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 inferrable (more frequent, probable, or 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 inferrability 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 counteracting the bias, and thus provide strong support for a critical role of inferrability 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

Speakers face a multitude of constraints when encoding their intended message as an actual utterance. On the one hand, speakers want to encode their meaning in a way that guarantees communicative success—it must be understood by the interlocutor. At the same time, speakers need to cope with difficulties associated with utterance planning and articulation. As researchers have shown, speakers regularly do this, e.g. by choosing shorter forms and/or elements that are readily retrieved and formulated (see, inter alia, Ferreira & Dell, 2000; MacDonald, 2013).

A body of psycholinguistic work sees this negotiation between communicative success and effort minimization as a guiding principle of the computational system underlying language production and comprehension. Specifically, it is expected that there is an efficiency-based trade-off between the amount of information encoded and the amount of linguistic signal expended by the speaker. (e.g., Aylett & Turk, 2004; Levy & Jaeger, 2007; Buz, Tanenhaus, & Jaeger, 2016). Communicative efficiency is predicted to be maximized when the speaker preferentially encodes components of meanings that are otherwise less likely to be inferred by the listener given prior expectations.

Against this backdrop, we consider the possible role of communicative efficiency in the organization of grammatical number marking. Grammatical number systems often have "markedness" contrasts between a default, uncoded value, and a value explicitly coded, e.g. through morphology. Typically, the singular value is uncoded while the plural value is coded, as in *dog* vs. *dogs*. One question we will ask is if a plural value for a referent is likely to be inferred, will a speaker encode it? While it has been long observed that languages have preferences for what information is coded in default forms as opposed to explicitly coded (e.g., Greenberg, 1966), the causes underlying these preferences have remained obscure. The design of our study allows us to take a step towards distinguishing what sort of information forms the basis for these preferences. In particular, we investigate whether the active ingredient is the predictability of linguistic forms, i.e. frequency of occurrence of some element in produced language, or if predictability is related to the meaning.

A case study: Optional plural marking

Unlike in English, grammatical encoding of plural meaning (e.g., *dog* vs. *dogs*) can be optional in some languages. Optional Plural Marking (OPM) is not uncommon cross-linguistically (e.g., Yucatec Maya (Butler, Bohnemeyer, & Jaeger, 2017)) and has been investigated in 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.

A class of proposals, elaborated for number marking more generally, grounds the encoding of number values in conceptual properties related to entities (Prasada, Ferenz, & Haskell, 2002; Wisniewski, Lamb, & Middleton, 2003). This viewpoint suggests that singular (or plural) values might be more conceptually consonant for some entity types than for others. For instance, entities that are typically conceptualized as individuals (e.g., large animals) tend to be referenced in language as singular, rather than plural. For these entities, their occurrence in plural is limited, and therefore, plural coding is the unexpected or "marked" value. Conversely, for entities that are often conceptualized as collectives (e.g., small insects), plural coding is the expected or "unmarked" value. In sum, on this view formal (morphological) markedness corresponds to conceptual markedness.

When combined with a framework such as the communicative efficiency hypothesis, this "markedness" of plural meaning can predict biases seen in language production. Put simply, learning and production 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. In relation to OPM, accounts based in communicative efficiency 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, Mifsud (1996) on Maltese, Stolz (2001) on Welsh, see (Grimm, 2012) for discussion). In these languages, referents that are likely to be conceptualized and manipulated as collectives (e.g., fruits, grains, vegetables) or as 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").

A difficulty arises, however, in determining "markedness" of plural meaning based on token counts of plural forms in a corpus. 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, in this approach, one can only infer the frequency of meaning (e.g., How often does one talk about pea(s) as singular or plural?) based on the frequency of form (e.g., How often does one use a singular or plural form for pea(s)?). In other words, there is no simple way of dissociating predictions of the communicative efficiency account from an account based on form frequency: speakers may be simply reproducing the patterns heard in the input (e.g., They are more likely to hear *psy* than *psy-en* when they see peas and are faithfully representing the pattern in their own production).

To address this problem, we present two production experiments using an artificial language learning paradigm. 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.

This setup pinpoints an instance where frequency (inferrability) of meaning can be examined independent of frequency of forms. For instance, Individuals are less likely to appear as multiples compared to Collectives. This makes the plural meaning less inferrable for Individuals than for Collectives without the overt marking. Therefore, conceptual-markedness based accounts would predict that learners should be more likely to use the plural marker with the Individuals rather than Collectives. Critically, this bias is not predictable based purely on the frequency of forms. Notice that, given the fact that Collectives are more likely to appear as multiples, a larger proportion of token counts of Collectives appear with the plural marker than Individuals. If learners are simply reproducing the patterns observed in the input, they should produce the optional plural marker more with Collectives than with Individuals.

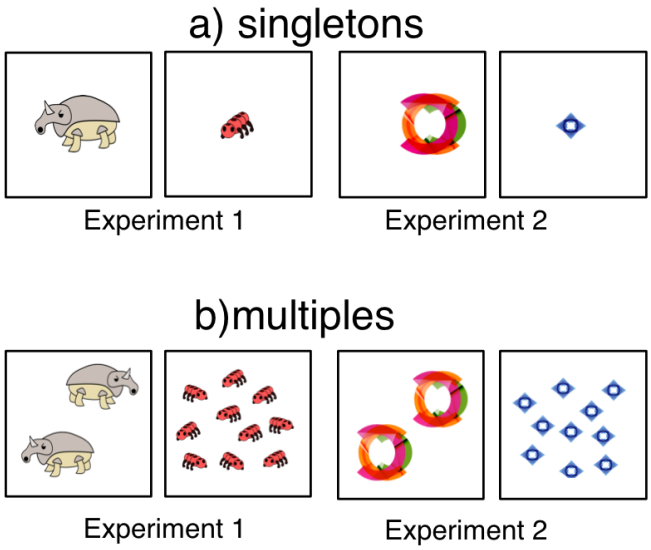

Figure 1: Sample images of visual stimuli in Experiments 1 and 2.

Results from this investigation may help to bridge the gap between the factors shaping sentence production in language processing and those that are shaping typological patterns. 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, Tily, and Gibson (2011); Jaeger (2013)). Recent work on learning biases during (miniature artificial) language acquisition has also found similar biases to be active during artificial language learning (e.g., Culbertson, Smolensky, & Legendre, 2012; Fedzechkina, Newport, & Jaeger, 2016; Smith & Wonnacott, 2010). 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 languages, suggesting a tight link between observations in lab-based studies and typological pattern found in existing languages (e.g., Aissen, 2003; Kurumada & Jaeger, 2015).

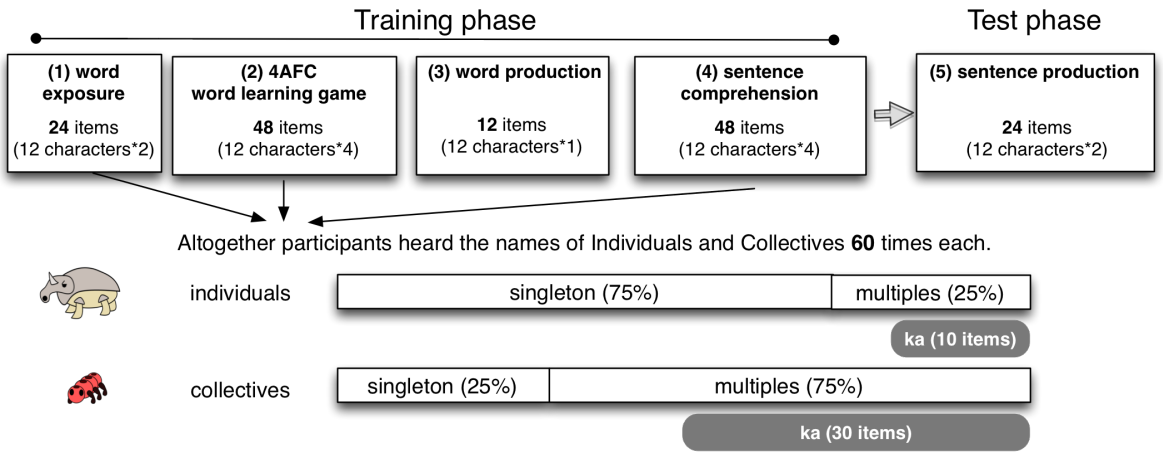

Figure 2: Schematic illustration of the flow of the experiment and proportions of singleton and multiple visual prompts.

Experiment 1

We employ a miniature artificial language learning paradigm modifying Fedzechkina, Jaeger, and Newport (2012). Participants first learn 12 nouns and then learn to produce intransitive sentences in response to video clip prompts. 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 48 native speakers of American English at University of Rochester participated in this study. They received $10 for their participation.

The language

Lexicon We constructed 12 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 character-noun combinations. All of the nouns were 1-2 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.

We included only one verb – *glim* – meaning "moving up and down." In constructed sentences, 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 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.

(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). Importantly, 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 used 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 had 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 93.9% (animals, 93%; insects, 94%) and all the subject means were well 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 six (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 116 (13%) 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 42 subjects and 773 sentences.

Proportions of participants' plural marker use in Experiment 1 are illustrated in Fig. 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 driving 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 inferrable (and conversely for insects). In Experiment 1, however, it is not clear if the inferrability 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 geometrical shapes to minimize 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

52 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). Individuals consisted of six relatively large geometrical shapes spatially distributed in a manner similar to how the animals were presented in Experiment 1. On the other hand, Collectives consisted 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 86% (Animals, 89%; Insects, 83%), suggesting that the word learning was slightly more difficult in Experiment

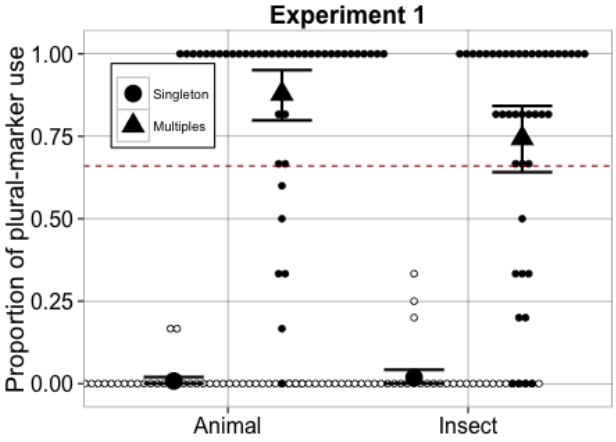

Figure 3: Proportions of plural marker use by conditions. Dots present by-participant averages (White = singleton visual prompt; Black = multiple visual prompt). Error-bars show 95% Confidence Intervals. Dotted line indicates the input ratio of the *-ka* marking for the multiples.

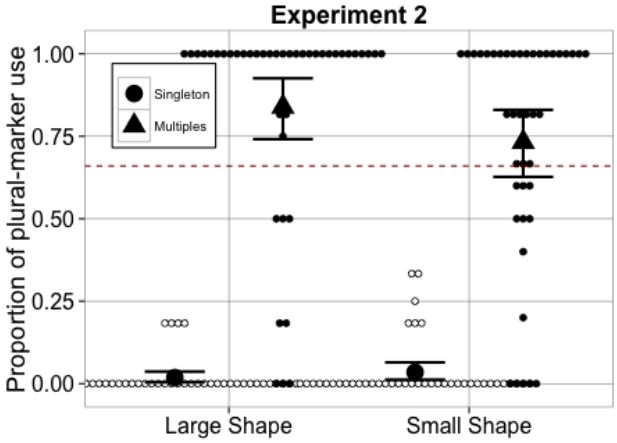

Figure 4: Proportions of plural marker use by conditions. Dots present by-participant averages (White = singleton visual prompt; Black = multiple visual prompt). Error-bars show 95% Confidence Intervals. Dotted line indicates the input ratio of the *-ka* marking for the multiples.

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 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 ten (19.2%) 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 151 (15.5%) sentences that included wrong nouns. The final dataset included 42 subjects and 823 sentences.

Proportions of participants' plural marker use in Experiment 2 are illustrated in Fig. 4. We constructed a combined model with the noun classes (Individuals vs. Collectives), visual prompts (singleton vs. multiples), and experiments as fixed effects and participants and items as random effects. As in Experiment 1, we found a significant main effect of visual prompts (= more *ka* use for multiples) ($p < .001$) and an interaction between the noun class and the visual prompts ($p < .002$), indicating an inverse conditioning of *-ka* production on plural inferrability. Importantly, there was no significant effect of the experiments. This suggests that the plural predictability is not necessarily tied to participants' prior knowledge of the semantic classes (animals vs. insects) and is learnable with respect to new classes of referents.

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 inferrable given the noun classes (e.g., animals vs. insects). The effect was present even with the 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. We thus argue that learners have implicit knowledge of the relative inferrability of plural meaning (e.g., How often do you describe animals/insects as singletons vs. multiples?), 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-inferrability items, they end up producing more plural marking for less inferrable items.

The current results constitute strong support for the view that language production is optimized to maximize the efficiency of information transmission (Levy & Jaeger, 2007, Jaeger, 2013). 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), since all the sentences were produced with the same verb and no participant failed to learn the verb.

It is an open question how learners compute the plural predictability. In the current experiment, we provided multiple cues to noun classes beyond the statistics of singleton vs. multiples. For instance, Individuals were always depicted larger in size than Collectives. In the sentence comprehension and production phases, each instance of Individuals moved independently while Collectives always showed a group motion. Future studies can manipulate these cues separately to delve into effects of spacio-temporal distributions of referents on conceptualization of noun classes and their plural inferrability.

Lastly, this study has broad implications for understand-

ing typologically attested morpho-syntactic variation. 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 inferrability of plurality information guides learners to restructure the input they receive, as would be expected if language users are biased towards communicatively efficient systems. Our results thus illuminate the critical role of distributional information of meanings on language learning, production, and typological variation across languages.

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: CUP.

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.

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

Grimm, S. (2012). *Number and Individuation*. Unpublished doctoral dissertation, Stanford University.

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. (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.

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

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449.

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

Wisniewski, E. J., Lamb, C. A., & Middleton, E. L. (2003). On the conceptual basis for the count and mass noun distinction. *Language and Cognitive Processes*, *18*, 583–624.

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