Communicative efficiency in language production and learning: Optional plural marking

Chigusa Kurumada
ckuruma2@ur.rochester.edu
Department of Brain and Cognitive Sciences, Meliora Hall
University of Rochester

Scott Grimm
scott.grimm@rochester.edu
Department of Linguistics, Lattimore Hall
University of Rochester

Abstract

Recent work suggests that language production exhibits a bias towards efficient information transmission. Speakers tend to provide more linguistic signal for meaning elements that are difficult to recover while reducing contextually inferable (more frequent, probable, expected) elements. This trade-off has been hypothesized to shape grammatical systems over generations, contributing to cross-linguistic patterns. We put this idea to an empirical test using miniature artificial language learning over variable input. Two experiments were conducted to demonstrate that the inferability of plurality information inversely predicts the likelihood of overt plural marking, as would be expected if learners prefer communicatively efficient systems. The results were obtained even with input frequency counts of the plural marker counteract the bias, and thus provide strong support for critical role of inferability of meaning in language learning, production, as well as in typologically attested variations.

Keywords: language production, artificial language learning, optional morphology, plural marking, communicative efficiency

Introduction

Producing language is a balancing act. On the one hand, the speaker is biased towards minimizing effort by choosing a shorter form and linguistic elements that are readily retrieved and formulated (e.g., Ferreira & Dell, 2000; MacDonald, 2013). On the other hand, the speaker's choices are, at least to some extent, optimized under considerations of communicative success. The speaker is more likely to encode a linguistic message that is otherwise less predictable or recoverable (e.g., Aylett & Turk, 2004; Levy & Jaeger, 2007; Buz, Tanenhaus, & Jaeger, 2016). For instance, in English, native speakers are more likely to produce the optional complementizer "that" after verbs that are less likely followed by a complement clause (e.g., I read (that) the president was arrested.) compared to those that are biased towards a complement clause continuation (e.g., I thought (that) the president was arrested) (Jaeger, 2010).

It has been argued that such speakers' preferences reflects a principle of the computational system underlying language production. That is, linguistic communication necessarily involves transmission of information through a noisy channel and the information is often degraded due to factors such as production/comprehension mistakes, ambient noise, and noise in the perceptual and neural mechanisms involved in language processing. The comprehender, therefore, has to *infer* a message sent by the speaker rather than simply recognizing and decoding the input (e.g., ? (?), see also Piantadosi, Tily, and Gibson (2011)). A trade-off between the amount of information and the amount of linguistic signal expended is expected when the speaker encodes a message in a way so that the listener has a higher chance of inferring it given the linguistic input and contextually shared knowledge. Put simply, the speaker should preferentially encode components of meanings that are otherwise less likely to be inferred by the listener given prior expectations.

An appealing property of this account is that it provides a potential explanation for typological patterns (although the link is tentative so far) beyond individual instances of sentence production. It has long been observed that the lexicon and grammar of languages across the world tend to exhibit many properties that would be expected if language was shaped by communicative pressures (e.g., Zipf (1949); Plotkin and Nowak (2000), also precisely those predicted by accounts of communicatively efficient language production Piantadosi et al. (2011); Jaeger (2013)). Recent work on learning biases during (miniature artificial) language acquisition has also found the same biases to be active during artificial language learning (e.g., Culbertson, Smolensky, & Legendre, 2012; Fedzechkina, Newport, & Jaeger, 2016). Fedzechkina et al. (2012) found that native speakers of American English, when learning a miniature language with an optional case marking morphology, restructure the input and condition the uses of the marker on factors such as Animacy. This is in line with patterns observed in existing optional (or more categorical) case-marking language, suggesting a tight link between observations in lab-based studies and typological pattern found in real languages e.g., (Aissen, 2003; Kurumada & Jaeger, 2015).

We provide a novel investigation of the possible role of communicative efficiency in grammatical *number* marking, in particular the acquisition of Optional Plural Marking (OPM). OPM is not uncommon cross-linguistically (e.g., Yucatec Maya (Butler, Bohnemeyer, & Jaeger, 2017)) and has been investigated linguistic work on grammatical systems (see Corbett (2000) and Haspelmath (2013) for general discussion). Yet, the mechanisms that predict when speakers would use (or would not use) the marker are not well understood.

Two classes of accounts have been put forward. One relies on form-based frequency of the input. That is, learners are more likely to hear the optional marker with a particular class of nouns and reproduce the distributional patterns in the their production (e.g., Tiersma, 1982; Haspelmath & Karjus, 2017). This is largely consistent with a general view

in language acquisition research: higher input frequency in the input often predicts earlier acquisition and higher usage frequency of the element.

A second set of accounts, relying on conceptual "markedness", make a distinct prediction. These accounts posit that singular (plural) values are conceptually consonant with some entity types more than others. For instance, entities that are typically conceptualized as individuals (e.g., large animals) tend to be referenced in language as singletons, rather than multiples. For these entities, their occurrence in multiples is limited, thus resulting in lower plural inferability, and therefore, plural coding is the unexpected or "marked" value. Conversely, entities that are often conceptualized as collectives (e.g., small insects) have high plural inferability.

We argue that the account based on conceptual markedness, or *meaning-based* predictability, accords with the communicative efficiency hypothesis. Put simply, learning and production of OPM is guided by a consideration to communicate the plural meaning most efficiently. That is, learners should prefer systems in which markedness of plural meaning is inversely correlated with the production of plural marking. Accounts based in communicative efficiency thus predict that, when learners of an OPM language refer to multiples of individualized items (e.g., large animals), they should be more likely to produce plural marking, compared to when referring to multiples of collective items (e.g., small insects).

Preliminary support for the conceptual markedness account comes from repeated observations across a number of studies on typologically-diverse languages which possess a singulative/collective morphology (e.g., Arensen (1998) on Murle, Grimm (2012) on Dagaare, Mifsud (1996) on Maltese, Stolz (2001) on Welsh). In these languages, referents that are likely to be conceptualized and manipulated as collectives (e.g., fruits, grains, vegetables) or a group/mass of individuals tend to be expressed with lexical items that have a plural meaning by default (e.g., *psy* "peas" in Welsh) and only through an additional singulative suffix can singletons be designated (e.g., *psy-en* "pea").

While effects of markedness on number-marking morphology have been hypothesized and widely discussed in linguistics, it is particularly difficult to differentiate predictability of forms and predictability (markedness/inferability) of meanings in a corpus based method. Haspelmath and Karjus (2017), for instance, collected token counts of singular vs. plural forms of a word (e.g., *psy-en* and *psy*) to argue that frequency asymmetries can predict the asymmetrical plural marking system such that the more frequent meaning (singular/plural) is often encoded in a simpler form. However, as in most of existing corpus-based approaches, one cannot easily dissociate the frequency of forms and the frequency of meanings. In other words, there is no simple way of measuring the inferability of meanings apart from the frequency of forms.

Here we aim to tease apart these two possible accounts using a miniature language learning paradigm. We present two production experiments on optional number-marking. Learners acquire 12 novel nouns and one novel verb to produce simple intransitive sentences with the Subject-Verb word order. As we describe below, the novel lexicon consists of two classes of referents: six Individuals and six Collectives that depict fictitious animals and insects, respectively. In the input, they were visually presented as either singletons or multiples at varying rates: Individuals are more likely to be singletons whereas Collectives are more likely to be multiples. Referents are optionally (stochastically) plural-marked and the probability of occurrence of the marker was constant across Individuals and Collectives. Notice that, given the fact that Collectives are more likely to appear as multiples, a larger proportion of token counts of Collectives appeared with the plural marker than Individuals.

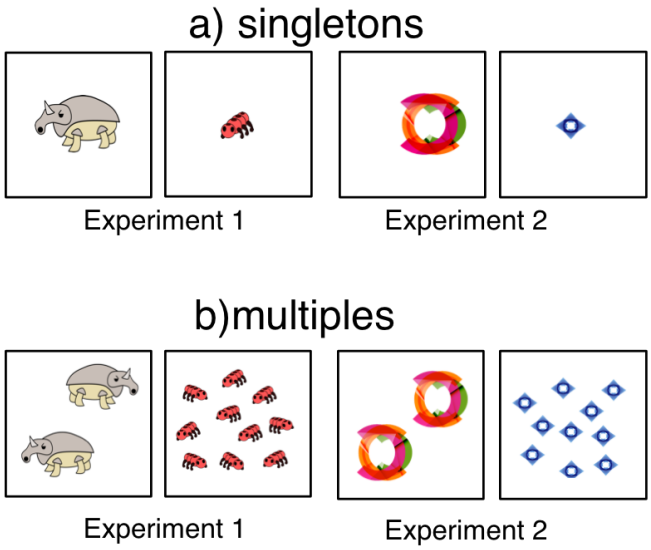

Figure 1: Sample images of visual stimuli in Experiments 1 and 2.

Frequency-based accounts therefore predict that learners of this miniature language should be more likely to use the plural marker with the Collectives rather than Individuals. On the other hand, conceptual-markedness based accounts would predict the opposite. Individuals are less likely to appear as multiples, which makes the plural meaning less inferable without the overt marking. Therefore, language should be more likely to use the plural marker with the Individuals rather than Collectives.

Experiment 1 in the current study directly pits the frequency vs. communicative-efficiency accounts against each other. Animals (visually and conceptually more individuated) are more likely to be presented as a singular referent. In contrast, insects (visually and conceptually more collective) are more likely to be presented as multiples. We test whether the inferability of plurality information affects the likelihood of overt marking (producing a plural marker), as would be expected if learners prefer communicatively efficient systems. Experiment 2 investigates if the inferability of plural meaning is learned through exposure within the current experiment or

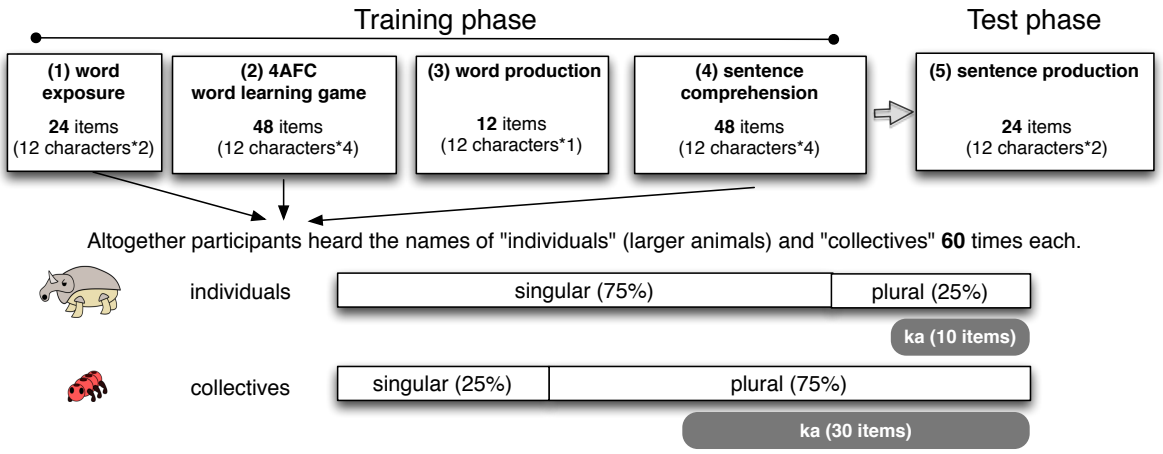

Figure 2: Schematic illustration of the flow of the experiment and proportions of singular and plural visual prompts.

influenced by participants' prior experiences.

Experiment 1

We employ a miniature artificial language learning paradigm modifying Fedzechkina, Jaeger, and Newport (2012). Learners first learn 12 nouns and then learn to produce intransitive sentences in response to prompt video clips. We manipulated visual features of the referents (e.g., size, group size, movements) as well as the probability with which Individuals (animals) and Collectives (insects) appear as singletons and multiples, respectively. If optional number-marking is affected by a preference for communicative efficiency, speakers should be more likely to produce responses with a plural-marker for Individual (animal) compared to Collective (insect) referents.

Methods

Participants 40 native speakers of American English at University of Rochester participated in this study. They received $10 for their participation.

The language

Lexicon We constructed twelve nonce nouns. Six of them denote large animal characters and the other six denote small insect characters (e.g., Fig.1). To ensure that results did not include spurious phonological effects, we created two versions of of character-noun combinations. All of the nouns were 1-2 syllables following the English phonotactics (e.g, *norg*, *velmick*, *zamper*). When characters were presented as multiples, the noun was optionally suffixed with the plural-marker (*-ka*) that optionally marked 2/3 of the time.

We included only one verb – *glim* – meaning "moving up and down". In a sentence, the verb followed a noun, constituting a SV (intransitive) word order (e.g., *Velmick-ka glim*).

Procedure

There were five phases in this experiment (Fig. 2). Participants went through (1) - (3) for six of the twelve noun types (three animals and three insects) and then repeated the same procedure to learn the other six words.

(1) Word exposure (12 characters * 2 = 24 trials total): During word exposure participants were presented with pictures of each of the characters. Participants were instructed to repeat the names of the characters aloud. In this phase, all the characters were presented as singletons. An animal was depicted approximately three times as large as an insect.

(2) 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 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, Fig. 1a), and 25% as multiples (Fig. 1b). Collectives had the inverse distribution (25% singleton, 75% multiples). Both Individual (animal) nouns and Collective (insect) nouns were followed by the plural-marker (*ka*) 2/3 of the time when occurring as multiples.

(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.

(4) Sentence comprehension (12 characters * 4 = 48 trials total): During the sentence comprehension phase, participants viewed short clips and heard their descriptions in the novel language. Participants were asked to repeat the sentences out loud. 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 multiples. Consequently, participants heard the animal and insect nouns with *ka* 10 times and 30 times, respectively by the end of this phase (Fig. 2). Critically, this means that input frequency biases

against the prediction of communicative-efficiency: the input in our experiment(s) provides more instances of training for plural-marked Collectives than Individuals.

(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 had to produce intransitive descriptions. In this phase, visual images 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.

Scoring

In the 4AFC comprehension test, participants' responses were scored as 'correct' if they matched the intended referent. Following the standard in similar studies (e.g., Fedzechkina et al. (2012)), we *a priori* decided to exclude participants who failed to achieve mean accuracy of 65% from all analyses.

We transcribed the production obtained in (5) and annotated if participants produced a given noun correctly and if a noun was produced with *ka* or not. In the comprehension test, participants responses were scored as "correct" if it matched the provided input while subtle phonological variations (e.g., *velmick* pronounced as *belmick*) were ignored.

Results and Discussion

Comprehension Accuracy To ensure that participants have achieved a sufficient level of accuracy in identifying referents, we first measured their performance in the 4AFC word learning game. The average rate of correct response was 74% and all the subject means were above the pre-determined cut-off rate of 65%. The mean accuracy of the word production phase (3) was above 85%. This suggests that the task was feasible and the lexicon was acquired reasonably well before participants performed the production task.

Plural marker use in Production We excluded five (12.5%) of the participants who failed to produce 50% of the sentences in the final sentence production phase. This was done to ensure that the data analyzed are produced by those who have mastered the language at a more or less sufficient level. All the results we report below remain unchanged, however, when we include all the participants. We then further removed 105 (14.5%) sentences that included wrong nouns such as a different character's name or a noun that did not belong to the learned lexicon. The final dataset included 35 subjects and 619 sentences.

Proportions of participants' plural marker use in Experiment 1 are illustrated in Figure 3. To analyzed the data, we used a mixed effect logit model in R, predicting the use of the optional plural marker. We included the noun classes (Individuals (animals) vs. Collectives (insects)) and visual prompts (singleton vs. multiples) as fixed effects and participants and items as random effects. The model included the maximal random effects structure justified by the data based on model comparison (Jaeger, 2008). There was an expected significant main effect of visual prompts such that participants were more likely to produce the optional plural marker *ka* for multiples ($p < .001$). Critically, the interaction between the noun class and the visual prompts was also significant ($p < .03$): Learners (inversely) conditioned plural production on plural inferability. 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).

Experiment 2

What is deriving the observed difference between Individuals and Collectives? Under our hypothesis, it is at least partially due to the expectation that animals are *less* likely to be represented with the plural meaning, and hence the meaning is less inferable. (And the opposite is the case for insects). In Experiment 1, however, it is not clear if the inferability of the plural meaning (the conditional probability of multiples given the referent) is learned within the experiment or it is carried over from participants' prior semantic knowledge that insects are more likely to occur, and be referred to, as multiples.

To separate these two factors, in Experiment 2, we used the lexical items from Experiment 1 while associating them with novel, semantically bleached, items to remove effects of prior semantic knowledge. If participants exhibit the same asymmetric use of the plural marker for Individuals and Collectives, that will yield support for the idea that the inferability is likely extrapolated in this experiment.

Participants

20 native speakers of American English at University of Rochester participated in this study. They received $10 for their participation.

The language

The lexicon was identical to that used in Experiment 1. The only difference is that the visual images consisted of 12 geometrical shapes with no commonly known names. To equate the visual features of the referents (e.g., size, spacial distributions, complexity of visual scenes), we created two classes of referents (Fig. 1). One of the classes (Individuals) consists of six relatively large geometrical shapes spatially distributed in a manner similar to how the animals were presented in Experiment 1. The other class (Collectives) consists of six smaller shapes that replace the insects in Experiment 1.

Procedure

The same as Experiment 1.

Results and discussion

Comprehension Accuracy The mean accuracy in the 4AFC task was 68%, suggesting that the word learning was slightly more difficult in Experiment 2 compared to Experiment 1, presumably due to the overall unfamiliarity with the geometrical shapes. One subject could not achieve the cut off rate of

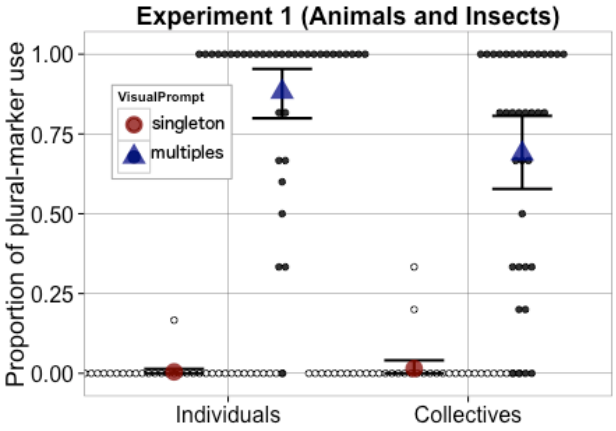

Figure 3: Proportions of plural marker use by conditions. Dots present by-participant averages. Error-bars show 95% Confidence Intervals.

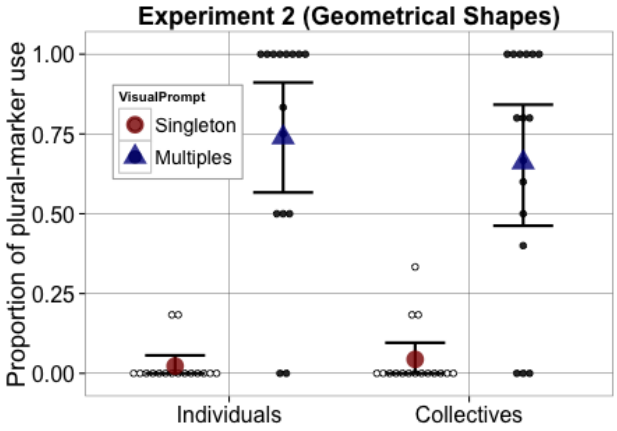

Figure 4: Proportions of plural marker use by conditions. Dots present by-participant averages. Error-bars show 95% Confidence Intervals.

65% and was removed from the analysis. The mean accuracy in the word production phase (3) was 80%.

Plural marker use in Production We excluded three (15.7%) of the participants who failed to produce 50% of the sentences in the final sentence production phase. As in Experiment 1, all the results we report below remain unchanged with the complete set of data. We then further removed 99 (25.9%) sentences that included wrong nouns. The final dataset included 16 subjects and 283 sentences.

Proportions of participants' plural marker use in Experiment 2 are illustrated in Figure 4. The same model from Experiment 1 was used to predict participants' use of *-ka*. While the main effect of visual prompt was significant ($p < .001$), the interaction between the noun classes and visual prompt was not ($p > .2$). Learners were equally likely to produce the optional plural marker for both Individuals and Collectives, suggesting that the effect in Experiment 1 was likely driven by prior semantic knowledge of the semantic classes (animals vs. insects). The relative conditional probability of multiples in the input was not sufficient to induce this effect, perhaps requiring longer exposure (to be tested).

General Discussion

Our results suggest that native speakers of American English prefer to produce an NP *without* overt marking of plurality when the meaning is more inferable given the semantics of the noun classes (e.g., animals vs. insects). The effect was not present when the nonce shapes were used even though within-experiment statistics as well as visual features of referents (size, spacial arrangements, movement patterns) were held constant. This suggests that learners have knowledge of the relative inferability of plural meaning for different types of referents (e.g., How often do you describe animals/insects as singletons vs. plural referents?), and this knowledge supports the learning of morphological systems of a novel language. Critically, English does not have the optional plural marking (OPM) system. Still, when native speakers of English are exposed to an OPM language with no bias to mark plurality for low-inferability items, they end up producing more plural marking for less inferable items. As such this is one of the first studies showing a systematic effect of semantic knowledge on the morpho-syntactic encoding of speech.

This body of research including the current study constitutes strong support for the view that language production is optimized to maximize the efficiency of information transmission (Levy & Jaeger, 2007, Jaeger, 2010). The asymmetrical uses (and non-uses) of *-ka* cannot be accounted for in terms of availability of an upcoming linguistic element or other sources of speaker-internal production or planning difficulties (Ferreira & Dell, 2000; MacDonald, 2013). All the sentences were produced with the same verb and no participant failed to learn to produce it.

It is possible, however, that the difference between Experiment 1 and Experiment 2 stems from the differential levels of mastery in word learning. Participants learned the nonce labels better in Experiment 1 than in Experiment 2, presumably due to more easily memorable visual referents (animals/insects as opposed to geometrical shapes). The facilitation in word learning might have made it easier for participants to modulate their production with respect to communicative considerations. In a future study, we intend to increase the amount of of 1) linguistic and 2) non-linguistic (inferability) information in the input in Experiment 2 to see if participants would show an asymetrical use of the OPM.

Lastly, this study has broad implications for understanding typologically attested morpho-syntactic variations. It has long been hypothesized that conceptual markedness plays a guiding role in grammaticalization of morpho-syntactic elements. The current experimental paradigm using an artificial language allows us to dissociate the effects of input in terms

of the predictability of forms (e.g., How often do you hear a particular noun with *-ka*?) and the predictability/inferability of meaning (e.g., How likely is it that a given referent is described as a singleton vs. multiples?), making it possible to test a multitude of hypotheses put forward about effects of meaning-based predictability. For instance, it has been observed that functionally paired objects (e.g., glasses, chopsticks, a set of pillars) and body-parts (e.g., eyes, ears, hands) are often conceptualized as plural by default, and hence likely encoded without any additional plural marking morphology (Haspelmath & Karjus, 2017). We can directly test this hypothesis in the current paradigm using objects that differ in their likelihood of appearing in pairs.

In summary, the current results yield support for the hypothesis that the inferability of plurality information guides learners to restructure the input they receive, as would be expected if language users are biased towards communicately efficient systems. Our results thus illuminate the critical role of distributional information of meanings on language learning, production and typological variation across languages.

Acknowledgments

Thanks to Wesley Orth, Joseph Plvan-Franke, Sahed Martinez, Kathleen Ward, and Jennifer Andrews for assistance with stimuli construction, data collection, and annotation, and to T. Florian Jaeger, Masha Fedzechkina, Kurumada-Tanenhaus Lab, and Experimental Semantics/Pragmatics group for valuable discussion.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory*, *21*, 435–483.

Arensen, J. E. (1998). Murle categorization. In G. J. Dimmendaal & M. Last (Eds.), *Surmic languages and cultures* (pp. 281–318). Rüdiger Köppe.

Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis. *Language and Speech*, *47*(1), 31–56.

Butler, L. K., Bohnemeyer, J. B., & Jaeger, T. F. (2017). Plural marking in Yucatec Maya at the syntax-processing interface. In A. Machicao y Priemer, A. Nolda, & A. Sioupi (Eds.), *Zwischen Kern und Peripherie (Studia Grammatica, volume 75)*. Berlin: Akademie-Verlag.

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.

Corbett, G. G. (2000). *Number*. Cambridge: Cambridge University Press.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*(3), 306–329.

Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *PNAS*, *44*(109), 17897–17902.

Fedzechkina, M., Newport, E. L., & Jaeger, T. F. (2016). Balancing effort and information transmission during language acquisition : Evidence from word order and case marking. *Cognitive Science*, 1–31.

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

Grimm, S. (2012). Individuation and inverse number marking in Dagaare. In D. Massam (Ed.), *Count and mass across languages*. Oxford: Oxford University Press.

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*.

Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *JML*, *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.

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

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)* (Vol. 19, p. 849-856). Cambridge, MA: MIT Press.

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

Mifsud, M. (1996). The collective in Maltese. *Rivista di Linguistica*, *8*(1), 29-51.

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). The communicative function of ambiguity in language. *Cognition*, *122*, 280–291.

Plotkin, J., & Nowak, M. (2000). Language evolution and information theory. *Journal of Theoretical Biology*, *205*(1), 147–159.

Stolz, T. (2001). Singulative-collective: Natural morphology and stable cases in Welsh number inflexion of nouns. *Sprachtypologie und Universalienforschung*, *54*(1), 52–76.

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

Zipf, G. K. (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Addison-Wesley.