

Word Order and Case Marking in Language Acquisition and Processing

Maryia Fedzechkina, T. Florian Jaeger, Elissa L. Newport

Department of Brain & Cognitive Sciences, University of Rochester

{mashaf@bcs.rochester.edu; fjaeger@bcs.rochester.edu; newport@bcs.rochester.edu}

Abstract

To understand a sentence, comprehenders must identify its actor and patient. In principle, these relationships can be signaled using a single cue, but most languages employ several redundant cues, including word order and case marking. In artificial language learning experiments we investigate word order and case as cues in processing and learning. Flexible word order languages are potentially ambiguous if no case-marking (or other cues) are employed to identify the doer of the action. We explore whether language learners have a bias towards ambiguity reduction and tend to fix word order in non-case-marking languages (cf. the historic change from Old English to contemporary English).

The Experiment

Lexicon

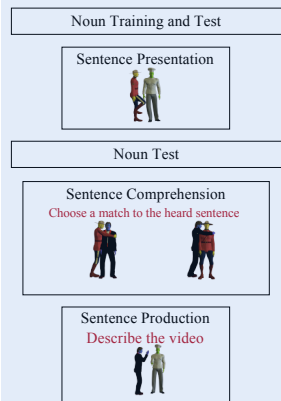

Noun: 6 humans
Verb: 10
Case-marker: “kah”

Grammar

Word Order	Object case-marked?	Number of verbs associated with dominant word order			
		100%	83%	50%	0%
Lg 1 SOV (63%)	no	3	3	2	2
Lg 2 OSV (37%)	yes	3	3	2	2

An artificial language learning study

- tested 19 monolingual English native-speakers
- over 4 consecutive days:

The ref is poking the hunter. The hunter is poking the ref.

LG 1 (NO Case):
Melnawg bliffen zamper. OR Bliffen melnawg zamper.

LG 2 (Case):
Melnawg bliffen kah zamper. Bliffen melnawg kah zamper.
Bliffen kah melnawg zamper. Melnawg kah bliffen zamper.

Results

Word Order Fixing

In the absence of case-marking there is a tendency to generalize the majority order in both comprehension and production (for comprehension, $p < 0.05$ and for production, $p < 0.09$). The tendency to generalize the dominant word order appears most strongly in the no-case condition when the majority word order is most variable in the input language. In contrast, when case marking is present, the word order variability of the input language is more accurately reproduced in learners' usage.

Potential Origin of Word Order Fixing

- more spread in word order consistency among the learners of the non-case-marking language
- a group of people who have a strong preference for the dominant word order in non-case-marking language

Fig. 3: Word order preferences

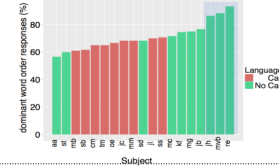


Fig. 1: Comprehension, final day of training

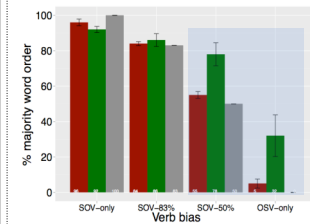
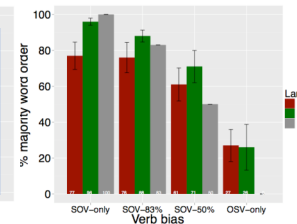


Fig. 2: Production, final day of training

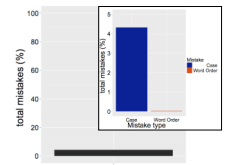


Accurate Acquisition

Both languages are acquired accurately (<5% mistakes), suggesting that the task is manageable.

Word order errors (using word orders not allowed in the language) are extremely rare; the majority of errors are case errors (e.g., incorrect use or omission of a case-marker).

Fig. 5: Total Mistakes



Processing Advantage of Case

There is no effect of case on reaction times. This result is, however, driven by the same learners of the non-case-marking language who fix word order (cf. Fig. 3). The effect of case on reaction time reaches significance ($p = 0.055$) once these fixed word order users are removed from the analysis.

Standard processing effects [1-3] also hold: Bias-consistent responses are significantly faster.

Fig. 5: Mean Reaction Times

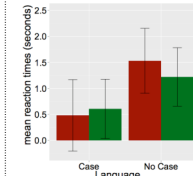
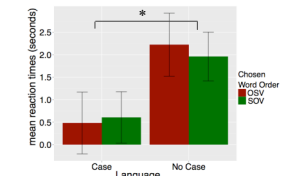


Fig. 6: Mean RTs, overgeneralizers removed



Discussion

In the absence of case-marking, learners tend to generalize the dominant word order in both comprehension and (somewhat less) in production, suggesting a bias to reduce ambiguity in the language.

This effect is partially driven by a group of learners of the non-case-marking language who show extreme word order fixing both in comprehension and production.

When case is present, learners use it to resolve ambiguity. We do not see any learners of the case-marking language who show word order generalization in the production task. If anything, they UNDERgeneralize (one interpretation being that learning word order is less important to them).

Having several redundant cues to meaning poses additional difficulty in language acquisition. In our simple setup where the case-marker is consistent and easily segmentable in the speech stream the majority (99%) of mistakes are associated with incorrect use of the case-marker. Having case in a language thus comes at a cost: Case takes longer to acquire.

Case provides a processing advantage for languages with flexible word order. Once the extreme word order generalizers are removed from the analysis, stimuli in the case-marking language are processed significantly faster.

References

- [1] Gagny, S. M., Furlow, N. J., Myers, E., and Jacoby, M. A. (1997). The contribution of orth bias and plausibility to the comprehension of temporally ambiguous sentences. *Journal of Memory and Language*, 37, 58-83.
- [2] Spelke, E., and Newport, E. L. (1986). The developing constraints on parsing decisions: The role of lexical biases and referential access in child and adult sentence processing. *Cognitive Psychology*, 18, 215-299.
- [3] Frazier, L., and Johnson, M. K. (1972). Sentence processing: Separating effects of lexical preference from garden-path. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1, 23-32.