Producing informative cues early: Evidence from a miniature artificial language

Maryia Fedzechkina[1], T. Florian Jaeger[2], John C. Trueswell[1]

[1]Psychology, University of Pennsylvania, [2]Brain & Cognitive Sciences, University of Rochester

{*mfedze@sas.upenn.edu; fjaeger@bcs.rochester.edu; trueswel@psych.upenn.edu*}

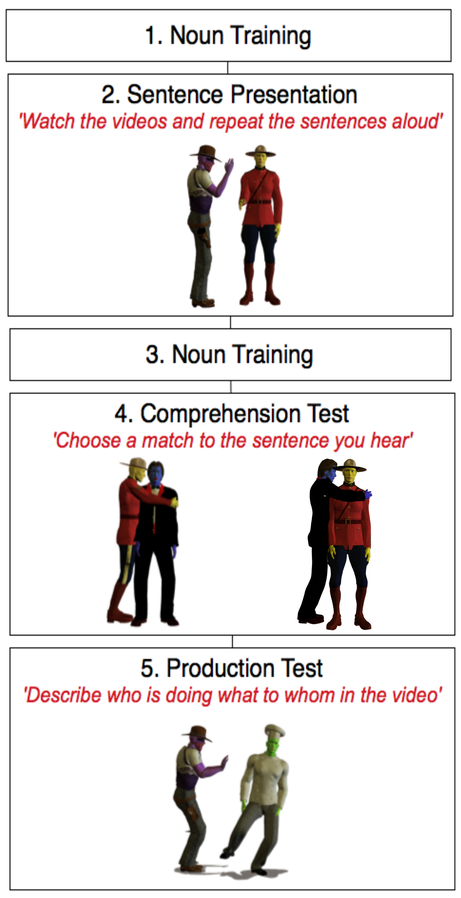
Abstract

The impact of processing constraints on sentence structure has been a topic of central interest in cognitive science. One proposal [1] suggests that language production system is organized to provide informative cues early. We experimentally test this hypothesis using a miniature artificial language learning paradigm. By using experimentally-designed languages, we can separate and orthogonally control properties that are potentially correlated in natural languages (e.g., cue informativity and cue order).

Experiment design

A miniature artificial language learning study

- 39 monolingual English speakers
- over 3 consecutive days



Miniature input grammars

Object-marking language
SOV/OSV *word order* ***50/50%***
Object *case-marking 67%*
(no subject case-marking)

Subject-marking language
SOV/OSV *word order* ***50/50%***
Subject *case-marking 67%*
(no object case-marking)

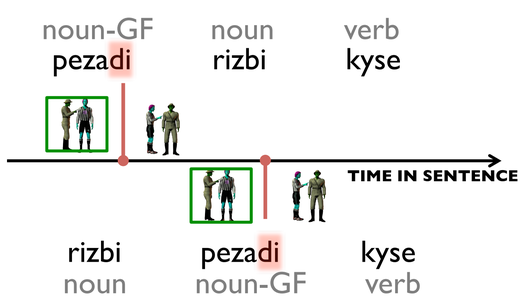
Miniature input lexicon

6 Nouns: {barsu, doakla, rizbi, lanfu, peza forpih}
4 Verbs: {tegud, moship, kyse, skroop}

Predictions

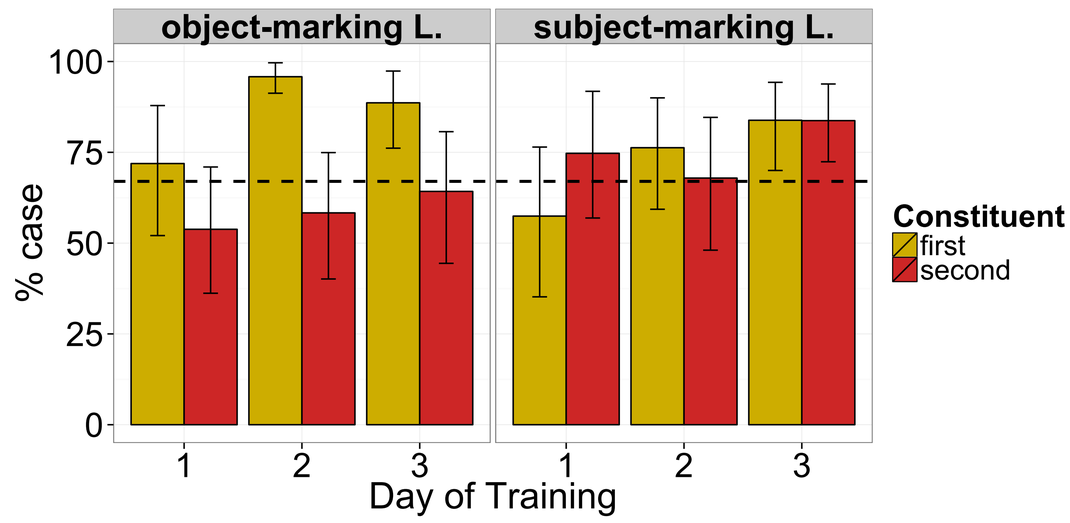
- More case on the 1st constituent in both languages (i.e., in OSV sentences in the object- and SOV sentences in the subject-marking language).
- BUT: non-English OSV word order might attract more case due to a preference to mark the less expected [2,3].

Earliest disambiguation with case on the 1st constituent



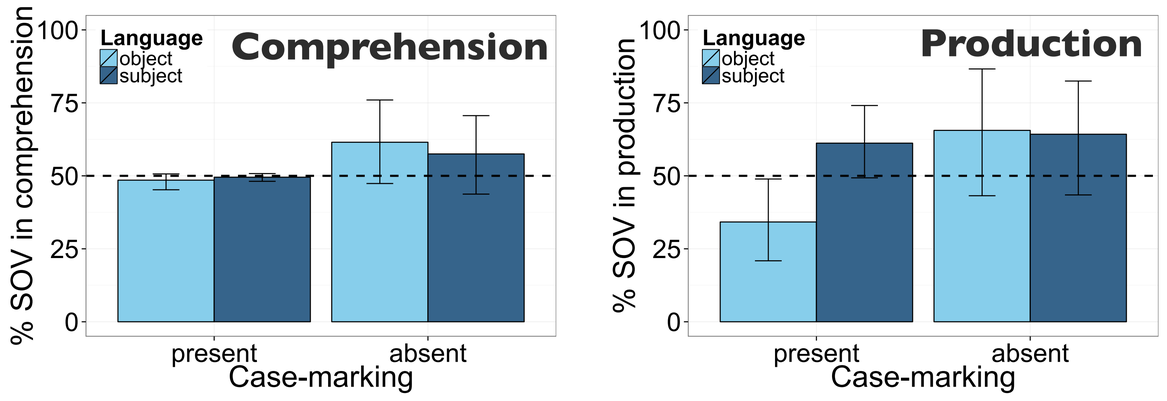
Results

Case-marker use in production



- Learners of the object-marking language tended to case-mark the first constituent significantly more often than the second one.
- Learners of the subject-marking language produced the same amount of case-marking on both constituents.

Word order preferences on the final day



- Learners of both languages had a non-significant baseline SOV preference on non-case-marked trials in comprehension and production.
- Learners of the object-marking language produced significantly more OSV for case-marked than for non-case-marked trials; learners of the subject-marking language did not have a differential word order preference based on the presence of case.

Discussion

- We present 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 [1].
- Learners' behavior is indicative of two biases influencing language production – a preference to provide informative cues early and a preference to case-mark the less expected.
- Since the two biases are closely intertwined in natural language, it would be difficult to study their respective influences in natural language production. Our findings underscore the potential of the miniature artificial language applications for language production research.

References

[1] Hawkins, J.A. (2004). *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

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

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

Acknowledgements

This work was supported in part by NSF CAREER grant IIS-1150028 to TFJ, and by NIH grant 1-R01-HD 37507 to JCT. We thank C. Schneider, K. Woodard, J. Padilla, and B. Oshiba for their help with data collection and annotation; C. Donatelli, I. Minkina, and A. Wood for their help creating video stimuli.