University Press.

Hudson Kam, C., & Newport, E. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, *1*, 151–195.

Hudson Kam, C. L. (2015). The impact of conditioning variables on the acquisition of variation in adult and child learners. *Language*, *91*, 906–937.

Hudson Kam, C. L., & Newport, E. L. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive psychology*, *59*, 30–66.

Jaeger, T. F. (2006). *Redundancy and Syntactic Reduction in Spontaneous Speech*. Phd thesis Stanford University Stanford, CA.

Jaeger, T. F. (2008). Categorical Data Analysis: Away from ANOVAs (transformation or not) and towards Logit Mixed Models. *Journal of Memory and Language*, *59*, 434–446.

Jaeger, T. F. (2010). Redundancy and reduction: speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23–62.

Jaeger, T. F. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, *4*, 1–4.

Jaeger, T. F., & Buz, E. (2017). Signal Reduction and Linguistic Encoding. In E. M. Fernandez, & H. S. Cairns (Eds.), *Handbook of Psycholinguistics*. Wiley-Blackwell.

Jakobson, R. (1932). Zur struktur des russischen verbums. In *Charisteria Guilelmo Mathesio Quinquagenario* (pp. 74–83). Prague: Prask Lingvistick Krouek. Reprinted in Selected Writings, Vol. II, The Hague, Mouton.

Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: Language users optimize a miniature lexicon for efficient communication. *Cognition*, *165*, 45 – 52.

Kean, M.-L. (1975). *The theory of markedness in generative grammar*. Ph.D. thesis MIT.

Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 10681–10686.

Kirby, S., Dowman, M., & Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proceedings of the National Academy of Sciences*, *104*, 5241–5245.

Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, *28*, 108–114.

Kurumada, C., & Jaeger, T. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, *83*, 152–178.

Levinson, S. (2000). *Presumptive Meanings*. The MIT Press.

Levy, R., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems (NIPS)* (pp. 849–856). Cambridge, MA: MIT Press volume 19.

Lockridge, C. B., & Brennan, S. E. (2002). Addressees' needs influence speakers' early syntactic choices. *Psychonomic Bulletin & Review*, *9*, 550–557.

MacDonald, M. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, *4*, 226.

Mesoudi, A., & Whiten, A. (2008). The multiple roles of cultural transmission experiments in understanding human cultural evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *363*, 3489–3501.

Middleton, E. L., Wisniewski, E. J., Trindel, K. A., & Imai, M. (2004). Separating the chaff from the oats: Evidence for a conceptual distinction between count noun and mass noun aggregates. *Journal of Memory and Language*, *50*, 371 – 394.

Mifsud, M. (1996). The collective in Maltese. *Rivista di Linguistica*, *8*, 29–51.

Moreton, E., & Pater, J. (2012). Structure and Substance in Artificial-Phonology Learning, Part II: Substance. *Language and Linguistics Compass*, *6*, 702–718.

Norcliffe, E., & Jaeger, T. F. (2016). Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition*, *8*, 167–205.

Pajak, B., Fine, A. B., Kleinschmidt, D. F., & Jaeger, T. F. (2016). Learning Additional Languages as Hierarchical Probabilistic Inference: Insights From First Language Processing. *Language Learning*, *66*, 900–944.

Pate, J. K., & Goldwater, S. (2015). Talkers account for listener and channel characteristics to communicate efficiently. *Journal of Memory and Language*, *78*, 1–17.

Peirce, C., Hartshorne, C., & Weiss, P. (1974). *Collected Papers of Charles Sanders Peirce*. Number 5-6 in Collected Papers of Charles Sanders Peirce. Belknap Press of Harvard University Press.

Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*, 3526–3529.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *The Journal of the Acoustical Society of America*, *118*, 2561.

Prasada, S., Ferenz, K., & Haskell, T. (2002). Conceiving of entities as objects and stuff. *Cognition*, *83*, 141–165.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: relating regularization to inductive biases through iterated learning. *Cognition*, *111*, 317–328.

Rose, D. (2017). *Predicting plurality: An examination of the effects of morphological predictability on the learning and realization of bound morphemes*. Ph.D. thesis University of Canterbury.

Samara, A., Smith, K., Brown, H., & Wonnacott, E. (2017). Acquiring variation in an artificial language: Children and adults are sensitive to socially conditioned linguistic variation. *Cognitive Psychology*, *94*, 85–114.

Schumacher, A. R., Pierrehumbert, J. B., & LaShell, P. (2014). Reconciling inconsistency in encoded morphological distinctions in an artificial language. *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, (pp. 2895–2900).

Seyfarth, S. (2014). Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition*, *133*, 140–155.

Singleton, J., & Newport, E. (2005). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive psychology*, *49*, 370–407.

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *372*.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*, 444–449.

Sóskuthy, M., & Hay, J. (2017). Changing word usage predicts changing word durations in New Zealand English. *Cognition*, *166*, 298 – 313.

Stolz, T. (2001). Singulative-collective: Natural morphology and stable cases in Welsh number inflexion of nouns. *Sprachtypologie und Universalienforschung*, *54*, 52–76.

Tiersma, P. (1982). Local and general markedness. *Language*, *58*, 832–849.

Tily, H. J., Frank, M. C., & Jaeger, T. F. (2011). The learnability of constructed languages reflects typological patterns. *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, (pp. 1364–1369).

Trubetzkoy, N. (1931). Die phonologischen systeme. *Travaux du Cercle Linguistique de Prague*, *4*, 96–116.

Von Heusinger, K., & Kaiser, G. A. (2005). The evolution of differential object marking in Spanish. In *Proceedings of the Workshop: Specificity and the Evolution/Emergence of Nominal Determination Systems in Romance* (pp. 33–69).

White, J. (2014). Evidence for a learning bias against saltatory phonological alternations. *Cognition*, *130*, 96–115.

Wierzbicka, A. (1988). *The Semantics of Grammar*. Amsterdam: John Benjamins.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley.