Functional pressures in (artificial) language learning

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

Why do languages share structural properties? Functionalists have argued that languages have evolved to suit the needs of their users [1, 2]. We investigate whether the biases operating in language acquisition [3, 4] are at least partially driven by functional pressures. In two artificial language learning experiments we explore the trade-off between word order and case-marking as well as optional case-marking. We explore whether language learners are biased against uncertainty in the mapping of form and meaning, showing a tendency to make word order a stronger cue to the intended meaning in no-case languages and to case-mark atypical (animate) objects if case-marking is optional in the language.
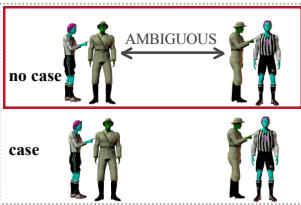
Experiment I

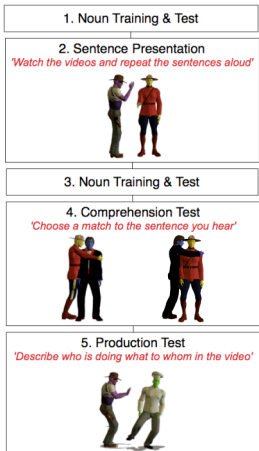**Input Lexicon**

Noun: 6 humans
Verb: 10
Case-marker: "kah"

Input 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

Results:

Word order regularization in the absence of case in comprehension ($p<0.05$) and production ($p<0.09$). This tendency appears most strongly when the dominant word order is most variable in the input language.

Word order and case interact in acquisition. Without case, learners regularize word order, suggesting a bias against systemic ambiguity. When case is available, learners use it to resolve ambiguity.
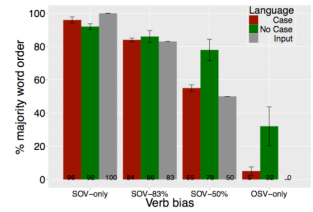

Fig. 1: Comprehension, final day of training

An artificial language learning study

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

The Procedure:

1. Noun Training & Test
2. Sentence Presentation
   'Watch the videos and repeat the sentences aloud'
3. Noun Training & Test
4. Comprehension Test
   'Choose a match to the sentence you hear'
5. Production Test
   'Describe who is doing what to whom in the video'

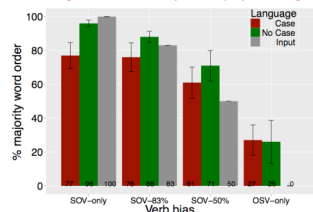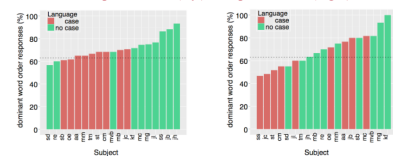*Fig. 2: Production, final day of training*

Fig. 3: Individual word order preferences in comprehension (left) and production (right)

Experiment II

Input Lexicon

Subject noun: 5 humans
Object noun: 5 humans, 5 inanimate
Verb: 8
Case-marker: "kah"

Input Grammar

Word Order	Object case-marked?	Animacy of object
SOV (60%) OSV (40%)	60%	human (50%) inanimate (50%)

An artificial language learning study:

- tested 19 monolingual native English speakers
- procedure as in Experiment I with minor changes:
  - no production test on Day 1
  - comprehension test -- 'click on the doer of the action'
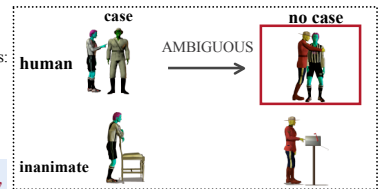
Differential object marking cross-linguistically

- found in natural languages (e.g., Japanese, Korean)
- governed by several hierarchies [5, 6]:
  - person
  - definiteness
  - **animacy: *human > animate > inanimate***

Results:

Language learners retain **significantly more case-markers on animate objects and in the non-canonical word order** where the uncertainty about the intended meaning is highest.

Effects replicate under less flexible word order (SOV 80%, OSV 20%), 10 subjects.
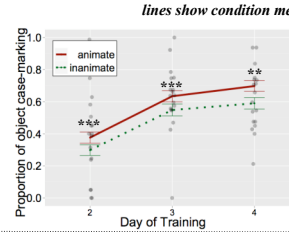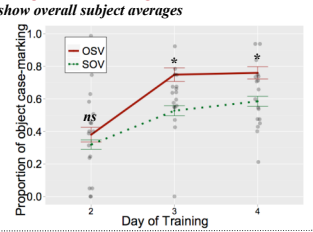

Fig. 4: Case-marking is sensitive to animacy
Fig. 5: Case-marking is sensitive to word order
lines show condition means, points show overall subject averages

Discussion

We have investigated the *trade-off between word order and case-marking* as well as *optional case-marking* and have found several striking results:

1. *Some learning biases are functional in nature. Language learners have a bias against systemic ambiguity* and tend to reduce ambiguity in language during acquisition by:

- regularizing word order in the absence of case (Exp I)
- retaining more case-marking on atypical (animate) objects (Exp II)
- retaining more case-marking on scrambled objects (Exp II)

This behavior cannot be accounted for by native-language (English) bias since it fails to explain differential word order regularization depending on the condition in Exp I and the fact that learners induce a structure into the language that is not present either in the input or their native language in Exp II.

2. *Learning biases mirror typologically frequent patterns:*

- loss of case and word order fixing from Old English to Modern English (Exp I)
- differential case-marking systems in Japanese, Korean and Hindi (Exp II)

Learning biases can thus offer an account of some of the structural similarities found in natural languages.

References

[1] Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner, & L. Gleitman (Ed.), Language acquisition: The state of the art. New York: CUP.
[2] Hawkins, J.A. (2004). Efficiency and Complexity in Grammars. Oxford: Oxford University Press.
[3] Newport, E.L., & Aslin, R.N. (2004). Learning at a distance: I. Statistical learning of non-adjacent dependencies. Cognitive Psychology, 48, 127-162.
[4] Culbertson, J., & Smolensky, P. (2009). Testing Greenberg's universal 18 using an artificial language learning paradigm. NELS40, Cambridge, MA.
[5] Aissen, J. (2003). Differential object marking: iconicity vs. economy. Natural Language and Linguistic Theory, 21, 435-483.
[6] Lee, H. (2006). Parallel optimization in case systems: Evidence from case ellipsis in Korean. Journal of East Asian Linguistics, 15, 69-96.