COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science (2016) 1–31

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12346

Balancing Effort and Information Transmission During Language Acquisition: Evidence From Word Order and Case Marking

Maryia Fedzechkina,[a] Elissa L. Newport,[b] T. Florian Jaeger[c,d,e]

[a]*Department of Psychology, University of Pennsylvania*
[b]*Department of Neurology, Center for Brain Plasticity and Recovery, Georgetown University*
[c]*Department of Brain & Cognitive Sciences, University of Rochester*
[d]*Department of Computer Science, University of Rochester*
[e]*Department of Linguistics, University of Rochester*

Received 13 April 2014; received in revised form 27 October 2015; accepted 5 November 2015

Abstract

Across languages of the world, some grammatical patterns have been argued to be more common than expected by chance. These are sometimes referred to as (statistical) *language universals*. One such universal is the correlation between constituent order freedom and the presence of a case system in a language. Here, we explore whether this correlation can be explained by a bias to balance production effort and informativity of cues to grammatical function. Two groups of learners were presented with miniature artificial languages containing optional case marking and either flexible or fixed constituent order. Learners of the flexible order language used case marking significantly more often. This result parallels the typological correlation between constituent order flexibility and the presence of case marking in a language and provides a possible explanation for the historical development of Old English to Modern English, from flexible constituent order with case marking to relatively fixed order without case marking. In addition, learners of the flexible order language conditioned case marking on constituent order, using more case marking with the cross-linguistically less frequent order, again mirroring typological data. These results suggest that some cross-linguistic generalizations originate in functionally motivated biases operating during language learning.

Keywords: Language acquisition; Learning biases; Language universals; Efficient information transfer; Communicative pressures

Correspondence should be sent to Maryia Fedzechkina, Institute for Research in Cognitive Science, University of Pennsylvania, 3401 Walnut St., Suite 400A, Philadelphia, PA 19104. E-mail: mfedze@sas.upenn.edu

1. Introduction

As observed by linguist Joseph Greenberg (1963), certain grammatical properties tend to co-occur across languages. These statistical tendencies that occur with higher than chance frequency cross-linguistically often take the form of implicational statements (*implicational language universals*[1]): If a language has property A, then it will most likely have property B. The explanation for such (hypothesized) universals has been heavily debated in linguistics and cognitive science. Most theories agree that linguistic universals originate at the level of the individual and suggest that language structures are shaped by biases and limitations of human cognitive systems (Chomsky, 1965; Christiansen & Chater, 2008; Deacon, 1997; Pinker, 1984; but see Dunn, Greenhill, Levinson, & Gray, 2011; Evans & Levinson, 2009, for arguments that language universals originate outside of the cognitive system). Views differ, however, as to whether these biases are specific to language (Chomsky, 1965; Fodor, 2001; Pinker, 1984) or rather arise from constraints shared by at least some other cognitive systems such as memory or perception (e.g., Bates & MacWhinney, 1982; Christiansen & Chater, 2008). Under the latter view, cognitively non-arbitrary pressures on language acquisition and use shape the historical development of languages over time, providing an explanation for cross-linguistic generalizations (Bever, 1970; Bybee, 2007; Christiansen & Chater, 2008; Deacon, 1997; Givón, 1991; Hawkins, 2004; Newport, 1981, 1990; Slobin, 1973; Zipf, 1949, for review, see Jaeger & Tily, 2010).

One example of a linguistic universal that has been of long-standing interest in linguistic typology is the inverse correlation between case and constituent order as means of encoding grammatical function assignment.[2] Typological and historical studies as well as computational simulations have supported the hypothesized trade-off between these two cues. For example, typological studies have found that languages with flexible constituent order often use morphological means, such as case, to mark grammatical function assignment (e.g., German, Japanese, and Russian), whereas languages with fixed constituent order (such as English and Mandarin) often lack productive case marking (Baugh & Cable, 1993; Blake, 2001; Sinnemäki, 2008; for a related discussion, see also Siewierska, 1998).[3] Historical studies have found that case marking and constituent order flexibility trade off over time, such as in the historical development from Old English (a language with flexible constituent order and rich case marking) to Modern English (a language with fixed constituent order and a rudimentary case system) (Marchand, 1951; Sapir, 1921; Traugott, 1972; see also Tily, 2010).

However, typological and historical correlations should be interpreted with caution. First, any approach relying solely on typological or historical data suffers from sparsity of independent data points: Most or even all languages are more or less directly genetically related. Language contact, sometimes extending over centuries, can further diminish independence. This drastically reduces the true effective sample size available for statistical tests of hypothesized universals (for related discussion, see Cysouw, 2010; Dryer, 1989; Jaeger, Graff, Croft, & Pontillo, 2011; Rafferty, Griffiths, & Klein, 2014). Indeed,

recent more statistically advanced analyses that adequately discount dependencies between languages have called into question the validity of many implicational universals (Dunn et al., 2011; but see Croft, Bhattacharya, Kleinschmidt, Smith, & Jaeger, 2011, for discussion). Specifically with regard to the correlation between case and constituent order freedom, some studies have found stronger correlations between constituent order flexibility and *other* morphological means (e.g., agreement, Siewierska, 1998). These studies have left open the question of whether the observed correlation between constituent order flexibility and case marking holds once these other factors are controlled for. Similarly, the historical link between the loss of case marking and the loss of word order flexibility has been called into question (Detges, 2009; Pintzuk, 2002).

More important, while typological studies can uncover a correlation between biases in individual users and patterns in cross-linguistic diversity, they *cannot* address questions about the underlying *causal* relationship between the two. Thus, despite the long-standing interest in the trade-off between constituent order and case marking, it is still unclear whether this phenomenon indeed originates in biases of individual language users.

Here, we draw on an emerging complement to typological data, which allows us to address questions about the causes of linguistic universals. The approach we employ is *miniature language learning*. In this paradigm, child or adult learners are exposed to experimentally designed languages that are small enough to be learned in the lab. In contrast to a typological approach, a miniature language learning paradigm does not suffer from data sparsity: Within the limitations of the paradigm, novel data can be generated (for a discussion of these limitations, see Fedzechkina, Newport, & Jaeger, in press). Perhaps more important, a miniature language learning approach also allows us to directly test hypotheses about the underlying causes of linguistic universals by studying patterns in the acquisition and use of novel miniature languages.

Recent findings from this paradigm provide behavioral evidence for language universals. In a number of such experiments, typologically frequent patterns are learned more easily or are preferred over cross-linguistically less frequent alternatives, thereby providing evidence for the hypothesis that patterns in cross-linguistic diversity originate as biases in language users. For example, Culbertson, Smolensky, and Legendre (2012) behaviorally replicate Greenberg's universal #18. In a miniature language learning experiment, they find that adult learners preferentially acquire languages in which adjective-noun and numeral-noun orders are harmonic (both noun-first or both noun-last); and Culbertson and Newport (2015) show that child learners do this even more strongly. A small but growing number of studies have observed similar parallels between typological data and biases operating during acquisition of miniature languages (e.g., for phonology: Finley & Badecker, 2008; Wilson, 2006; word formation: Newport & Aslin, 2004; morphology: Fedzechkina, Jaeger, & Newport, 2012; Hudson Kam & Newport, 2005, 2009; Hupp, Sloutsky, & Culicover, 2009; St Clair, Monaghan, & Ramscar, 2009; syntax: Culbertson et al., 2012; Culbertson & Newport, 2015; Morgan, Meier, & Newport, 1987; Tily, Frank, & Jaeger, 2011).

Only a few artificial language learning studies have directly addressed the role of functionally motivated learning biases in shaping language structures. In previous work, we

have studied one instance in which we hypothesized that grammatical properties associated with efficient communication originate at least in part in acquisition. Fedzechkina et al. (2012) presented adult participants with miniature languages that had flexible constituent order and optional case marking on either the object or the subject of the sentence, which also varied in animacy. Natural languages tend to use case marking efficiently—that is, they typically condition case marking on semantic properties of the referent, such as animacy, and employ overt case marking when these semantic properties are more likely to bias the listener away from the intended interpretation (cf. Kurumada & Jaeger, 2015). The artificial languages used by Fedzechkina et al. (2012) differed from naturally occurring ones in that they did not condition the use of case marking on the referent's animacy. However, learners did not acquire these inefficient morphological systems veridically. Instead, learners deviated from the input and used more case marking for animate (atypical) objects than for inanimate (typical) objects, making the case systems more efficient (Gibson et al., 2013; Jaeger, 2010; Kurumada & Jaeger, 2015) and more in line with typological data (Aissen, 2003; Fry, 2003; Lee, 2006; Silverstein, 1976).

Here, we extend this line of research to ask whether typological patterns of cue trade-offs at different levels of linguistic organization (e.g., syntax and morphology) can also be explained by preferences for efficient communicative systems. We present new evidence that the inverse correlation between case and constituent order flexibility emerges during language learning. This, we argue, can be understood as a trade-off between effort and robust information transmission (cf., Jäger, 2007; Kurumada & Jaeger, 2015; Zipf, 1949). In languages with relatively fixed constituent order, the ordering of arguments is highly informative about their grammatical function (which is the subject and which is the object of the clause). Hence, there is little uncertainty about the intended meaning after hearing a sentence in such a language (formally, the conditional entropy over grammatical function assignments is low), and morphological marking would provide little additional information beyond that conveyed by constituent order. In languages with flexible constituent order, however, the ordering of arguments alone leaves greater uncertainty about the intended grammatical function assignment. For such languages, morphological marking is an informative cue to grammatical function. In both types of languages the effort required for the production of morphological marking is identical; but the *utility* of morphological marking—that is, the trade-off between production effort and the likelihood of successful message transmission—is higher in flexible constituent order languages (for related arguments, see also Haspelmath, 1999; Hawkins, 2004; Jäger, 2007; Kurumada & Jaeger, 2015; Tily, 2010).

Similarly—and discussed in more detail below—there is an inverse correlation between case marking and constituent order flexibility and learnability: Flexible constituent order languages are learned more successfully by connectionist networks if case marking is available (Lupyan & Christiansen, 2002; Van Everbroeck, 2003). The current work tests whether these simulations are supported by data from human language learners.

As in our earlier work (Fedzechkina et al., 2012), we test whether learners exposed to a communicatively imperfect input language will deviate from their input to "improve"

the language, and whether they will do so by reducing average production effort, reducing average uncertainty about the intended meaning, or both. The present experiment tests this prediction by directly investigating the trade-off between constituent order (a syntactic cue) and case marking (a morphological cue). We further investigate *how* case is used in the flexible constituent order language. We ask, in particular, whether learners increase communicative success by favoring robust information transfer over effort and regularize case marking in the language overall, or whether they favor an efficient balance between these two goals by conditioning case marking on constituent order.

2. Method

2.1. Participants

Monolingual native speakers of English were recruited from the University of Rochester. Recruiting and execution of this study were approved by the Research Subjects Review Board of the University of Rochester. Following our previous work (Fedzechkina et al., 2012), recruitment continued until the number of participants who successfully learned the miniature languages reached 20 in each condition. A total of 52 volunteers were recruited for the experiment, all of whom received $25 for their time. Twelve participants were excluded from the analysis for the following reasons: failure to achieve 65% accuracy on the comprehension test (10 participants); computer error (1 participant); being bilingual (1 participant). This left the data from 40 participants for the analysis.

2.2. Design and materials

Each of the two miniature artificial languages contained 10 novel content words (4 verbs and 6 nouns) and a case marker "kah" (see Table 1). All words were phonotactically legal non-words of English. Individual words were synthesized using the AT&T speech synthesizer (voice "Crystal") and concatenated into sentences with 35 ms silence between the words using Praat (Boersma, 2001). This procedure ensured that the stimuli did not contain prosodic cues to sentence meaning. All sentences described short videos created using Poser Pro software that depicted simple transitive events such as "hug" or

Table 1
The artificial lexicon

Nouns	Verbs	Case Marker
glim	geed	kah
flugit	kleidum	
spad	zamper	
bliffen	shen	
norg		
melnawg		

"poke" performed by two male actors such as "chef" or "referee" (see Fig. 1 and Fig. 2 for example stimuli).

All verbs occurred equally frequently within each language overall and with each constituent order allowed by the language. All nouns occurred equally often in the subject and object position with each verb.

Participants were randomly assigned to one of the two language conditions. Both languages contained optional case marking: 67% of all objects were marked with a case maker "kah" and 33% of objects had no overt marking. Subjects were never case marked in either of the languages. Both languages had head-final constituent order (i.e., the verb followed both the subject and the object). This constituent order was chosen since it is cross-linguistically more common in languages with a case system (Dryer & Haspelmath, 2011; Greenberg, 1963).

The two languages differed in their constituent order consistency (and therefore in the amount of information that constituent order conveyed about sentence meaning). In the *flexible constituent order language,* subject-object-verb (SOV) and object-subject-verb (OSV) orders occurred equally frequently in the input. Thus, in this language, constituent order was uninformative about grammatical function assignment, and case marking added important information for decoding sentence meaning. The *fixed constituent order language* did not contain constituent order variation: SOV constituent order occurred in

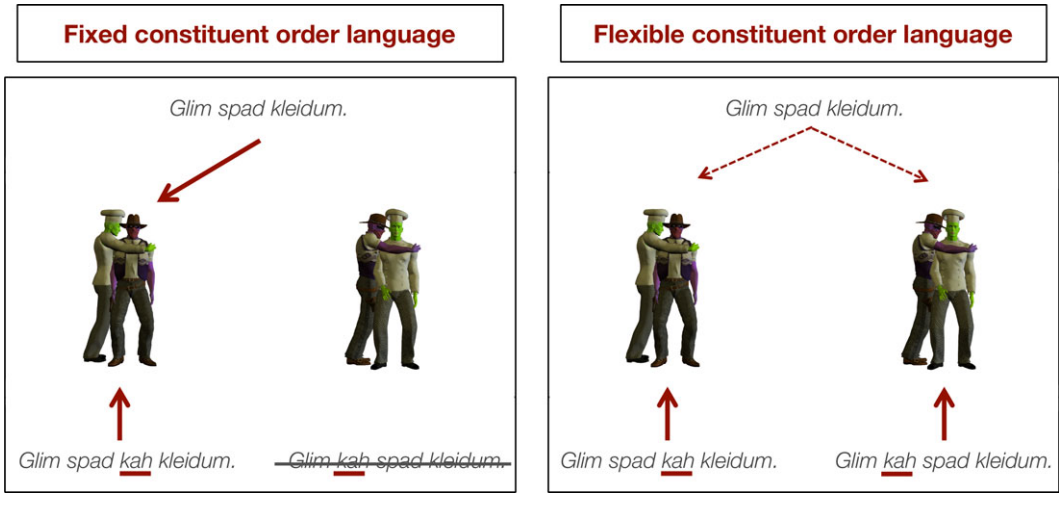

Fig. 1. Illustration of form to grammatical function mappings in two extreme scenarios (completely fixed and completely free constituent order languages) in the experiment. Pictures are still images of sample videos with their descriptions in the miniature language (not shown to participants). Arrows indicate form to meaning mappings in the two languages. Solid arrows indicate one-to-one form-meaning mappings such as in the absence of constituent order variation or in the presence of case marking (underlined). Dashed arrows indicate one-to-many form-meaning mappings, such as in the absence of case marking for variable constituent order. The strikethrough form refers to the ungrammatical (OSV) sentences in the fixed (SOV-only) language.

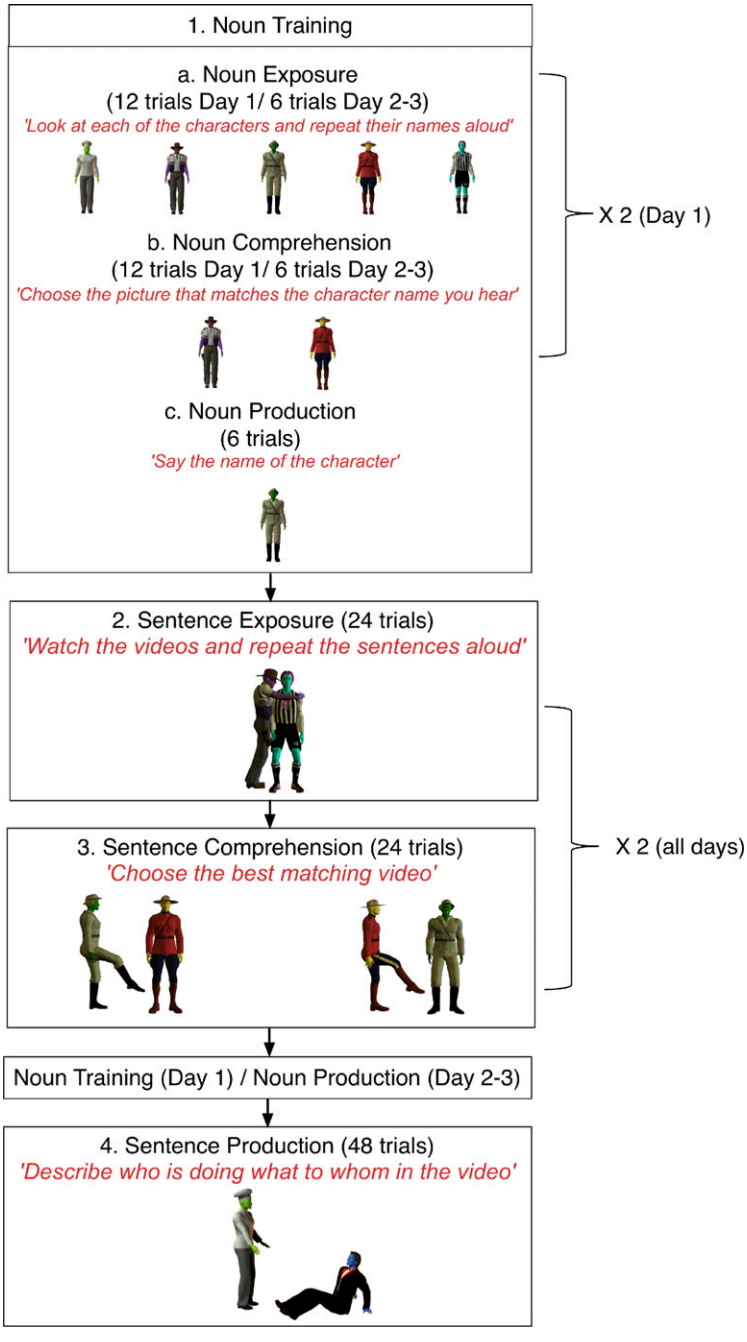

Fig. 2. Experimental procedure. Pictures are still images of sample videos used in the experiment.

100% of the input sentences. In this language, constituent order was highly informative and always disambiguated grammatical function assignment; case marking was a redundant cue to grammatical function.

If language acquisition is indeed biased toward efficient communicative systems, we expect this to be reflected in the languages that learners acquire. This makes two predictions. First and most important, we predicted that learners would use more case marking when it is informative of grammatical function assignment (flexible constituent order language) than when it is a redundant cue (fixed constituent order language). Second, if learners strongly disprefer to expend effort when it is not required for successful communication, we might see that the redundant case marking is more frequently omitted by learners of the fixed constituent order language, compared to the input language.

2.3. Procedure

The procedure was adopted from Hudson Kam and Newport (2005, 2009). At the beginning of the experiment, participants were informed that they would be learning a novel "alien" language by watching short videos and hearing their descriptions in this language, but they received no explicit instructions about the structure of the language.

Participants completed three experimental sessions (each lasting 30–35 min) spread over 3 days, with at most 1 day between the sessions. During each visit, participants completed a series of exposure and test blocks that focused on noun and sentence learning. All sessions followed the same overall procedure; the number of blocks, however, differed from session to session (Fig. 2).

2.3.1. Noun exposure and tests

2.3.1.1. Noun exposure: Each experimental session began with a *noun exposure block*, in which participants were presented with pictures of characters accompanied by their names in the novel language and were instructed to repeat the names to facilitate learning.

2.3.1.2. Noun comprehension: The two-alternative forced-choice *noun comprehension block* followed. Participants heard a name of a character in the novel language accompanied by two pictures and were asked to choose the correct picture. Feedback was provided after each trial.

2.3.1.3. Noun production: Participants were shown pictures of characters one at a time and asked to provide a label for them in the novel language. Feedback on correctness was provided on each trial.

Noun exposure and comprehension blocks included 12 trials each on Day 1 and 6 trials each on Days 2–3. The noun production block included 6 trials on all days of training. On Day 1, the three blocks were repeated immediately after completion of the noun production test. The noun exposure and comprehension blocks were also presented immediately before the sentence production test. On Days 2–3, participants completed only the noun production block before the sentence production block.

2.3.2. Sentence exposure and tests

2.3.2.1. Sentence exposure: The *sentence exposure block* followed noun exposure and tests on all days of training. Participants were shown short computer-generated videos accompanied by their descriptions in the novel language and were asked to repeat the descriptions out loud to facilitate learning. The video and sound stimuli were presented simultaneously; the description was played as the action unfolded. Participants were allowed to view the action and listen to the novel description as many times as they liked during the first sentence exposure block on Day 1; replay was disabled for all other blocks during the experiment.

2.3.2.2. Sentence comprehension: On all days of training, sentence exposure was followed by a *sentence comprehension block*. On each trial, participants were shown two previously unseen videos accompanied by a sentence in the novel language. Both videos depicted the same action performed by the same characters, but the roles of the actor and patient were reversed in the two videos. Participants were asked to choose the video that best matched the sentence they heard. No feedback on correctness was provided during the test. On all days of training, participants were presented with two sets of two sentence exposure blocks and one sentence comprehension block (24 trials each).

2.3.2.3. Sentence production: Each experimental session ended with a *sentence production block* (48 trials total). Participants viewed previously unseen videos and were asked to describe them in the novel language. To facilitate production, participants were auditorily presented with a novel verb prompt. No feedback was provided during this test.

3. Results

Before turning to the central prediction of this study—differential case marker use depending on constituent order flexibility—we describe how our data were scored and discuss learners' overall performance. Then we analyze case marking preferences in participants' productions. To more fully capture how language learners are affected by the hypothesized bias to trade off robust information transmission against production effort, we provide several complementary analyses. First, we present the most common type of analysis—analyzing patterns of case use across participants. This provides a clear picture of how learners change the use of case marking in their productions compared to the input they receive. This analysis, however, does not capture the *joint* effect of constituent order and case marking in the hypothesized trade-off. There are many ways to make the input languages more communicatively efficient, and the most efficient use of case marking also depends on how each individual learner uses constituent order. For example, some learners might create more communicatively efficient languages by fixing constituent order and dropping case marking, whereas others might maintain constituent order variation and add case marking. Indeed, we observe considerable individual differences between learners in both constituent order and case marking preferences. This leads

us to suggest a unifying information-theoretic approach that will directly assess the extent to which participants trade off effort against robust information transmission through the *joint* use of constituent order and case marking.

3.1. Scoring

3.1.1. Comprehension accuracy

Participants who did not achieve 65% accuracy on the comprehension test on the final day of training were removed from the analysis. For this purpose we analyzed responses only on case marked (i.e., unambiguous) trials. This excluded 10 participants, all in the flexible order language. This is not surprising since participants were monolingual native speakers of English, a language that has no constituent order variation or case marking. This makes the flexible order language considerably harder to learn.

For the remaining 40 participants, mean comprehension accuracy was 97% across languages (99% for the fixed constituent order language and 96% for the flexible constituent order language) on the final day of training. The results reported below did not change when participants who failed to pass the 65% comprehension accuracy criterion were included in the analysis.

3.1.2. Production accuracy

For each trial, we scored constituent order used in the utterance, the presence of case marking on the object, lexical (using incorrect vocabulary), and grammatical mistakes (using a constituent order not allowed by the language or using a case marker on a constituent other than the object). On a small number of utterances, participants mispronounced the name of a referent or an action. If it was impossible to determine the constituent order used in the utterance (e.g., when both referent names were mispronounced), the production was scored as both lexically and grammatically incorrect. If it was still possible to determine the constituent order used in the utterance (e.g., the name of only one referent was incorrect), the production was coded as a lexical mistake and was scored for grammatical accuracy.

Both languages were acquired with a high degree of accuracy. On the final day of training, participants made 1.3% lexical mistakes across languages (1% in the fixed constituent order language and 1.6% in the flexible constituent order language) and 6.2% grammatical mistakes (all in the flexible constituent order language), suggesting that the task was feasible for our participants. All analyses reported below are based on grammatically correct productions only. The same results were obtained when productions containing lexical mistakes were also removed from the analysis.

3.2. Constituent order in production

Since learners in our experiment received no instruction as to which structures to use in their own productions, they could have made languages more efficient by changing the amount of constituent order flexibility allowed in the languages (e.g., by making the flexi-

ble constituent order language more fixed or the opposite) or by differentially using case marking depending on the language condition. Learners did not vary the constituent order properties of the input languages (see Fig. 3): Participants in both language conditions maintained the input constituent order proportions on all days of training.

Having established that learners acquired the meaning of case marking in our experiment and matched the constituent order distribution in the input, we now turn to the analysis of case marker preferences. We first explore the main prediction of this study—*whether* learners use case marking differentially depending on the amount of constituent order variability in the language. We then turn to a more detailed analysis of *how* case marking is used.

3.3. Case marker use in production

The central prediction of this study is that language learners are biased against excessive redundancy in linguistic systems and use additional cues to grammatical function only if the existing cues do not provide sufficient information for successful recovery of sentence meaning. Note that while we expect learners to trade off the information provided by cues to sentence meaning, we do not expect them to completely remove all redundancy from the grammatical system. That is, we hypothesize differential case marker use depending on the amount of constituent order flexibility allowed by the language, but not the categorical absence of case marking in the fixed constituent order language. Some amount of redundancy is expected in an efficient linguistic system since communication takes place in the presence of noise and thus each of the probabilistic cues to grammatical function assignment has a certain probability of being misperceived during communication.

To test this prediction, we used a mixed logit model (Breslow & Clayton, 1993; Jaeger, 2008) to regress the presence of case marking on the object onto full factorial design (all main effects and interactions) of language condition (fixed vs. flexible constituent

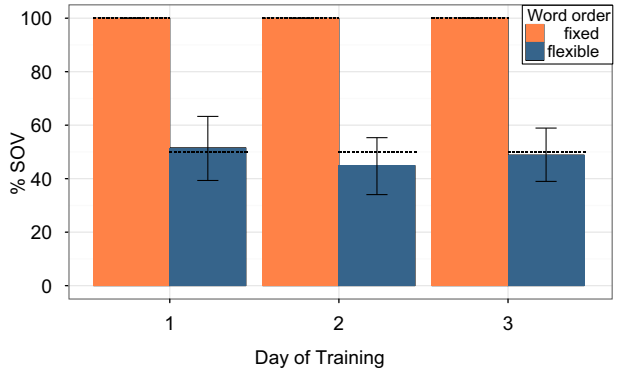

Fig. 3. Constituent order use by language condition. The error bars represent bootstrapped 95% confidence intervals. The dashed lines indicate the input proportions (different across conditions).

order) and day of training (1–3). All analyses reported below contained the maximal random effects structure justified by the data based on backward model comparison. All results also hold when the fullest converging random effects structure is used.

There was a significant main effect of language condition (see Fig. 4): Learners of the flexible constituent order language used significantly more case marking in their productions than learners of the fixed constituent order language ($\hat{\beta} = 1.45$, $z = 2.24$, $p < .05$). Language condition interacted with Day 2 ($\hat{\beta} = 0.46$, $z = 3.4$, $p < .001$) and Day 3 of training ($\hat{\beta} = 0.25$, $z = 2.75$, $p < .01$). Simple effects test revealed that the difference between the two language conditions (fixed vs. flexible constituent order) was significant on Day 2 ($\hat{\beta} = 1.65$, $z = 2.5$, $p < .05$) and Day 3 ($\hat{\beta} = 1.94$, $z = 2.72$, $p < .01$) of training. Thus, as predicted, learners of the flexible constituent order language were more likely to use case marking.

Second, we also predicted that learners of the fixed constituent order language would reduce the amount of case marking compared to the input, as the production of case consumes effort but adds no information above that already conveyed by constituent order. Indeed, learners of the fixed constituent order language used case marking in their own productions significantly below the input on all days of training, supporting our hypothesis (Day 1: 50% case marking in production, marginally lower than 67% input proportion [Wilcoxon Signed-Rank test over by-participant proportions: $V = 59$, $Z = -1.73$, $p = .08$]; Day 2: 45% case marking in production, significantly lower than the input [$V = 44$, $Z = -2.06$, $p < .05$]; Day 3: 41% case marking in production, significantly lower than the input [$V = 37$, $Z = -2.56$, $p < .05$]).

With these two basic predictions confirmed, we next investigated in more detail *how* speakers of the flexible order language used case marking.

3.4. How is case marking used in the flexible order language?

There are at least two ways in which learners could increase the robustness of information transmission in the flexible order language. First, learners can regularize case mark-

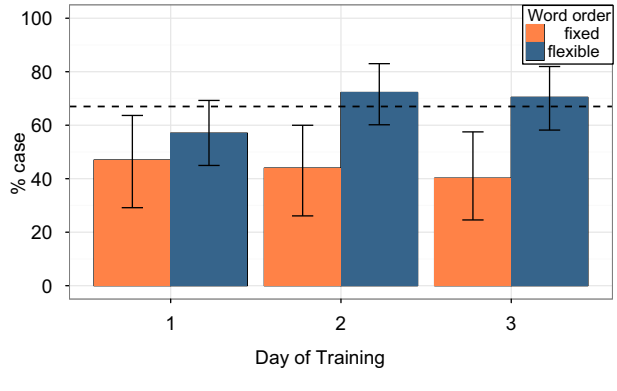

Fig. 4. Case marker use by language condition. The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the input proportion (equal across conditions).

ing in the language overall (i.e., produce it more frequently than in the input overall). This change would increase production effort compared to the input since all objects will be case-marked. Alternatively, learners could condition case marking on constituent order —for example, by always using it with one variant but never with the other (perfectly asymmetric case marking). This would result in low uncertainty about the intended meaning since the absence of case marking in this instance is highly informative of grammatical function assignment. Crucially, this latter strategy is more efficient than the former since it reduces case marking to a subset of all sentential objects. This makes asymmetric case marking a more robust system without a concomitant increase in effort compared to the input. Note that even imperfect case marking asymmetry (i.e., marking objects more often in one constituent order than the other) can reduce the uncertainty about grammatical function assignment compared to perfectly symmetric case marking. Next we explore which of these strategies learners used in our experiment.

Overall, the use of case by learners of the flexible constituent order language was the same as in the input on all days of training, suggesting that learners did not adopt the full case regularization strategy (Day 1: 55% case marking in production, not significantly different from 67% input proportion [Wilcoxon Signed-Rank test over by-participant proportions: $V = 69$, $Z = -1.34$, $p = .18$]; Day 2: 72% case marking in production, not significantly different from the input [$V = 142$, $Z = 1.38$, $p = .17$]; Day 3: 71% case marking in production, not significantly different from the input [$V = 130$, $Z = 0.93$, $p = .36$]).

Did learners prefer a more efficient strategy of conditioning case on constituent order, leading to asymmetric case marking? We regressed the presence of case marking onto constituent order (SOV/OSV), day of training (1–3), and their interaction. Learners did not use case marking uniformly across the two constituent orders, but instead produced case significantly more often in OSV sentences ($\hat{\beta} = 1.11$, $z = 17$, $p < .001$; see Fig. 5). A significantly higher proportion of object case marking in OSV sentences compared to SOV sentences was observed on every day of training (simple effects for Day 1: [$\hat{\beta} = 1.53$, $z = 12.6$, $p < .001$]; Day 2 [$\hat{\beta} = 0.9$, $z = 8.6$, $p < .001$]; Day 3 [$\hat{\beta} = 0.91$, $z = 8.23$, $p < .001$]). We also compared learners' case use to the input: They produced significantly more case marking than the input in OSV sentences and tended to match the input proportion of case marking in SOV sentences on most days of training (for OSV sentences: Day 1: 81% case marking in production, significantly higher than 67% input proportion [Wilcoxon Signed-Rank test over by-participant proportions: $V = 165$, $Z = -2.28$, $p < .05$]; Day 2: 86% case marking in production, significantly higher than the input [$V = 177$, $Z = 3.31$, $p < .001$]; Day 3: 85% case marking in production, significantly higher than the input [$V = 158$, $Z = 2.56$, $p < .05$]; for SOV sentences: Day 1: 28% case marking in production, significantly lower than 67% input proportion [$V = 33$, $Z = -2.7$, $p < .01$]; Day 2: 54% case marking in production, not significantly different from the input [$V = 60$, $Z = -1.41$, $p = .16$]; Day 3: 55% case marking in production, not significantly different from the input [$V = 69$, $Z = -1.04$, $p = .3$]).

The analyses presented so far provide evidence in support of the hypothesis that learners trade off robust information transmission against production effort. In understanding

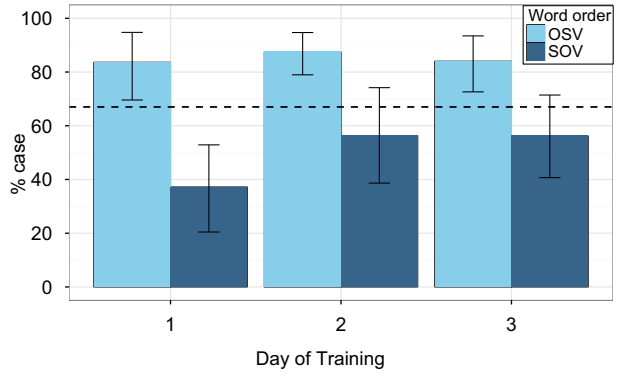

Fig. 5. Case marker use by sentence constituent order (flexible constituent order language). The error bars represent bootstrapped 95% confidence intervals. The dashed line represents the input proportion (equal across constituent orders).

how languages change over time as a function of this pressure, it is also of interest to understand individual learners' behavior. It is possible, for example, that all learners follow the same overall preference in constituent order and case marking use. Alternatively, all learners could follow the same general bias to balance effort and informativity, but do so in a variety of different ways. This would highlight the role of individual innovations in language change. Yet another possibility is that the across-participant averages we have discussed so far draw a misleading picture: It is theoretically possible that individual participants never exhibit the hypothesized trade-off.

In the next two sections, we first illustrate that a high degree of between-participant variability applies to our data, as is typical in this type of experiments (Culbertson et al., 2012; Fedzechkina et al., 2012; Hudson Kam & Newport, 2005, 2009): Learners had dramatically different case marking and constituent order preferences. We then present a novel information-theoretic perspective on the data showing that the superficially differing individual strategies are guided by a single underlying principle of trading off effort against robust information transmission.

3.5. Individual learners' preferences

There was considerable between-participant variation both in constituent order and case marking preferences. We describe these variable patterns in turn. While the majority of learners of the flexible order language matched the input distribution of constituent order (see Fig. 6), a number of learners substantially deviated from the input either in the direction of SOV or OSV. In particular, three learners adopted SOV as the dominant constituent order in their productions and five learners regularized OSV constituent order; the remaining participants tended to stay close to the input. There was no between-participant variability in the fixed constituent order language: All learners used exclusively SOV constituent order in their productions. This learning outcome is in line with prior work suggesting that learners acquire consistent input very well and are not likely to introduce

innovations into a perfectly consistent linguistic system (e.g., St Clair et al., 2009; Tily et al., 2011).

There was also considerable variation in case marker preferences among individual participants in the two languages (see Fig. 7). While the majority of learners of the fixed constituent order language produced case marking below the input (and eight learners never used case marking at all), three participants produced substantially more case marking than the input proportion. Similarly, in the flexible constituent order language, most learners tended to produce case above the input (and several learners used it in all their productions), whereas one learner never used case and five learners produced case marking below the input.

3.6. A unifying information-theoretic perspective on individual differences

Individual preferences discussed above provide further support for the hypothesis that learners trade off production effort against robust information transmission. While the majority of learners of the fixed constituent order language produce case below the input proportion, learners of the flexible order language produce case above the input. These data, however, do not show a complete picture since communicative efficiency of the acquired languages ultimately depends on learners' *joint* case and constituent order preferences. For example, the efficiency of case marker regularization in the flexible constituent order language would vary depending on learners' constituent order preferences: While it is efficient to maintain or regularize case in the presence of constituent order variability, it would not be an efficient innovation if constituent order variability is eliminated.

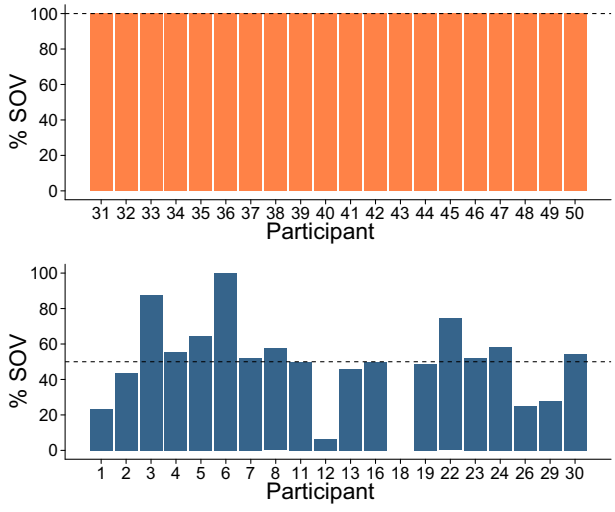

Fig. 6. Constituent order preferences of individual learners on the final (third) day of training in the fixed (top panel) and flexible (bottom panel) constituent order languages. The dashed lines indicate the input proportion.

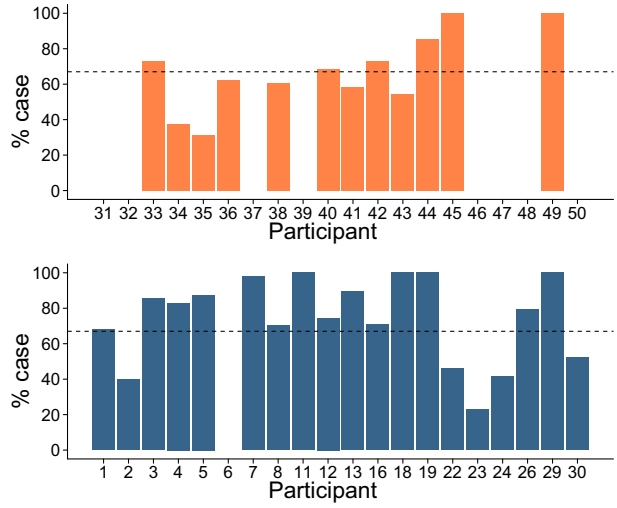

Fig. 7. Case marker preferences of individual learners on the final (third) day of training in the fixed (top panel) and flexible (bottom panel) constituent order languages. The dashed lines indicate the input proportion.

To address learners' joint case marker and constituent order preferences, we introduce a new perspective on our data that lets us directly assess whether learners strive to balance effort and uncertainty about the intended meaning, independent of the *means* by which they achieve it (i.e., independent of whether they, for example, fix constituent order, regularize case marking, or condition case marking on constituent order). For this purpose, we estimated production effort and the amount of uncertainty about grammatical function assignment for each language acquired by individual participants.

We formalize production effort as the average number of syllables per sentence.[4] As shown in Fig. 8, the two input languages in our experiment have the same amount of effort (ranging from 3 to 7 syllables per sentence, with a mean of 5.15) since they use the same artificial lexicon and have the same amount of case marking in the input.

The average uncertainty about grammatical function assignment experienced by the listener who has a perfect knowledge of the grammar used by the speaker was captured as conditional entropy over grammatical function assignments (GF):

$$H(\text{GF}|\text{sent.form}) = -\sum_{\text{GFs}} \sum_{\text{sent.forms}} \text{p}(\text{sent.form},\text{GF}) * \log_2 \text{p}(\text{GF}|\text{sent.form})$$

where the sum is over grammatical function assignments (subject-object, object-subject) and three possible sentence forms in the languages (NP1-no case NP2-case; NP1-case NP2-no case; NP1-no case NP2-no case).

In our languages, the conditional entropy of grammatical function assignment was 0 bits for sentences with case-marked objects and for all sentences if there was no constituent order variation in the language.[5] Thus, the fixed order input language had mini-

mal weighted conditional entropy of 0 bits as it contained no constituent order variation. In the flexible order input language, 67% of sentences were case-marked and thus resulted in the conditional entropy of 0 bits. Also, 33% of input sentences were not case-marked and thus resulted in the conditional entropy of 1 bit, since SOV and OSV constituent order variants occurred equally often in the input. Thus, the average weighted conditional entropy of the flexible constituent order language was 0.33 bits (1*0.33 + 0*0.67).

As expected under our hypothesis, learners systematically deviated from the input toward languages that balanced effort and uncertainty about sentence meaning more efficiently (see Fig. 8). As predicted, by dropping redundant case marking in their productions, learners of the fixed constituent order language converged on a language that had lower average effort compared to the input. The output language produced by learners of the flexible constituent order language tended to have lower weighted conditional entropy of grammatical function assignment compared to the input without a concomitant increase in effort. This behavior highlights the benefit of asymmetric case marking: For the same overall effort, entropy reduction is larger for systems that have asymmetric case marking compared to systems with symmetric case marking (i.e., case marking independent of constituent order) for the same overall constituent order proportion (see Fig. 8).[6]

3.6.1. Between-participant variability revisited from an information-theoretic perspective

The analysis of individual participants' productions further highlights the benefits of understanding learners' behavior as a function of effort and uncertainty reduction. As Fig. 9 shows, the majority of learners of the fixed constituent order language followed

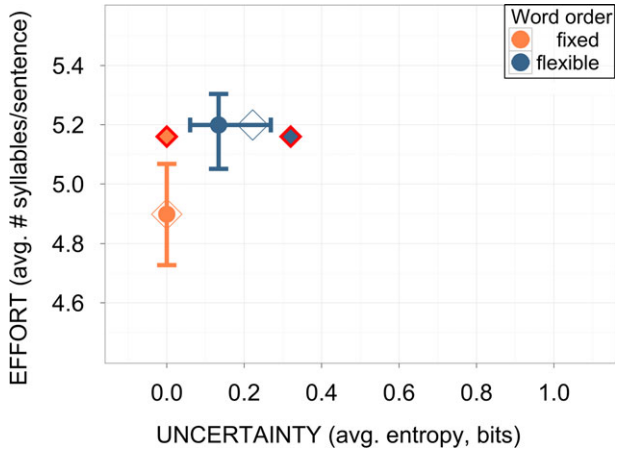

Fig. 8. Uncertainty versus production effort trade-off on the final (third) day of training. Diamonds with red borders represent input languages. Solid circles represent mean output languages produced by participants. Open diamonds represent what the output languages produced by participants would look like if they contained the same overall proportions of constituent order and case marking as the actual output, but used them independently (i.e., did not condition case marking on constituent order). The error bars represent bootstrapped 95% confidence intervals.

our prediction and reduced effort without increasing uncertainty. All learners of this language produced languages that had zero conditional entropy of grammatical function assignment, and 14 (of 20) learners reduced effort compared to the input.

In the flexible constituent order language, there was an overall strong bias to reduce uncertainty: 12 of 20 participants produced languages that had zero conditional entropy of grammatical function assignment (see Fig. 9) and 5 of the remaining 8 learners produced languages with lower uncertainty than in the input. Only 3 participants increased conditional entropy of grammatical function assignment. However, all of them did so while decreasing effort.

Overall, the preference to reduce uncertainty in grammatical function assignment was stronger than a preference to reduce effort: Effort increases were acceptable as long as conditional entropy of grammatical function assignment was reduced. However, if uncertainty about grammatical function assignment was increased, there was a clear preference to reduce effort.

In summary, instead of veridically reproducing the input language, which did not trade off effort against robust information transmission efficiently, participants in our experiment introduced innovations into the acquired language that increased its efficiency. Learners did not differ in *whether* they followed this bias; they only differed in *how they weighted* the two opposing pressures. This suggests that, despite the fact that learners employ vastly differing strategies, their learning outcomes are guided by a deeper abstract principle of a trade-off between robust information transmission and effort.

4. Discussion

Our findings add to the growing body of research showing that learners preferentially acquire and regularize typologically frequent patterns (Christiansen, 2000; Culbertson et al., 2012; Fedzechkina et al., 2012; Finley & Badecker, 2008; Morgan et al., 1987; Newport & Aslin, 2004). Learning outcomes in our experiment parallel synchronic and diachronic cross-linguistic phenomena, thereby providing converging evidence that these patterns are not due to chance.

Even though learners of both languages were exposed to the same amount of case marking in the input, we observed differential learning outcomes depending on the language condition that followed the typological correlation between constituent order flexibility and case marking. Learners of the fixed order language tended to drop case marking in their productions, paralleling the lack of case systems in fixed constituent order languages like English or French. Learners of the flexible order language, however, retained case marking in their productions, mirroring flexible constituent order languages like Latin or Russian that typically have a case system.

Learning outcomes in our experiment also parallel diachronic typological patterns, such as, the change from Old English (a language with flexible constituent order and a rich case system) to Modern English (a language with fixed constituent order and no case system). Whether the historical loss of case marking precipitated (Marchand, 1951; Sapir,

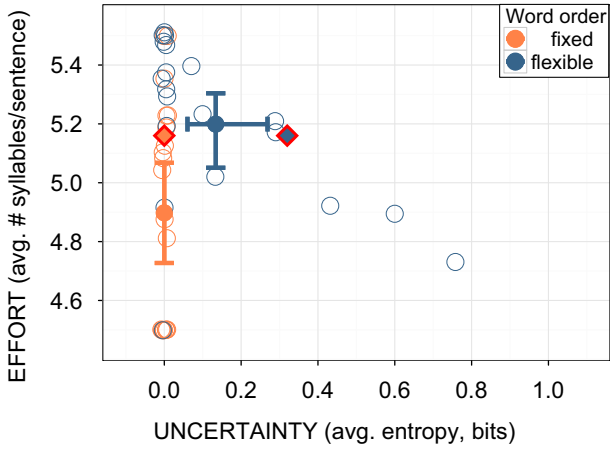

Fig. 9. Uncertainty versus production effort trade-off on the final (third) day of training. Diamonds with red borders represent input languages. Solid circles represent mean output languages produced by participants. Open circles represent language outputs produced by individual learners. The error bars represent bootstrapped 95% confidence intervals.

1921) or followed constituent order fixing (Lehnert, 1957) is a matter of debate, and our current results are unable to tease apart these two alternatives. Under our hypothesis, however, both paths should result in the same outcome: Learners of a flexible order language should maintain constituent order flexibility if this language has a case system (cf. Old English) and gradually lose it if this language has no case system (cf. Modern English).

Interestingly, learners of the flexible order language did not fix or even regularize constituent order. One reason for this might be that changes to constituent order result in larger perceptual deviance from the input distributions, but future work is necessary to address this question.

The overall picture emerging from individual learners' performance suggests that the preference to balance robust information transmission and effort is indeed in play during language acquisition (Aylett & Turk, 2004; Gibson et al., 2013; Jaeger, 2010; Jäger, 2007; Kurumada & Jaeger, 2015; see also Lindblom, 1990). Our findings provide insight into linguistic diversity: For any grammatical system, the general principle of trading off robust information transmission and effort can manifest itself in a variety of ways. Which of them would survive at the population level and eventually become part of the grammar depends on a variety of factors, including the inherent efficiency of the innovation, other properties of the linguistic system, historical factors such as language contact (Heine, 2008), and social factors such as social influences of a person creating the innovation (Nettle, 1999).

In the remainder of the article, we first discuss some of our findings in more detail, and in particular learners' preference to use more case marking in OSV sentences. We then address outstanding questions for future work. Specifically, we situate our find-

ings in a broader literature on deviations from the input in adult artificial language learning and consider possible mechanisms underlying the learning outcomes observed in our experiment. We conclude with a discussion of how the logic of the arguments presented in this paper can be extended to a broader range of cues to grammatical function assignment, beyond the interaction between case marking and constituent order.

4.1. Conditioned variability

Learners of the flexible order language chose to condition case marking on constituent order rather than to regularize case marking in the language to 100%. These results corroborate prior work showing that adult learners prefer linguistic systems that contain conditioned variability over systems with no variability (Smith & Wonnacott, 2010; but see Hudson Kam & Newport, 2005, 2009, for the opposite findings in children). This learning outcome conceptually replicates our previous findings with different miniature languages (Fedzechkina et al., 2012), where learners conditioned case marking on the arguments' animacy. Learning behavior in our experiment points to a preference for another cross-linguistically frequent property of human language—a tendency to avoid free variation in which several competing forms carry the same meaning in similar contexts. Instead, languages tend to have predictable variation by conditioning the use of competing forms on semantic, pragmatic, phonological, and other factors (Givón, 1985).

Conditioned variability in our experiment increases communicative success without a concomitant increase in production effort. In a language that conditions case marking on constituent order, not all sentential objects need to be marked but sentence meaning can be reliably inferred when case markers are not present. While the same increase in communicative success would be gained by conditioning case marking on either SOV or OSV order, learners in our experiment consistently used case more often in OSV sentences. What can account for this preference?

It is possible that this behavior is associated with OSV being a non-canonical constituent order. Even though OSV and SOV constituent orders occurred equally frequently in the input and in participants' productions, OSV constituent order is typologically rare and is uncommon in English. The increase in case marking in OSV sentences could thus be a consequence of increased utterance planning difficulty for non-canonical structures (Ferreira, 1994; Urosevic, Carello, Savic, Lukatela, & Turvey, 1981). Alternatively, producers might prefer to mark structures that are less expected for the comprehender (in this case, less expected cross-linguistically or in the native language), thereby facilitating comprehension (cf. Aissen, 2003; Bybee & Hopper, 2001; Haiman, 1983; Jaeger, 2013; Jäger, 2007). Finally, the preference to use more case in OSV sentences might reflect a bias to maximize the number of linguistic dependencies processed at each point in time (cf. "Maximize On-line Processing" principle that is observed both in natural language production and typological distributions; Hawkins, 1994, 2004; Nichols, 1986). Because our input languages only have object case

marking, case would provide the earliest point of disambiguation in OSV sentences and allow for correct parsing commitments early on. Thus, it is possible that learners preferentially restructured the input language in such a way as to put more informative cues earlier in the sentence. This interpretation of our results parallels the developmental literature showing that young learners are highly sensitive to the order of cues to sentence meaning and disproportionally rely on early arriving cues when making parsing decisions (Choi & Trueswell, 2010; Snedeker & Trueswell, 2004; Trueswell, Sekerina, Hill, & Logrip, 1999). Similarly, early arriving cues also appear to be acquired more easily and accurately both in first language acquisition (Trueswell, Kaufman, Hafri, & Lidz, 2012) and in laboratory miniature language learning experiments (Pozzan & Trueswell, in press). The three explanations outlined above are mutually compatible, and future work is necessary to assess their validity.

4.2. When are deviations from the input expected?

Our results add to the growing body of research investigating the circumstances under which learners are likely to deviate from the input they receive. Adult learners in our experiment did not veridically reproduce the inefficient variation in the input, but tended to consistently deviate from the input toward more efficient linguistic systems. In fact, 60% of learners of the flexible constituent order language converged on output languages that had zero conditional entropy about grammatical function assignment. This behavior is atypical given that prior work in miniature artificial language learning has found that adult learners generally match the input they receive (Hudson Kam & Newport, 2005, 2009; Tily et al., 2011). These studies, however, differ from present work in several important respects.

First, a large body of work suggests that learners are not likely to introduce innovations into perfectly consistent languages (traditionally used to study the relationship between language learning and language structure), at least not within the short amount of time spent in the laboratory (e.g., Christiansen, 2000; St Clair et al., 2009; Tily et al., 2011). Unlike these studies, our input languages incorporate variability, somewhat reminiscent of the situation of a pidgin, to make learning biases manifest within the short period of time available in an experiment.

Studies using the paradigm similar to ours, in which learners are exposed to languages where several grammatical structures are available to express the same meaning, have produced somewhat mixed results with regard to adult learners' behavior. Adult learners tend to match the statistics of the miniature artificial grammar in experiments of this type both in linguistic (Hudson Kam & Newport, 2005, 2009) and non-linguistic (Ferdinand, Thompson, Kirby, & Smith, 2013) domains. For example, Hudson Kam and Newport (2005) presented adults and children with miniature artificial languages that had fixed constituent order and unpredictable morphological variation: Nouns were followed by determiners ("ka" and "po") that varied probabilistically in the input. While young learners readily regularized the inconsistent input they received and used one of the determiners in almost all of their productions, adults

mostly reproduced the unpredictable variation present in the input. This bias against deviating from the input can, however, be overcome in adult learners if input languages are fairly complex (e.g., languages in which learners need to keep track of multiple frequencies simultaneously), while being still learnable within a short period of time (Ferdinand et al., 2013; Hudson Kam & Newport, 2009).

Crucially, the miniature artificial languages used in our experiment were different from those used in most previous work in that the design of our experiment provides a *functional motivation* to deviate from the input. Our work thus adds to the growing body of literature suggesting that in the presence of such motivation adult learners tend to deviate from the input provided that the input languages are carefully designed to be complex enough to observe such deviations (Culbertson & Legendre, 2010; Culbertson et al., 2012; Fedzechkina et al., 2012; Hudson Kam & Newport, 2009).

4.3. The mechanism underlying learning outcomes

Our findings raise questions about the precise nature of the mechanism underlying learners' preferences to efficiently trade off cues to grammatical function assignment.

The learning outcomes in our experiment closely resemble the patterns found in online productions of adult speakers (Aylett & Turk, 2006; Frank & Jaeger, 2008; Gomez Gallo, Jaeger, & Smyth, 2008; Jaeger, 2010; Kurumada & Jaeger, 2015; Lindblom, 1990; van Son & Pols, 2003; van Son & van Santen, 2005). Particularly relevant to the work presented here is a recent study by Kurumada and Jaeger (2015). They found that speakers of Japanese were more likely to produce object case marking when sentence properties (such as animacy of the object or plausibility of grammatical function assignment) were likely to bias the listener away from the intended interpretation. This suggests that learners' preferences in our experiments could originate in the human production system (see also Norcliffe & Jaeger, 2015, for similar evidence from optional head-marking in Yucatec Maya; for a cross-linguistic review, see Jaeger & Buz, in press).

It is also possible that the patterns observed in our experiments originate in learning that is not specific to production: Learners may have misinterpreted some of the sentences they were exposed to, altering the characteristics of the input from which they learned. Where might such "misinterpretations" arise? One possibility is that learners might have misinterpreted some of the sentences during initial perception (cf. Guy, 1996; Ohala, 1989). A promising account along these lines is presented by simulations of case marking and constituent order interactions (Lupyan & Christiansen, 2002; Van Everbroeck, 2003). For example, in Van Everbroeck's simulations, the model was trained to determine grammatical function of each sequentially presented word in a sentence. When tested on previously unseen non-case-marked sentences, the model performed well on most fixed constituent orders, but showed somewhat poorer performance at discriminating constituents in fixed SOV order. For flexible constituent order languages, however, the model could not reliably distinguish between constituents in the absence of other cues to grammatical function assignment (e.g., case). Adding case marking dramatically improved network performance. Thus, the perceptual account can explain several cross-linguistic universals: the trade-off

between case and constituent order (Blake, 2001; Sapir, 1921) as well as the prevalence of case marking in verb-final languages (Dryer & Haspelmath, 2011; Greenberg, 1963).

How likely is such an account to explain our results? On the one hand, the rarity of object-before-subject sequences in English (the native language of our learners) is likely to further increase error rates for OSV sentences (Christianson, Hollingwoth, Halliwell, & Ferreira, 2001; Ferreira, 2003). On the other hand, two properties of our experimental design make an account that grounds misinterpretations in perception somewhat implausible. First, unlike the connectionist simulations described above, the meaning of the training sentences was always illustrated by the accompanying video in our experiment, thereby unambiguously conveying the intended grammatical function assignment. Recall that sentence descriptions were played as the action unfolded, thus making it less probable that learners have formed incorrect interpretations of sentence meaning based on the novel descriptions alone. Second, there was no time pressure. This makes it somewhat unlikely that misinterpretations during perception were sufficiently frequent to create the observed effect.

It is, however, possible that "misinterpretations" arise later during memory encoding or retrieval. A recurring finding in verbal and non-verbal short-term memory is that items interfere with each other during encoding and retrieval (Jonides et al., 2008). This interference is increased if the items that need to be retrieved or encoded are in some way similar to each other (e.g., structurally, semantically, spatially, etc.). This type of interference has been shown to affect sentence comprehension as well (Lewis, 1996; Lewis & Nakayama, 2001; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & Lewis, 2003). Increasing syntactic or semantic similarity between the target and preceding or following material poses additional difficulty during retrieval (as manifested in longer reading times and increased comprehension errors). Since all visual referents in our experiment were human and male, it is likely that participants experienced this type of interference and thus misassigned grammatical roles in a certain number of trials. Such misassignments could have introduced changes into the input distribution when it was consolidated in memory. We note, however, that any explanation along these lines would need to be extended further to also capture the findings from optional case marking in both miniature and natural languages (Fedzechkina et al., 2012; Kurumada & Jaeger, 2015; Lee, 2006), whereas these findings follow straightforwardly under the hypothesis that learners are biased toward communicatively efficient languages.

The hypothesis that the outcomes observed in our experiment arise outside of the production system (whether in perception or memory) is broadly compatible with functional theories of efficient communication. It complements work that has focused on the organization of human production system as the locus of biases for efficient communication (e.g., Aylett & Turk, 2004; Jaeger, 2013; Kurumada & Jaeger, 2015; Lindblom, 1990).

4.4. The scope of explanation

While the main focus of this study is on the interaction between constituent order and case marking, the arguments presented here can be extended beyond these particular cues

to grammatical function assignment. This view can be summarized as a more general implicational universal concerning the trade-off between cues to grammatical function assignment proposed in the functional literature (Bates & MacWhinney, 1982, 1989; Givón, 1991), which is also broadly compatible with information-theoretic accounts (e.g., Gibson et al., 2013; Jaeger, 2006, 2013; Levy, 2005). If an existing cue to grammatical function assignment is highly informative, other cues would be redundant and thus could be omitted; if existing cues do not contain enough information to successfully communicate who is doing what to whom in a sentence, additional cues should be recruited to maintain successful communication.

In this broad interpretation, this view captures the avoidance of excessive redundancy in encoding grammatical function assignment cross-linguistically: While natural languages employ a variety of strategies to identify the subject and the object of the sentence (e.g., constituent order, case marking, agreement, prosody, animacy, etc.), it is highly uncommon to consistently employ all attested means of grammatical function assignment within the same language. Our account also makes explicit predictions for languages that go against the correlation between constituent order flexibility and the presence of case, such as some highly flexible constituent order languages that have no case marking (e.g., Abkhaz, Papago). Such languages are expected to have cues other than case marking to encode grammatical function assignment. Indeed, natural languages typically preserve the distinction between the subject and object in the sentence (Wasow, 2015) and often recruit such cues as animacy (Prat-Sala & Branigan, 2000), agreement (Siewierska, 1998), or pragmatic information (Mithun, 1987) when constituent order is flexible.

Note that while we argue against excessive unnecessary redundancy in linguistic systems, some amount of redundancy is expected under our account (for a related discussion, see also Piantadosi, Tily, and Gibson 2011a). Human communication takes place in the presence of noise. Thus, listeners need to infer the speaker's intended message from multiple probabilistic cues, each of which can be corrupted during its transmission via the noisy channel with a certain probability. For example, case markers and constituent order might not be successfully perceived every time—they can be lost or misheard. This has consequences for when case marking would be expected in a language. Even when constituent order is highly informative, case marking will often further increase robust information transmission (especially in those cases when the message is less expected). This might explain why fixed constituent order languages sometimes have case and agreement systems (Siewierska, 1998) or why many languages with optional case marking (e.g., Japanese, Korean, Hindi) will optionally mark both grammatical subjects and objects (for a review, see Malchukov, 2008).

5. Conclusion

To summarize, our findings suggest that the well-documented inverse correlation between constituent order flexibility and the presence of a case system can be explained by learners' preference for grammatical systems that encode linguistic information effi-

ciently. More generally, our results provide additional support for the hypothesis that at least some cross-lexical and grammatical properties of languages represent efficient trade-offs between effort and robust information transmission (Maurits, Perfors, & Navarro, 2010; Piantadosi, Tily, & Gibson, 2011b; Qian & Jaeger, 2012). Our results also contribute to a growing body of work demonstrating the potential of using miniature artificial language learning to study the relationship between learning biases and language structures.

Acknowledgments

We thank T. Stanley and S. Krishnan for their help with data collection and data coding; C. Anzalone, G. Smith, E. Williams, and B. Chu for their help with data coding; C. Donatelli, I. Minkina and A. Wood for their help with video stimuli; L. Emberson, P. Reeder, and J. P. Blevins for valuable feedback on the manuscript. This work was supported in part by NSF grant IIS-1150028, as well as an Alfred P. Sloan Fellowship BR2011-092 to T. F. Jaeger, and by NIH grant DC00167 and HD037082 to E. L. Newport.

Notes

1. Throughout this paper, we use the term "language universals" to refer to cross-linguistic generalizations regardless of whether they are statistical or absolute.
2. We focus exclusively on one function of case and constituent order—encoding of grammatical function assignment—even though both cues typically convey additional information in natural languages (Du Bois, 1987). An example is Japanese, in which some case markers are associated with focus, topic, or discourse-marking (Shimojo, 2006; Yatanabe, 1999), or Russian, where constituent order can express pragmatic information (Yokoyama, 1986).
3. This inverse relation is by no means absolute. For example, though rare, there are some languages that arguably have fixed constituent order *and* case marking such as Icelandic (Kiparsky, 1997). The existence of exceptions does not, however, argue against the existence of statistical tendencies. There is also a substantial number of languages with flexible constituent order that employ means other than case to mark grammatical relations (e.g., head-marking in many Balkan languages; Friedman, 2006). This is compatible with the perspective we explore here, which predicts a trade-off between different means of encoding grammatical function assignment. We return to this point in Section 4.4.
4. Our choice of complexity measure (e.g., syllable count vs. word count) does not affect the results presented here since they are perfectly correlated in our language. More generally, measures of complexity are typically highly correlated in natural language as well (Szmrecsanyi, 2004).

5. We make a simplifying assumption of noise-free recognition of word sequences and case marking. However, the perspective assumed here is readily extendable to a more plausible assumption of noisy acoustic input and noisy recognition (e.g., Kleinschmidt & Jaeger, 2015; Levy, 2011; Levy, Bicknell, Slattery, & Rayner, 2009; Norris & McQueen, 2008).
6. See Chapter 2 in Fedzechkina (2014) for data from a language with flexible but informative constituent order (75/25% SOV/OSV), which provides further support for the conclusions drawn here.

References

Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language & Linguistic Theory*, *21*(3), 435–483. doi:10.1023/A:1024109008573.

Aylett, M. P., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56. doi:10.1177/00238309040470010201.

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*, 3048. doi:10.1121/1.2188331.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). Cambridge, UK: Cambridge University Press.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing* (pp. 3–73). New York: Cambridge University Press.

Baugh, A., & Cable, T. (1993). *A history of the English language* (4th ed). London: Routledge.

Bever, T. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–362). New York: John Wiley & Sons.

Blake, B. J. (2001). *Case*. Cambridge, UK: Cambridge University Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association*, *88*, 9–25. doi:10.1080/01621459.1993.10594284.

Bybee, J. L. (2007). *Frequency of use and the organization of language*. New York: Oxford University Press.

Bybee, J. L., & Hopper, P. (2001). Introduction to frequency and the emergence of linguistic structure. In J. L. Bybee & P. Hoppers (Eds.), *Frequency and the emergence of linguistic structure* (pp. 1–24). Amsterdam: John Benjamins.

Choi, Y., & Trueswell, J. (2010). Children's (in)ability to recover from garden paths in a verb-final language: Evidence for developing control in sentence processing. *Journal of Experimental Child Psychology*, *106* (1), 41–61. doi:10.1016/j.jecp.2010.01.003.

Chomsky, N. (1965). *Aspects of the theory of syntax (Vol. 119)*: Cambridge, MA: The MIT Press.

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

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral Brain Science*, *31*(5), 489–509. doi:10.1017/S0140525X08004998.

Christianson, K., Hollingwoth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic-roles assigned along the garden path linger. *Cognitive Psychology*, *42*, 368–407. doi:doi:10.1006/cogp.2001.0752.

Croft, W., Bhattacharya, T., Kleinschmidt, D., Smith, D. E., & Jaeger, T. F. (2011). Greenbergian universals, diachrony and statistical analyses. *Linguistic Typology*, *15*(2), 433–453. doi:10.1515/lity.2011.029.

Culbertson, J., & Legendre, G. (2010). Investigating the evolution of agreement systems using an artificial language learning paradigm. In D. Bailey & V. Teliga (Eds.), *Proceedings of the 39th Western conference on linguistics* (pp. 46–58). Fresno: California State University, Fresno.

Culbertson, J., & Newport, E. (2015). Harmonic biases in child learners: In support of language universals. *Cognition*, *139*, 71–82. doi:10.1016/j.cognition.2015.02.007.

Culbertson, J., Smolensky, P., & Legendre, G. (2012). Learning biases predict a word order universal. *Cognition*, *122*, 306–329. doi:10.1016/j.cognition.2011.10.017.

Cysouw, M. (2010). On the probability distribution of typological frequencies. In C. Ebert, G. Jäger, & J. Michaelis (Eds.), *The mathematics of language* (pp. 29–35). Berlin: Springer.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York: WW Norton & Company.

Detges, U. (2009). How useful is case morphology? The loss of the Old French two-case system within a theory of preferred argument structure. In J. Barðdal & S. Chelliah (Eds.), *The role of semantic, pragmatic, and discourse factors in the development of case*. Amsterdam: John Benjamins.

Dryer, M. (1989). Large linguistic areas and language sampling. *Studies in Language*, *13*, 257–292. doi:10.1075/sl.13.2.03dry.

Dryer, M., & Haspelmath, M. (Eds.). (2011). *The world atlas of language structures online*. Munich: Max Planck Digital Library.

Du Bois, J. W. (1987). The discourse basis of ergativity. *Language*, *63*(4), 805–855. doi:10.2307/415719.

Dunn, M., Greenhill, S. J., Levinson, S. C., & Gray, R. D. (2011). Evolved structure of language shows lineage-specific trends in word-order universals. *Nature*, *473*(7345), 79–82. doi:10.1038/nature09923.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–492. doi:10.1017/S0140525X0999094X.

Fedzechkina, M. (2014). Communicative efficiency, language learning, and language universals. PhD thesis, University of Rochester.

Fedzechkina, M., Jaeger, T., & Newport, E. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(44), 17897–17902. doi:10.1073/pnas.1215776109.

Fedzechkina, M., Newport, E., & Jaeger, T. (in press). Miniature artificial language learning as a complement to typological data. In A. Tyler, L. Ortega, & M. Uno (Eds.), *The usage-based study of language learning and multilingualism*. Washington, DC: Georgetown University Press.

Ferdinand, V., Thompson, B., Kirby, S., & Smith, K. (2013). Regularization behavior in a non-linguistic domain. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, 436–441.

Ferreira, F. (1994). Choice of passive voice is affected by verb type animacy. *Journal of Memory and Language*, *33*, 715–736. doi:10.1006/jmla.1994.1034.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive Psychology*, *47*, 164–203. doi:10.1016/S0010-0285(03)00005-7.

Finley, S., & Badecker, W. (2008). Substantive biases for vowel harmony languages. In N. Abner & J. Bishop (Eds.), *27th west coast conference on formal linguistics* (pp. 168–176). Somerville, MA: Cascadilla Press.

Fodor, J. D. (2001). Setting syntactic parameters. In M. Baltin & C. Collins (Eds.), *The handbook of contemporary syntactic theory*. 730–767. Oxford: Blackwell.

Frank, A., & Jaeger, T. (2008). Speaking Rationally: Uniform information density as an optimal strategy for language production. Paper presented at the the 30th Annual Meeting of the Cognitive Science Society, Austin, TX.

Friedman, V. (2006). Balkans as a linguistic area. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd ed). Vol. 1 (pp. 657–672). Oxford, UK: Elsevier.

Fry, J. (2003). Ellipsis and wa-marking in Japanese conversation. Ph.D. dissertation, Stanford University.

Gibson, E., Piantadosi, S., Brink, K., Bergen, L., Lim, E., & Saxe, R. (2013). A noisy-channel account of cross-linguistic word order variation. *Psychological Science*, *4*(7), 1079–1088. doi:10.1177/0956797612463705.

Givón, T. (1985). Function, structure, and language acquisition. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition*. Vol. *2* (pp. 1005–1028). Hillsdale, NJ: Erlbaum.

Givón, T. (1991). Markedness in grammar: Distributional, communicative and cognitive correlates of syntactic structure. *Studies in Language*, *15*(2), 335–370. doi:10.1075/sl.15.2.05giv.

Gomez Gallo, C., Jaeger, T., & Smyth, R. (2008). Incremental syntactic planning across clauses. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci08)* (pp. 845–850).

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

Guy, R. (1996). Form and function in linguistic variation. In R. Guy, D. Feagin, D. Schiffrin, & J. Baugh (Eds.), *Towards a social science of language: Papers in honor of William Labov* (pp. 221–252). Amsterdam/Philadelphia: John Benjamins.

Haiman, J. (1983). Iconic and economic motivation. *Language*, *56*, 515–540. doi:10.2307/413373.

Haspelmath, M. (1999). Optimality and diachronic adaptation. *Zeitschrift für Sprachwissenschaft*, *18*(2), 180–205. doi:10.1515/zfsw.1999.18.2.180.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, UK: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammar*. Oxford, UK: Oxford University Press.

Heine, B. (2008). Contact-induced word order change without word order change. In P. Siemund & N. Kintana (Eds.), *Language contact and contact languages*. Amsterdam: Benjamins.

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*(2), 151–195. doi:10.1080/15475441.2005.9684215.

Hudson Kam, C., & Newport, E. (2009). Getting it right by getting it wrong: When learners change languages. *Cognitive Psychology*, *59*(1), 30–66.

Hupp, J. M., Sloutsky, V. M., & Culicover, P. W. (2009). Evidence for a domain-general mechanism underlying the suffixation preference in language. *Language and Cognitive Processes*, *24*(6), 876–909. doi:10.1080/01690960902719267.

Jaeger, T. (2006). Redundancy and syntactic reduction in spontaneous speech. PhD dissertation, Stanford University.

Jaeger, T. (2008). Categorical data analysis: away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446. doi:10.1016/j.jml.2007.11.007.

Jaeger, T. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, *61*, 23–62. doi:10.1016/j.cogpsych.2010.02.002.

Jaeger, T. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, *4*(230). doi:10.3389/fpsyg.2013.00230.

Jaeger, T., & Buz, E. (in press). Signal reduction and linguistic encoding. In E. M. Fernandez & H. S. Cairns (Eds.), *Handbook of psycholinguistics*. Chichester, UK: Wiley-Blackwell.

Jaeger, T., Graff, P., Croft, W., & Pontillo, D. (2011). Mixed effects models for genetic and areal dependencies in linguistic typology: Commentary on Atkinson. *Linguistic Typology*, *15*(2), 281–319. doi:10.1515/lity.2011.021.

Jaeger, T., & Tily, H. (2011). On language "utility": Processing complexity and communicative efficiency. *Wiley Interdisciplinary Reviews. Cognitive Science*, *2*(3), 323–335. doi:10.1002/wcs.126.

Jäger, G. (2007). Evolutionary game theory and typology: A case study. *Language*, *83*(1), 74–109. doi:10.1353/lan.2007.0020.

Jonides, J., Lewis, R., Nee, D., Lustig, C., Berman, M., & Moore, K. (2008). The mind and brain of short-term memory. *Annual Review of Psychology*, *59*(193–224), doi:10.1146/annurev.psych.59.103006.093615.

Kiparsky, P. (1997). The rise of positional licensing. In A. van Kemenade & N. Vincent (Eds.), *Parameters of morphosyntactic change* (pp. 460–494). Cambridge, UK: Cambridge University Press.

Kleinschmidt, D., & Jaeger, T. F. (2015). Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. *Psychological Review*, *122*(2), 148–203. doi:10.3758/s13423-015-0943-z.

Kurumada, C., & Jaeger, T. (2015). Communicative efficiency in language production: Optional case-marking in Japanese. *Journal of Memory and Language*, *83*, 152–178. doi:10.1016/j.jml.2015.03.003.

Lee, H. (2006). Parallel optimization in case systems: Evidence from case ellipsis in Korean. *Journal of East Asian Linguistics*, *15*, 69–96. doi:10.1007/s10831-005-3004-1.

Lehnert, M. (1957). The interrelation between form and function in the development of the English language. *Zeitschrift für Anglistik und Amerikanistik*, *5*, 43–56.

Levy, R. (2005). *Probabilistic models of word order and syntactic discontinuity*. Ph.D. dissertation, Stanford University.

Levy, R. (2011). Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results. In D. Lin (Ed.), *Proceedings of the 49th annual meeting of the Association for Computational Linguistics: Human language technologies* (pp. 1055–1065). Stroudsburg, PA: Association for Computational Linguistics.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086–21090. doi:10.1073/pnas.0907664106.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25*(1), 93–115. doi:10.1007/BF01708421.

Lewis, R. L., & Nakayama, M. (2001). Syntactic and positional similarity effects in the processing of Japanese embeddings. In M. Nakayama (Ed.), *Sentence processing in East Asian languages* (pp. 85–113). Stanford, CA: CLSI Publications.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science*, *10*(10), 447–454. doi:10.1016/j.tics.2006.08.007.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. *Speech Production and Speech Modelling*, *55*, 403–439. doi:10.1007/978-94-009-2037-8_16.

Lupyan, G., & Christiansen, M. H. (2002). *Case, word order, and language learnability: Insights from connectionist modeling*. Paper presented at the 24th Annual Conference of the Cognitive Science Society.

Malchukov, A. L. (2008). Animacy and asymmetries in differential case marking. *Lingua*, *118*, 203–221. doi:10.1016/j.lingua.2007.02.005.

Marchand, H. (1951). The syntactical change from inflectional to word order system and some effects of this change on the relation verb-object in English. *Anglia*, *70*, 70–89.

Maurits, L., Perfors, A., & Navarro, D. (2010). Why are some word orders more common than others? A uniform information density account. In J. Lafferty, C. Williams, J. Shawe-Taylor, R. Zemel, & A. Culotta (Eds.), *Advances in neural information processing systems 23* (pp. 1585–1593). Vancouver: British Columbia.

Mithun, M. (1987). Is basic word order universal? In R. S. Tomlin (Ed.), *Coherence and grounding in discourse*. Amsterdam: John Benjamins.

Morgan, J. L., Meier, R. P., & Newport, E. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology*, *19*(4), 498–550. doi:10.1016/0010-0285(87)90017-x.

Nettle, D. (1999). Using social impact theory to simulate language change. *Lingua*, *108*, 95–117. doi:10.1016/S0024-3841(98)00046-1.

Newport, E. (1981). Constraints on structure: evidence from American Sign Language and language learning. In W. A. Collins (Ed.), *Aspects of the development of competence. minnesota symposium on child psychology* (Vol. 14, pp. 93–124). New York: Erlbaum.

Newport, E. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*(1), 11–28. doi:10.1207/s15516709cog1401_2.

Newport, E., & Aslin, R. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162. doi: 10.1016/S0010-0285(03)00128-2-

Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, 56–119. doi: 10.2307/415601

Norcliffe, E., & Jaeger, T. (2015). Predicting head-marking variability in Yucatec Maya relative clause production. *Language and Cognition*, FirstView(12), 1–39. doi: 10.1017/langcog.2014.39.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357. doi:10.1037/0033-295X.115.2.357.

Ohala, J. (1989). Discussion of Lindblom's "Phonetic invariance and the adaptive nature of speech." In B. Elsendoorn & H. Bouma (Eds.), *Working models of human perception* (pp. 175–183). London: Academic Press.

Piantadosi, S., Tily, H., & Gibson, E. (2011a). The communicative function of ambiguity in language. *Cognition*, *122*, 280–291. doi:10.1016/j.cognition.2011.10.004.

Piantadosi, S., Tily, H., & Gibson, E. (2011b). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(9), 3526. doi:10.1073/pnas.1012551108.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pintzuk, S. (2002). Morphological case and word order in Old English. *Language Sciences*, *24*, 381–395. doi:10.1016/S0388-0001(01)00039-0.

Pozzan, L., & Trueswell, J. (in press). Revise and resubmit: how real-time parsing limitations influence grammar acquisition. *Cognitive Psychology*.

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*, 168–182. doi:10.1006/jmla.1999.2668.

Qian, T., & Jaeger, T. F. (2012). Cue effectiveness in communicatively efficient discourse production. *Cognitive Science*, *36*, 1312–1336. doi:10.1111/j.1551-6709.2012.01256.x.

Rafferty, A., Griffiths, T., & Klein, D. (2014). Analyzing the rate at which languages lose the influence of a common ancestor. *Cognitive Science*, *38*(7), 1406–1431. doi:10.1111/cogs.12112.

Sapir, E. (1921). *Language: an introduction to the study of speech*. New York: Harcourt, Brace.

Shimojo, M. (2006). Properties of particle "omission" revisited. Paper presented at the Toronto working papers in Linguistics.

Siewierska, A. (1998). Variation in major constituent order. In A. Siewierska (Ed.), *Constituent order in the languages of Europe* (pp. 465–529). Berlin: Mouton de Gruyter.

Silverstein, M. (1976). Hierarchy of features and ergativity. In R. M. W. Dixon (Ed.), *Grammatical categories in Australian languages* (pp. 112–171). Canberra: Australian Institute of Aboriginal Studies.

Sinnemäki, K. (2008). Complexity trade-offs in core argument marking. In F. Karlsson, M. Miestamo, & K. Sinnemäki (Eds.), *Language complexity: Typology, contact, change* (pp. 67–88). Amsterdam: John Benjamins.

Slobin, D. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development*. New York: Holt, Rinehart, Winston.

Smith, K., & Wonnacott, E. (2010). Eliminating unpredictable variation through iterated learning. *Cognition*, *116*(3), 444–449. doi:10.1016/j.cognition.2010.06.004.

Snedeker, J., & Trueswell, J. (2004). The developing constraints on parsing decisions: The role of lexical biases and referential scenes in child and adult sentence processing. *Cognitive Psychology*, *49*(3), 238–299. doi:10.1016/j.cogpsych.2004.03.001.

van Son, R. J. J. H., & Pols, L. C. W. (2003). How efficient is speech? Paper presented at the Proc of the Institute of Phonetic Sciences, Amsterdam.

van Son, R. J. J. H., & van Santen, J. P. H. (2005). Duration and spectral balance of intervocalic consonants: a case for efficient communication. *Speech Communication*, *47*, 464–484. doi:10.1016/j.specom.2005.06.005.

St Clair, M. C., Monaghan, P., & Ramscar, M. (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science*, *33*(7), 1317–1329. doi:10.1111/j.1551-6709.2009.01065.x.

Szmrecsanyi, B. (2004). *On operationalizing syntactic complexity*. Paper presented at the JADT, Louvain-la-Neuve, Belgium.

Tily, H. (2010). *The role of processing complexity in word order variation and change*. Ph.D. dissertation, Stanford University.

Tily, H., Frank, M., & Jaeger, T. (2011). The learnability of constructed languages reflects typological patterns. In L. Carlson, C. Hoelscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the Cognitive Science Society* (pp. 1364–1369). Boston, MA: Cognitive Science Society.

Traugott, E. C. (1972). *A history of English syntax*. New York: Holt, Rinehart and Winston.

Trueswell, J., Kaufman, D., Hafri, A., & Lidz, J. (2012). *Development of parsing abilities interacts with grammar learning: Evidence from Tagalog and Kannada*. Paper presented at the Proceedings of BUCLD 36, Sommerville, MA.

Trueswell, J., Sekerina, I., Hill, N., & Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*, 89–134. doi:10.1016/S0010-0277(99)00032-3.

Urosevic, Z., Carello, C., Savic, M., Lukatela, G., & Turvey, M. (1981). Some word-order effects in Serbo-Croat. *Language and Speech*, *29*(2), 177–195.

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316. doi:10.1016/S0749-596X(03)00081-0.

Van Everbroeck, E. (2003). Language type frequency and learnability from a connectionist perspective. *Linguistic Typology*, *7*(1), 1–50. doi:10.1515/lity.2003.011.

Wasow, T. (2015). Ambiguity avoidance is overrated. In S. Winkler (Ed.), *Ambiguity: Language and communication* (pp. 21–51). Berlin: DeGruyter.

Wilson, C. (2006). Learning phonology with substantive bias: An experimental and computational study of velar palatalization. *Cognitive Science*, *30*(5), 945–982. doi:10.1207/s15516709cog0000_89.

Yatanabe, S. (1999). Particle ellipsis and focus projection in Japanese. *Language, Information, Text*, *6*, 79–104.

Yokoyama, O. (1986). *Discourse and word order*. Amsterdam/Philadelphia: John Benjamins.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Cambridge, MA: Addison-Wesley.