Production is biased to provide informative cues early: Evidence from miniature artificial languages

Maryia Fedzechkina (mfedze@sas.upenn.edu)
Department of Psychology, University of Pennsylvania,
Philadelphia, PA 19104 USA

T. Florian Jaeger (fjaeger@bcs.rochester.edu)
Department of Brain & Cognitive Sciences, University of Rochester
Rochester, NY 14627 USA

John C. Trueswell (trueswel@psych.upenn.edu)
Department of Psychology, University of Pennsylvania,
Philadelphia, PA 19104 USA

Abstract

The role of processing constraints on sentence structure has been a topic of central interest in cognitive science. One proposal (Hawkins, 2004) suggests that language production system is organized to facilitate efficient parsing. We experimentally test this hypothesis using a miniature artificial language learning paradigm. Our findings support this account. Even though the input languages did not favor early placement of cues to grammatical function assignment (case and word order), participants used these cues in their own productions significantly more often in such a way as to allow early correct parsing commitments. This preference interacted with a bias to mark the less expected: Participants tended to use more case-marking in non-English OSV sentences. Our results underscore the potential of miniature artificial learning for language production research.

Keywords: language acquisition; language processing; language production; artificial language learning

Introduction

To infer sentence meaning, listeners must integrate a variety of probabilistic linguistic and non-linguistic cues. This integration happens incrementally – listeners do not wait for speakers to finish their utterances, but instead form provisional structural hypotheses based on the cues available at a given moment during sentence comprehension (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Due to this incremental nature of sentence comprehension, listeners sometimes face situations when their initial parsing commitments turn out to be incorrect and need to be revised as more information becomes available (Frazier & Rayner, 1987). Since revisions are associated with longer processing times, the order in which cues are presented to the parser has a potential impact on processing efficiency: Informative cues that allow recognition of sentence structure early on would on average increase parsing speed.

One proposal –the *performance-grammar correspondence hypothesis* (Hawkins, 2004) -- argues that the human production system supports efficient parsing. It does so in part by ordering informative cues early in the sentence ('Maximize Online Processing', MaOP). Primary support for this account comes from typological correlations:

Grammatical structures that permit more efficient integration of words into the structural representation tend to be cross-linguistically more common. For example, the correlation between marking type (case vs. agreement) and the position of the verb in a sentence is claimed to stem from considerations of processing efficiency: Verb-final languages typically have case-marking on nouns, while verb-initial languages tend to have rich verb agreement since it provides earlier disambiguation points in the respective languages (Hawkins, 2004). This explanation, however, is still under debate as a recent study found only partial support for the claim that language structures favor early parsing decisions (Dryer, 2002).

Cross-linguistic correlations, however, need to be interpreted with caution. Typological analyses are known to suffer from sparsity of independent data points. Languages that evolved from common ancestors or remained in contact for a long time often share lexical and structural properties. These genetic and areal influences drastically reduce the effective sample size of languages available for analysis. Thus, statistical tests might not have enough power to uncover hypothesized universals (Dryer, 1989; Jaeger, Graff, Croft, & Pontillo, 2011).

While direct experimental tests of the MaOP hypothesis are missing, evidence consistent with this account comes from the developmental literature. Young learners are highly sensitive to the order of cues to sentence meaning and disproportionally rely on early arriving cues when making parsing decisions. Initial parsing commitments based on early cues, once formed, are difficult to revise for young learners (Choi & Trueswell, 2010; Trueswell, Sekerina, Hill, & Logrip, 1999). Cues that are available early in parsing also appear to be acquired earlier: First and second language learners show better command of causative verb morphology in verb-initial languages, where it is appears early in the sentence compared to verb-final languages, where it is available later (Pozzan & Trueswell, under review; Trueswell, Kaufman, Hafri, & Lidz, 2012).

Here we use a miniature artificial language learning paradigm to experimentally test the prediction that the language production system is organized to provide

informative cues early. Using experimentally designed languages allows us to separate and orthogonally control properties that are potentially correlated in natural languages (e.g., cue informativity and cue order). Furthermore, this paradigm has been successfully used to study processing influences on grammar acquisition (Pozzan & Trueswell, under review) and online processing of newly acquired grammatical structures (Wonnacott, Newport, & Tanenhaus, 2008), which underscores its suitability to study parsing-related preferences in production.

In the experiment, we expose learners to miniature artificial languages that do not have an a priori preference to provide informative cues early (i.e., informative cues to grammatical function assignment appear early or late with the same frequency in the input). If the production system is indeed organized to facilitate early parsing decisions, we would expect learners to exhibit a (potentially small) bias to deviate from the statistics of the input towards using informative cues early in the sentence.

The experiment

Participants

56 monolingual native speakers of English were recruited from the University of Pennsylvania community to participate in the experiment. They received credit in introductory psychology classes for their participation. 17 participants were excluded from all further analyses for the following reasons: failure to complete all sessions of the experiment (4 participants); previous familiarity with the phenomenon under investigation (1 participant); failure to achieve 70% comprehension accuracy on the final day of testing (12 participants, see Scoring below). This left 39 participants for the final analysis (20 in the subject-marking and 19 in the object-marking language).

Input languages

The two languages used in the experiment contained 10 novel words (6 animate nouns and 4 transitive verbs), all of which conformed to English phonotactics, and a suffix 'di' that acted as a nominative or accusative case-marker depending on the language (see Table 1). All sentences described simple transitive events such as 'hug' or 'poke' performed by two male actors such as 'chef' or 'referee' (see Figure 1 for example stimuli). All verbs occurred equally frequently within each language overall and with each word order variant within each language. All nouns occurred equally often in the subject and object position with each verb

Participants were assigned to learn one of the two miniature languages. Both languages were verb-final (i.e., the verb followed both the subject and the object). This word order was chosen as it is cross-linguistically more common in languages with a case system (Greenberg, 1963). Word order was flexible: Subject-object-verb (SOV) and object-subject-verb (OSV) orders occurred equally frequently in both languages. Both languages had optional

case-marking – a case-marker (suffix 'di') was present in 67% of the input sentences and absent in the remaining 33%. The two languages differed in the locus and function of case-marking. In the *subject-marking language*, the case-marker optionally marked the subject and never the object. In contrast, in the *object-marking language*, the case-marker optionally occurred on the object and never on the subject.

Table 1: The artificial lexicon used in the experiment.

Nouns	Verbs	Case-marker
barsu	moship	di
doakla	kyse	
rizbi	skroop	
lanfu	tegud	
peza		
forpih		

If the production system exhibits a bias to facilitate correct parsing commitments early on, we expect learners to use cues to grammatical function assignment in such a way as to allow early disambiguation. Since word order was uninformative about grammatical function assignment in the experiment, case-marking (when present) provided important disambiguating information in both languages. We thus expect learners to produce more case-marking on the first constituent in a sentence regardless of its locus (subject vs. object) in the newly acquired language.

The procedure

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. No further instructions about the structure of the language were provided.

Participants were trained and tested on one of the two miniature languages during three visits to the lab (each lasting about 40-45 minutes) with at most one day between the visits. During every visit, learners were exposed to a mixture of noun and sentence exposure blocks, comprehension tests (noun and sentence), and short vocabulary tests. Each session concluded with a sentence production test. The same overall procedure was followed on all days. On the first day, however, the amount of noun exposure and vocabulary testing was doubled compared to subsequent days.

Every session began with *noun exposure*. Learners saw a picture of each character labeled in isolation and were encouraged to repeat the names to facilitate learning. After the initial familiarization, learners were presented with pictures of four characters accompanied with a label corresponding to one of them. Learners were instructed to click on the correct picture. In order to facilitate vocabulary learning, which is a prerequisite for the success of the experiment, but not itself of interest here, feedback was provided on every trial. Noun learning concluded with a

noun production test, during which learners were asked to provide labels for each character presented in isolation.
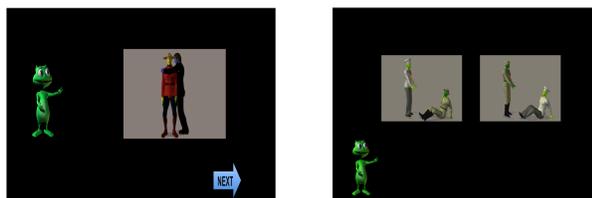

Figure 1: Screenshots of sentence exposure (left panel) and sentence comprehension (right panel) trials.

During *sentence exposure* that followed, learners watched short videos (2 blocks of 24 trials) depicting simple transitive events (e.g., the chef hugging the mountie) accompanied by their descriptions in the novel language. Participants were instructed to repeat the descriptions to facilitate learning. Learners could replay the video and sound as many times as they wished during the first training block on day 1; no further replays were allowed throughout the experiment. Sentence exposure was followed by a comprehension test (24 trials total), during which learners heard a sentence in the novel language and were shown two videos depicting the same referents and action with the reversed order of the actor and patient in the two videos. Participants were asked to click on the video that matched the description. No feedback was provided on sentence comprehension trials. After competing the test, participants completed two more sentence exposure and one more sentence comprehension blocks.

Each session ended in a sentence production test (48 trials), during which participants watched previously unseen videos depicting familiar characters and actions and described them in the novel language. Participants were given an auditory verb prompt that they could use in their sentence to facilitate production. No other instructions on form or content of the descriptions were provided.

Scoring

During every session, we collected measures of accuracy on vocabulary tests, comprehension tests, sentence production accuracy, as well as case and word order preferences in sentence production.

In the sentence comprehension test, we analyzed participants' responses on case-marked (i.e., unambiguous) trials. Since word order was uninformative in our experiment, this measure indicated how well learners have acquired the meaning of case-marking. 12 participants (5 in the subject-marking and 7 in the object-marking language) who failed to achieve 70% accuracy were excluded from all analyses. The results reported below hold regardless of this exclusion.

All production trials (noun and sentence production) were orthographically transcribed by two transcribers and then automatically scored using a Python script. A noun or verb was considered correctly pronounced if it fell within Levenstein distance of 2 from the target label (i.e., we allowed at most 2 character insertions, deletions, or substitutions in a word). For example, 'toagla' was still considered a correct label for 'doakla', but 'toagli' was not.

In the sentence production test, we scored sentence word order, presence and position of case-marking as well as lexical (using the wrong name for a referent or an action) and grammatical mistakes (using word order other than SOV or OSV or using the case-marker on the wrong constituent). If the name of only one referent was incorrect and it was still possible to determine sentence word order, productions were scored as overall correct but containing a lexical error. Such productions were included in the analyses below. Productions containing grammatical errors were excluded from all analyses.

Results

Accuracy of acquisition For participants included in the analysis, vocabulary test performance was at ceiling on the final day of training (99% correct in the object-marking and 100% correct in the subject-marking language). Similarly, the number of lexical mistakes during sentence production was below 3% across the two languages on the final day of training (2.2% in the object-marking and 2.9% in the subject-marking language). The number of grammatical mistakes made by the learners was below 1% in both languages (0.33% in the object-marking and 0.42% in the subject-marking language) on the final day of training. This performance suggests that the task was feasible for our participants.

We now turn to the main question of our study – whether language production system is biased towards providing informative cues early.

Do learners provide case-marking early in sentences? To address this question, we examined learners' case-marker use in production. We used a mixed logit model (Jaeger, 2008) to predict the presence of case-marking based on its locus in the language (sum-coded, subject vs. object), day of training (Helmert coded, 2 vs. 1, 3 vs. 1 and 2), constituent position (sum-coded, first vs. second), and all interactions between these factors. The model included maximal converging random effects structure. Fixed effect correlations greater than 0.4 were observed between multiple predictors in the model. The results reported here did not change after the stepwise model reduction was used to reduce collinearity.

Learners' overall proportion of case-marking did not significantly differ across the two languages ($\chi^2(2)=0.11$) or days of training ($\chi^2(2)=0.13$). Neither did language and day of training interact ($\chi^2(2)>0.18$). Learners did, however, use significantly more case-marking on the first constituent as compared to the second one (main effect of constituent position, ß=0.86, z=2.05, p<0.05). Constituent position interacted with language (ß=1.02, z=2.92, p<0.01). Simple effects test revealed that the effect of constituent position on case-marker use was significant in the object-marking

language (ß=1.88, z=3.34, p<0.001), but not in the subject-marking language (p>0.7). Thus, only learners of the object-marking language preferentially case-marked the first constituent (see Figure 2).
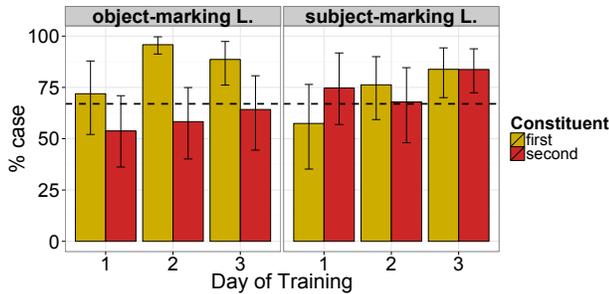

Figure 2: Case-marker use in production in the object-marking (left panel) and subject-marking (right panel) language. The dashed line indicates the input (equal for both languages).

We further compared participants' use of case-marking to the input proportion in the two languages (see Figure 3). On the final day of training, learners of both languages produced significantly more case-marking compared to the input (Subject-marking language, day 3: 83% case-marking in production, significantly higher than the input [χ^2 (1)=15.81, p<0.001]; Object-marking language, day 3: 83% case-marking in production, significantly higher than the input [χ^2 (1)=15.72, p<0.0001]).
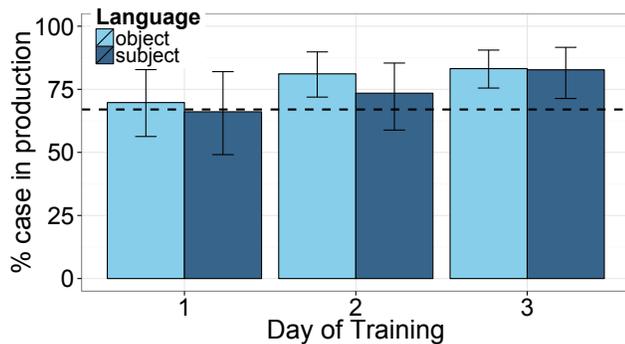

Figure 3: Overall case-marker use in production. The dashed line indicates the input (equal across the two languages).

Thus, while learners of both languages used significantly more case-marking in their own productions compared to the input, only learners of the object-marking preferentially case-marked the first constituent. Learners of the subject-marking language did not use case-marking asymmetrically and marked both constituents equally often.

What can account for this difference in the preference to case-mark the first constituent across the two languages? One possibility is that it is due to another production preference – a preference to mark the less expected (Jaeger, 2013; Piantadosi, Tily, & Gibson, 2011). Recall that the two languages in our experiment differed in the word order that allowed the earliest availability of case-marking during parsing (SOV in the subject-marking and OSV in the object-marking language). OSV word order is, however, cross-linguistically rare and, more importantly, highly uncommon in the native language of our participants (English), while SOV word order is more in line with participants' native language biases (since the subject in English typically precedes the object). A preference to provide case-marking early and a preference to mark the less expected non-English OSV word order work in the same direction for the object-marking language (both favor OSV) and in opposite directions for the subject-marking language (a preference to mark the atypical favors OSV while a preference for early disambiguation favors SOV). Thus, if the two biases are equally strong in production, one would expect the same degree of case-marking on both constituents in the subject-marking language and a relatively strong bias to mark the first constituent in the object-marking language, as observed in our experiment.

To further probe this question, we examined learners' comprehension performance on ambiguous trials (i.e., trials without case-marking): If comprehenders interpret such sentences preferentially as SOV, this would suggest that learners of both languages indeed treat OSV word order as less typical. In a mixed logit model, we predicted SOV use based on the language (sum-coded, subject vs. object), day of training (Helmert coded, 2 vs. 1, 3 vs. 1 and 2), and their interaction. The model included maximal converging random effects structure. Fixed effect correlations greater than 0.4 were observed between multiple predictors in the model. The results reported here did not change after the stepwise model reduction was used to reduce collinearity.
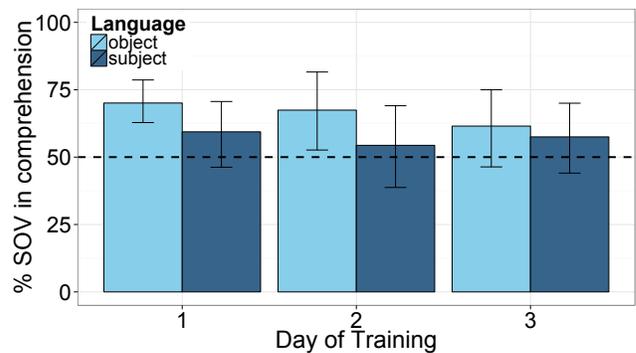

Figure 4: % SOV word order in comprehension of non-case-marked sentences. The dashed line indicates the input (same across the two languages).

The analysis revealed a baseline SOV preference in both languages as indicated by a significant positive intercept (ß=0.7, z=2.5, p<0.05). There was no significant main effect of language (p=0.18), day (χ^2 (2) = 0.83), or language by case interaction (χ^2 (2) = 0.43), see Figure 4. Simple effects test revealed no significant effect of language on any day of

training (p's>0.17 for all days). Thus, comprehension performance suggests that SOV interpretations for ambiguous sentences were preferred across conditions.

Summary Learners of the two languages in our experiment showed different patterns of case-marker use. As expected under our hypothesis, learners of the object-marking language preferentially marked the first constituent. Learners of the subject-marking language, however, case-marked both constituents equally often. This difference in case-marker use across the two languages is likely to reflect two production biases – a bias to mark the atypical (i.e., non-English OSV order dispreferred in comprehension) and a bias to provide informative cues early. Since the two biases work in opposite directions in the subject-marking language, the preference to case-mark the first constituent is expected to be less pronounced in this language.

Discussion

In a miniature artificial language learning experiment, we found that learners preferentially used cues to grammatical function assignment in such a way as to promote correct parsing commitments early. Learners of both languages produced case-marking significantly more often than it was present in the input. Learners of the object-marking language tended to case-mark the first constituent significantly more often than the second one, while learners of the subject-marking language produced the same amount of case-marking on both constituents. We argued that this behavior was indicative of two biases influencing language production – a preference to provide informative cues early and a preference to case-mark the less expected.

Our findings thus add to the growing body of research investigating the role of cue order in language acquisition and use. To the best of our knowledge, the current experiment constitutes the first direct experimental evidence in support of the hypothesis that the human production system is organized to produce informative cues earlier in the sentence (Hawkins, 2004). A close parallel between learners' preferences in production and patterns observed in typology suggests that some typological properties observed cross-linguistically stem from biases associated with incremental language processing. Finally, our findings parallel a variety of developmental phenomena, such as young learners' disproportional reliance of early cues during parsing (Choi & Trueswell, 2010; Trueswell et al., 1999) and a faster time line of early cue acquisition (Pozzan & Trueswell, under review; Trueswell et al., 2012).

By using experimentally designed miniature artificial languages, we were able to establish that a preference to provide informative cues early was not the only bias the production system caters to. Replicating earlier work (Fedzechkina, Jaeger, & Newport, 2012; Fedzechkina, Newport, & Jaeger, under review), we also observed a preference to mark the atypical in participants' production. Learners tended to use more case-marking in the less expected (non-English) OSV sentences. This preference interacted with a bias to provide informative cues early, resulting in a stronger preference to case-mark the first constituent in the object-marking language.

Since the two biases are closely intertwined in natural language, it would be difficult or perhaps even impossible to isolate and experimentally study their respective influences using natural language stimuli. Our findings, thus, underscore the potential of the miniature artificial language learning paradigm, which has previously been applied to language acquisition and in rare cases language comprehension, for language production research.

Finally, our findings raise a question of native language influences in miniature artificial language learning experiments. Learners in our experiment had a preference to mark the less typical OSV word order, which likely reflects a transfer from participants' native language (English). While our data do not speak to the exact circumstances under which such transfer is likely to occur or directly investigate the degree of such transfer, they point out that the influence of native language background is a concern for miniature artificial language experiments and that learners' native language background needs to be taken into account when interpreting experimental results (see Goldberg, 2013 for similar and additional arguments).

Conclusions

We used a miniature artificial language learning paradigm to experimentally investigate the hypothesis that language production is organized to provide informative cues early. Our findings support this hypothesis: Learners tended to use case-marking in their productions in such a way as to allow early commitments to the correct parse. These outcomes parallel a variety of developmental phenomena and patterns observed in typology.

Acknowledgments

References

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. *J Exp Psychol, 106*(1), 41-61.

Dryer, M. (1989). Large linguistic areas and language sampling. *Studies in Language, 13*, 257-292.

Dryer, M. (2002). Case Distinctions, Rich Verb Agreement, and Word Order Type (Comment on Hawkins' paper). *Theoretical Linguistics, 28*, 151-157.

Fedzechkina, M., Jaeger, T., & Newport, E. (2012). Language learners restructure their input to

facilitate efficient communication. *Proc Natl Acad Sci USA, 109*(44), 17897-17902.

Fedzechkina, M., Newport, E., & Jaeger, T. (under review). Balancing effort and information transmission during language acquisition: Evidence from word order and case-marking.

Frazier, L., & Rayner, K. (1987). Resolution of syntactic category ambiguities: Eye movements in parsing lexically ambiguous sentences. *Journal of Memory and Language, 26*(5), 505-526.

Goldberg, A. (2013). Substantive learning bias or an effect of similarity? Comment on Culbertson, Smolensky, Legendre (2012). *Cognition, 127*(3), 420-426.

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): MIT Press.

Hawkins, J. A. (2004). *Efficiency and Complexity in Grammar*. Oxford: Oxford University Press.

Jaeger, T. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang, 59*, 434-446.

Jaeger, T. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology, 4*(230).

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.

Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proc Natl Acad Sci USA, 108*(9), 3526.

Pozzan, L., & Trueswell, J. (under review). Revise and resubmit: How real-time parsing limitations influence grammar acquisition.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632.

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.

Wonnacott, E., Newport, E., & Tanenhaus, M. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychol, 56*(3), 165-209.