What can we learn from bound learners?

Barend Beekhuizen

Leiden University Center for Linguistics
Leiden University

Institute of Logic, Language and Computation
University of Amsterdam

11 March 2013

Introduction

- Marr's (1982) three levels:
  - computational (CL)
  - algorithmic (AL)
  - implementational

Introduction

- Marr's (1982) three levels:
  - computational (CL)
  - algorithmic (AL)
  - implementational
- Typical question: Can a function be calculated at all? (CL)
- Does not (have to) consider AL properties of the system

Introduction

- Marr's (1982) three levels:
  - computational (CL)
  - algorithmic (AL)
  - implementational
- Typical question: Can a function be calculated at all? (CL)
- Does not (have to) consider AL properties of the system
- But: can the function be calculated given AL properties?
- Moreover, are there situations where AL properties explain certain behavior?
- Let's call a learner bound if it is constrained on the AL, i.e. in its representations and processing algorithms.

Constraints on the AL

- Are there situations where AL properties explain certain behavior (in language acquisition)?
- Evidence from different domains:

domain	reference	constraints on
decision making	Gigerenzer & Selten (2001)	search, decision
garden-path sentences	Ferreira & Patson (2007)	ability to track multiple analyses
word-meaning acquisition	Medina et al. (2012)	ability to track multiple lexical semantic analyses
production of root infinitives	Freudenthal et al. (2007)	working memory
grammar learning	Elman (1993)	working memory

Bound learners: theoretical perspective

Implications

What does this mean for modeling the acquisition of grammar?

- Learning is often taken to involve optimizing some function
- Is this the correct way of looking at it?

Bound learners: theoretical perspective

Implications

What does this mean for modeling the acquisition of grammar?

- Learning is often taken to involve optimizing some function
- Is this the correct way of looking at it?
- Optimization is often algorithmically very heavy
- Herbert Simon's (1955) idea of satisficing:
  making a decision that meets an aspiration level rather than optimizes the decision
- Bounded Rationality: domain-specific fast and frugal heuristics

Bound learners: theoretical perspective

Implications

What does this mean for modeling the acquisition of grammar?

- Learning is often taken to involve optimizing some function
- Is this the correct way of looking at it?
- Optimization is often algorithmically very heavy
- Herbert Simon's (1955) idea of satisficing:
  making a decision that meets an aspiration level rather than optimizes the decision
- Bounded Rationality: domain-specific fast and frugal heuristics
- For language acquisition: using dumb, heuristic strategies to learn, that stop when an "aspiration level" is met
- Social approach: Learner tries to get by (communicatively)

Types of constraints

Assumption

Assume an incremental learner trying to build up some sort of grammar.

- Types of constraints
  - Working memory: only process one new word per utterance and words to the right of that
    → Natural "starting small" heuristic (Elman 1993; Spitkovsky et al. 2009)
  - Search: only add one new syntactic rule per utterance
    → No full hypothesis space (U-DOP)
    → Danger of getting stuck in bad part of hypothesis space?
    → Starting small alleviates?
  - Parse: minimal attachment, late closure (with revision?)
  - Abstraction: lazy (only when analogy can be made)

Types of phenomena

Phenomena

What kind of phenomena are likely candidates for an explanation in terms of a bound learner?

- Recent Minimalist explanations of certain production phenomena (Yang & Roeper 2012)
- E.g. argument drop within *wh*-questions; order asymmetries; argument realization

Types of phenomena

Phenomena

What kind of phenomena are likely candidates for an explanation in terms of a bound learner?

- Recent Minimalist explanations of certain production phenomena (Yang & Roeper 2012)
- E.g. argument drop within *wh*-questions; order asymmetries; argument realization
- Generally: deviations in production from the adult grammar and input
- And their development/convergence
- E.g. over- and undergeneralization, chunks that are compositional
- Explain as interactions between bound learning algorithms and input.

Why take this approach?

- Desirability of this type of explanation: acknowledges limitations of processing system and explains behavior.
- Seems promising for developmental patterns
- Seems promising for deviations from input data
- Provides a general learning answer to parameter-setting explanations like Yang & Roeper's

Thank you

- Elman, J.L. (1993). Learning and development in neural networks: The importance of starting small. Cognition, 48, 71-99.
- Ferreira, F., & Patson, N. D. (2007). The "Good Enough" Approach to Language Comprehension. *Language and Linguistic Compass*, 1(1-2), 7183.
- Freudenthal, D., Pine, J. M., Aguado-Orea, J., & Gobet, F. (2007). Modeling the Developmental Patterning of Finiteness Marking in English, Dutch, German, and Spanish Using MOSAIC. *Cognitive science*, 31(2), 31141.
- Gigerenzer, G. & Selten, R. (2002) (eds.). *Bounded Rationality. The Adaptive Toolbox*. Cambridge, MA: MIT Press.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. New York: Henry Holt.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *PNAS* 108(22), 90149.
- Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69, 99118.
- Spitkovsky, V.I., Alshawi H., and Jurafsky, D. (2009). Baby Steps: How "Less is More" in Unsupervised Dependency Parsing. *NIPS, Workshop on Grammar Induction, Representation of Language and Language Learning*.
- Yang, C. & Roeper, T. (2012). Minimalism and Language Acquisition. *The Oxford Handbook of Linguistic Minimalism*