An extensible framework for mechanistic processing models

From representational syntax-semantics theories to quantitative model comparison

Adrian Brasoveanu & Jakub Dotlačil

MathPsych/ICCM 2018 · University of Wisconsin Madison
July 23, 2018

Overarching goal

- build formally and computationally explicit processing models for natural language syntax and semantics
- specifically, cognitively realistic models for incremental parsing of discourse representations structures (DRSs, Kamp 1981; Kamp and Reyle 1993) or similar representations
- the semantic and syntactic representations are created in parallel

Main goal for today:

- modeling *syntactic* representations

An extensible framework for processing models

- **mechanistic** models of language processing
- common approach: use an independently motivated, general cognitive architecture
- parsing easy to embed in hybrid cognitive architectures
  - Soar

    Hale, 2014; Young and Lewis, 1999
  - ACT-R

    Dillon et al., 2013; Engelmann et al., 2013; Kush, 2013; Lewis and Vasishth, 2005; Nicenboim and Vasishth, 2018; Rij, 2012; Taatgen and Anderson, 2002; Vasishth et al., 2008

The current state of ACT-R language modeling

- Mainly used to model recall of syntactic structures

  Dillon et al., 2013; Engelmann et al., 2013; Lewis and Vasishth, 2005; Nicenboim and Vasishth, 2018; Vasishth et al., 2008
- This focus on recall-related modeling does not take advantage of the generality of ACT-R as a cognitive architecture and its “no magic” policy
- Implemented fully in LISP (not a very popular programming choice now)
- ACT-R comes with many parameters; these are set to their default values or manually changed
- Modeling is hard to replicate; systematic quantitative model comparsion hard to perform

This talk

- A new Python3 implementation of ACT-R (pyactr; Brasoveanu and Dotlačil in prep.)
- ACT-R + Bayes: ACT-R models embedded in Bayesian models, hence systematic exploration of parameter values, model comparison, modeling easy to replicate
- the ACT-R component: a working, extensible parsing framework for syntax and semantics, with visual and motor interfaces (today, only syntax)
- modular structure: alternative models for peripherals (visual, motor) & other components possible

Showcasing the framework

Grodner and Gibson (2005, Exp. 1), also used in Lewis and Vasishth, 2005

- Grodner and Gibson (2005, Exp. 1): self-paced reading, matrix subject is modified by a subject or object-extracted relative clause (RC)

(1) The <u>reporter who sent the photographer to the editor hoped</u> for a story.

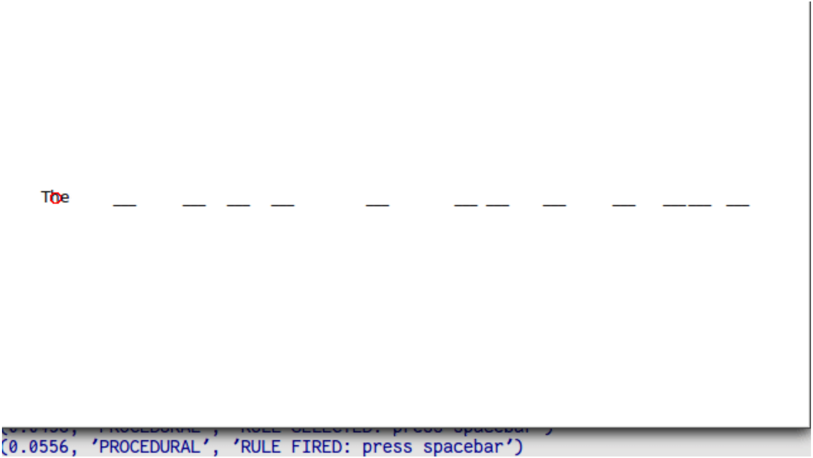
(2) The <u>reporter who the photographer sent to the editor hoped</u> for a story.

9 ROIs: word 2 through word 10 (underlined above)

Demo of an ACT-R model for subj and obj gap RCs

(open the slides with Adobe Acrobat Reader to see the movie)

Red circle is the visual focus. Temporal trace incrementally produced by the model is visible in the background.



Memory in ACT-R

Two types of memory:

- DECLARATIVE MEMORY: knowledge of facts
  facts represented as chunks (attribute-value matrices)

ISA:	word
FORM:	car
MEANING:	⟦car⟧
CATEGORY:	noun
NUMBER:	sg

- PROCEDURAL MEMORY: behavior as a series of productions
  productions – conditionalized actions

Goal> | TASK: reading, FORM: car | ⇒ Goal> | TASK: retrieving category | Retrieval> | ISA: word, FORM: car |

Modules and buffers in ACT-R

- ACT-R mind is composed of modules, which include declarative and procedural memory
- Modules are not directly accessible – they can only be accessed through buffers
- Buffers represent agent's current state; productions fire based on contents of buffers
- Buffers can hold only one chunk
- Only one production can fire at any given time

An eager left-corner parser in ACT-R

Building on Lewis and Vasishth, 2005; Resnik, 1992

Parser components:

- lexical knowledge, knowledge of intermediate parse states / syntactic structures → declarative memory
- knowledge of grammar → procedural memory (common, e.g., Lewis and Vasishth 2005, but not only choice, e.g., Reitter et al. 2011)
- expectations about upcoming syntactic categories, which guide parsing → goal buffer
- information about the current syntactic parse → imaginal buffer
- visual information from environment → visual buffer
- key press commands → manual buffer
- visual module – EMMA Salvucci, 2001 (other choices possible)
- motor module – EPIC Kieras and Meyer, 1996; Meyer and Kieras, 1997

An eager left-corner parser in ACT-R

Rules:
S → NP VP
NP → Det N
VP → V

Visual input:

- A boy sleeps.

`A -- ----.`

Input

- Stack: S (Goal)
- Found: *a*, Det
  (Visual + Retrieval)

Output

- Stack: N NP S (Goal)
- Structure:

S
|
NP
|
Det
|
a

An eager left-corner parser in ACT-R

Rules:
S → NP VP
NP → Det N
VP → V

Visual input:

- A boy sleeps.

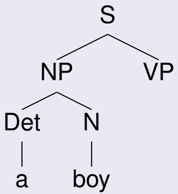
`- boy ----.`

Input

- Stack: N NP S (Goal)
- Found: *boy*, N
  (Visual + Retrieval)

Output

- Stack: VP (Goal)
- Structure:



An eager left-corner parser in ACT-R

Rules:
S → NP VP
NP → Det N
VP → V

Visual input:

- A boy sleeps.

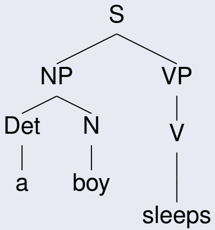
```
- -- sleeps.
```

Input

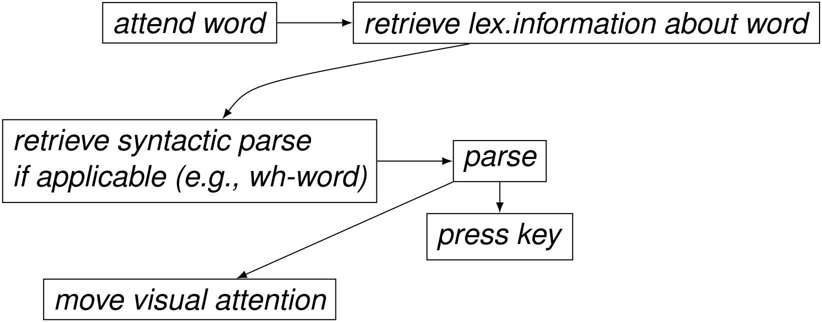
- Stack: VP (Goal)
- Found: *sleeps*, V
  (Visual + Retrieval)

Output

- Stack: {} (Goal)
- Structure:



Flow chart of parsing process per word



Parameters – visual encoding (EMMA)

- Visual encoding (T_{enc}) dependent on visual distance d and object properties, D

$$T_{enc} = K \cdot D \cdot e^{kd} \text{(parameter } k \text{ – angle)}$$

- D = word length, $K = 0.01$
- k – estimated to show that parameters for peripherals can be estimated at the same time as the more commonly estimated parameters associated with declarative and procedural memory

Parameters – rule firing and memory recall

- Rule firing = r (parameter r)
- Retrieval latency is a function of activation, modulated by parameters F (latency factor) and f (latency exponent)

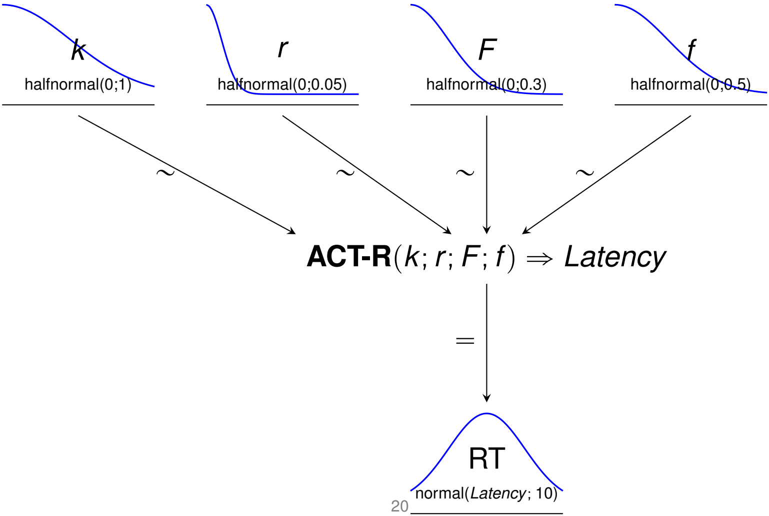
$$T = F \cdot e^{-f \cdot A}$$

- from chunk activation, A, we only consider base activation, which is a function of time elapsed since previous word usages
- estimating r – needed because our processing models incorporate linguistic theories in a fairly transparent way; so: we need to fire more rules per word / region of interest (ROI) than possible with the 50 ms default; production compilation should increase r closer to its ACT-R default
- latency exponent f – estimated because crucial in estimating latencies in lexical decision tasks (e.g., the tasks in Murray and Forster 2004)

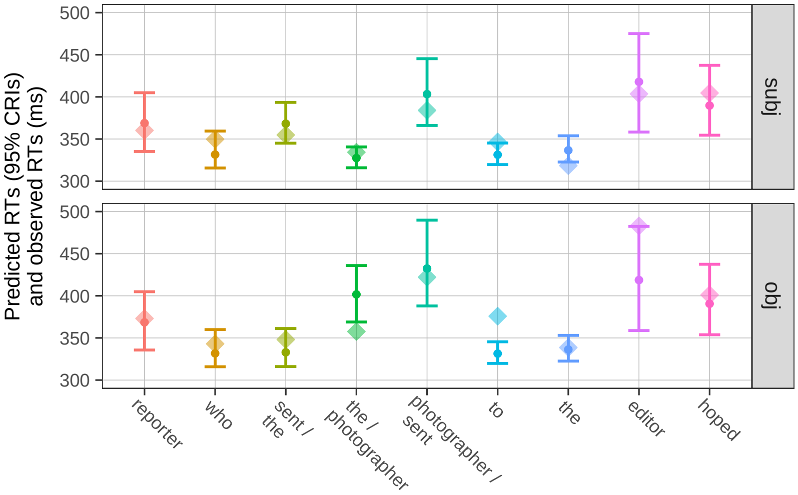
Estimation

- The model is fit to data by estimating the 4 free parameters (k, r, F, f)
- Standardly, relying on default values or manually changing the values; subjective & time consuming
- **pyactr** enables us to easily interface ACT-R models with standard statistical estimation methods implemented in widely-used Python3 libraries
- we use ACT-R models as the likelihood component of full Bayesian models, and fit the ACT-R parameters to experimental data

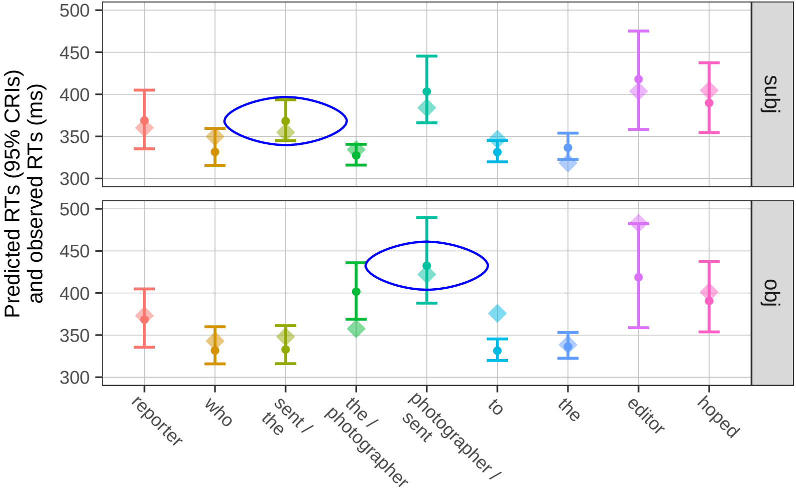
Bayesian model structure



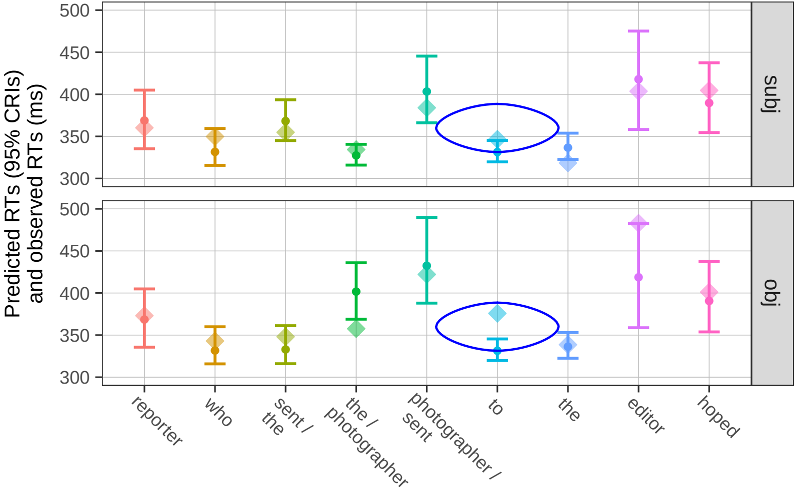
Posterior predictions (Model 1)



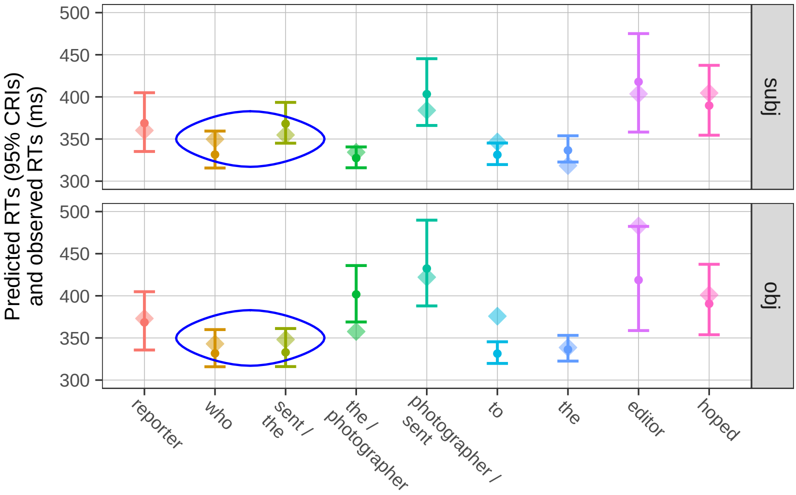
Posterior predictions (Model 1)



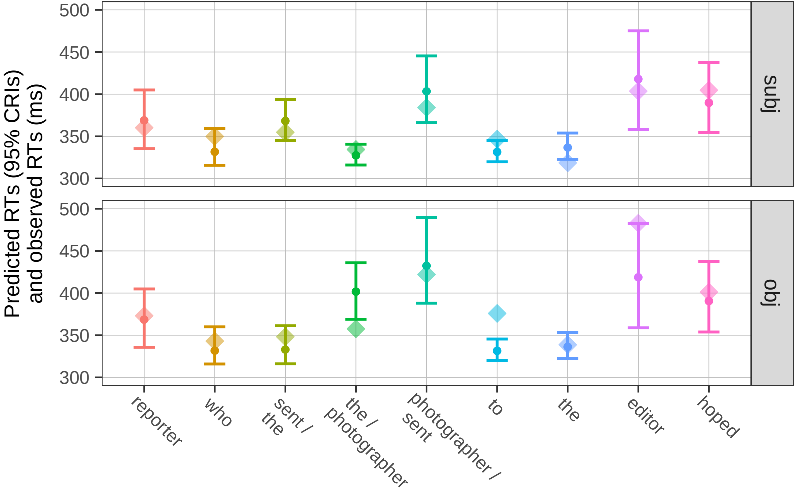
Posterior predictions (Model 1)



Posterior predictions (Model 1)

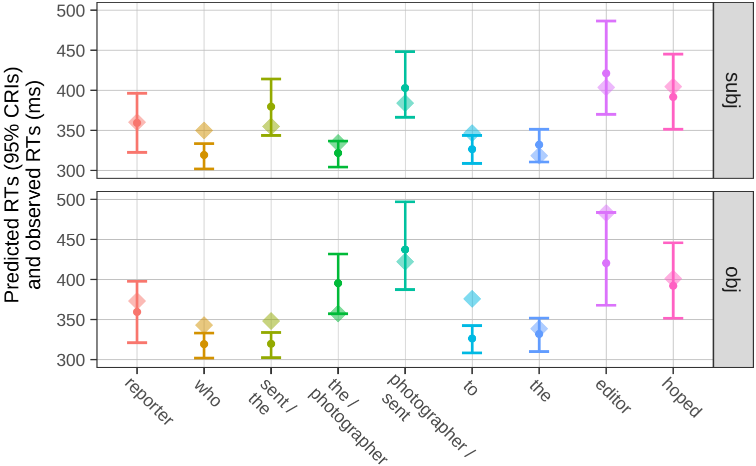


Posterior predictions (Model 1)

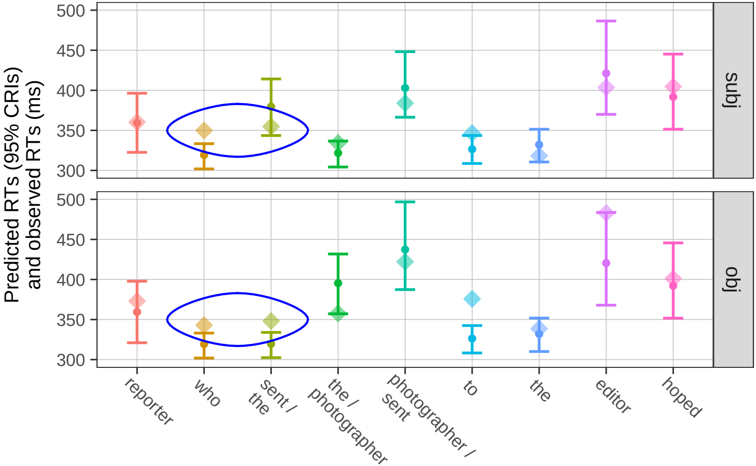


Model 2: no postulated subject gaps

Model 2: no postulated subject gaps

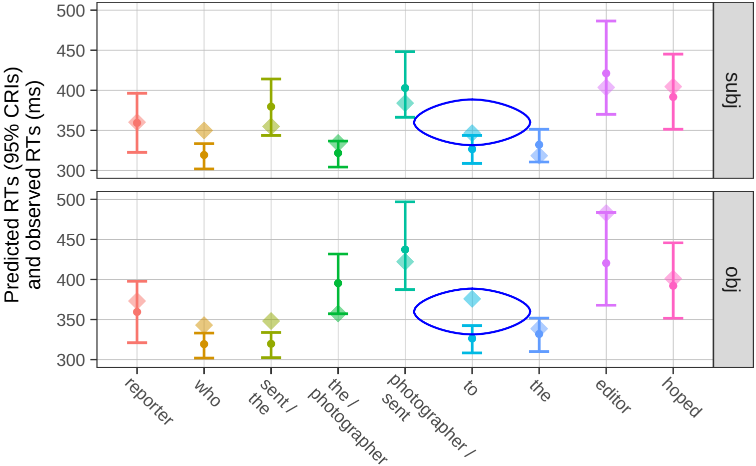


Model 2: no postulated subject gaps

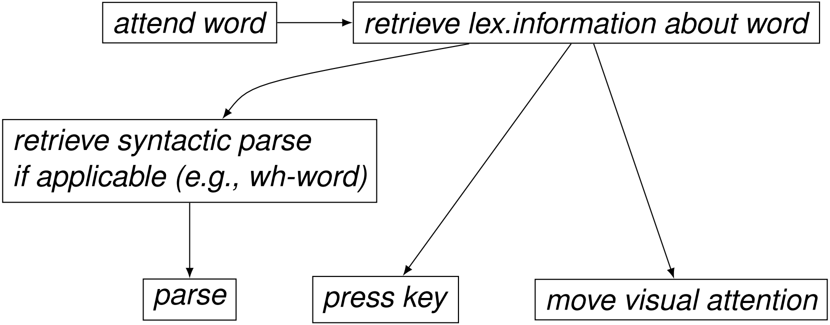


wh-word and following word not modeled well; Model 1 better

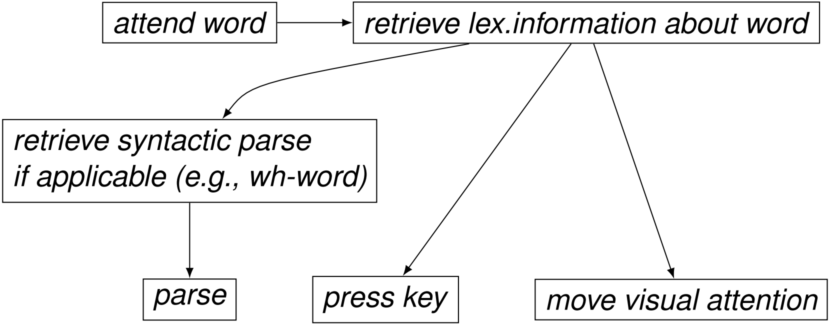
Model 2: no postulated subject gaps



Model 3: parallel reader

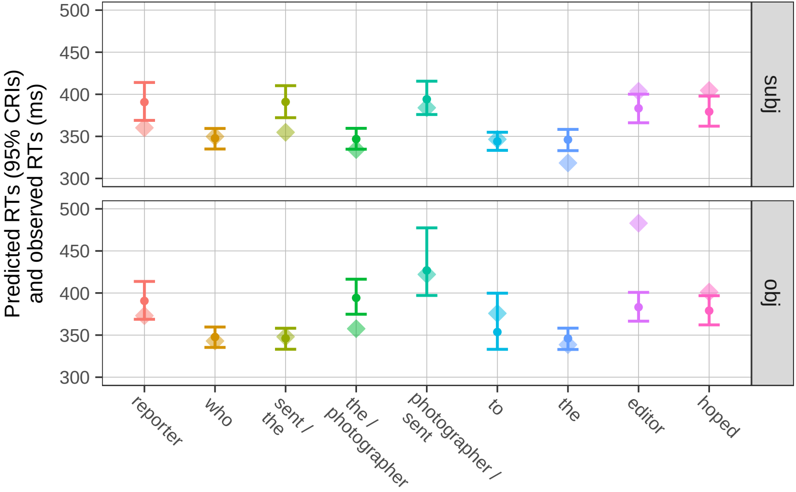


Model 3: parallel reader

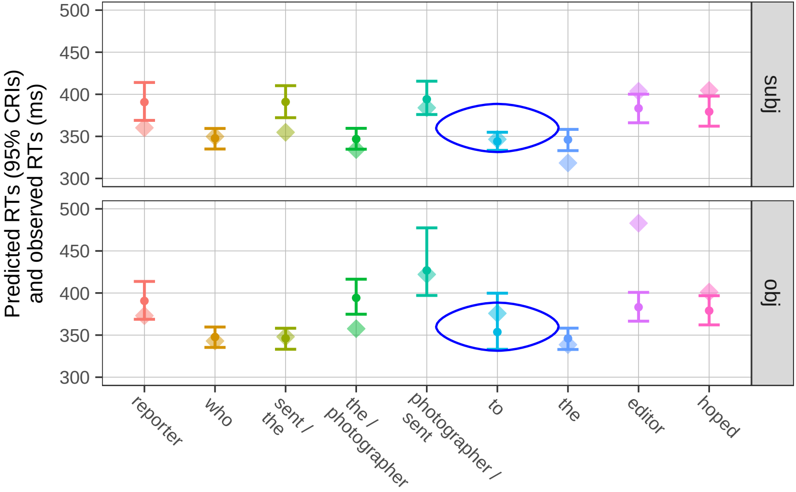


- Model 1 completes all available parsing before key press (serial)
- Model 3: first lexical retrieval, then structure building & key press in parallel
- Outcome: spillover on word after object gap captured

Model 3: spillover after object gap captured



Model 3: spillover after object gap captured



WAIC-based model comparison

	$WAIC_1$	$WAIC_2$
Model 1 (subject gaps)	388	1469
Model 2 (no subject gaps)	433	1613
Model 3 (‘parallel’ reader)	390	553

- Model 1 is better than Model 2 with respect to both $WAIC_1$ and $WAIC_2$
- increase in precision for Model 3 is clearly visible in its much lower $WAIC_2$ value (which is variance based)

Conclusion

- we introduced a modular and extensible framework for mechanistic processing models
- case study: an incremental left-corner parser with visual and motor interfaces for subject/object gap relative clauses
- framework used to quantitatively compare hypotheses about processing, e.g., predictively postulating subject gaps
- systematic across-the-board model comparison via Bayes factors is possible in this framework
- framework can model other tasks (eye tracking, lexical decision)
- individual differences can be modeled by adding random effects

Acknowledgments

We are grateful to Amanda Rysling, Donka Farkas, Abel Rodriguez, Matt Wagers, the UCSC S-lab audience (March 2018) and our ICCM 2018 anonymous reviewers for comments and discussion.

We want to thank Ted Gibson and Dan Grodner for providing the items and full datasets for the two experiments reported in their paper (Grodner and Gibson 2005). Jakub Dotlačil was supported by the NWO VENI 275-80-005 grant. The usual disclaimers apply.

References I

- Brasoveanu, Adrian and Jakub Dotlačil (in prep.). *Formal Linguistics and Cognitive Architecture*. Language, Cognition, and Mind (LCAM) Series. The *pyactr* library (Python3 ACT-R) is available here: `https://github.com/jakdot/pyactr`. Dordrecht: Springer.
- Dillon, Brian et al. (2013). "Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence". In: *Journal of Memory and Language* 69.2, pp. 85–103.
- Engelmann, Felix et al. (2013). "A Framework for Modeling the Interaction of Syntactic Processing and Eye Movement Control". In: *Topics in Cognitive Science* 5.3, pp. 452–474. DOI: `10.1111/tops.12026`.
- Grodner, Daniel and Edward Gibson (2005). "Consequences of the Serial Nature of Linguistic Input for Sentenial Complexity". In: *Cognitive Science* 29, pp. 261–291.
- Hale, John T. (2014). *Automaton Theories of Human Sentence Comprehension*. Stanford: CSLI Publications.

References II

- Kamp, Hans (1981). “A Theory of Truth and Semantic Representation”. In: *Formal Methods in the Study of Language*. Ed. by Jeroen Groenendijk et al. Amsterdam: Mathematical Centre Tracts, pp. 277–322.
- Kamp, Hans and Uwe Reyle (1993). *From Discourse to Logic. Introduction to Model theoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Dordrecht: Kluwer.
- Kieras, David E and David E Meyer (1996). “The EPIC architecture: Principles of operation”. Unpublished manuscript from ftp://ftp. eecs. umich. edu/people/kieras/EPICarch. ps.
- Kush, Dave W (2013). “Respecting relations: Memory access and antecedent retrieval in incremental sentence processing”. PhD thesis. University of Maryland, College Park.
- Lewis, Richard and Shravan Vasishth (2005). “An activation-based model of sentence processing as skilled memory retrieval”. In: *Cognitive Science* 29, pp. 1–45.

References III

Meyer, David E and David E Kieras (1997). "A computational theory of executive cognitive processes and multiple-task performance: Part I. Basic mechanisms.". In: *Psychological review* 104.1, p. 3.

Murray, Wayne S and Kenneth I Forster (2004). "Serial mechanisms in lexical access: the rank hypothesis.". In: *Psychological Review* 111.3, p. 721.

Nicenboim, Bruno and Shravan Vasishth (2018). "Models of retrieval in sentence comprehension: A computational evaluation using Bayesian hierarchical modeling". In: *Journal of Memory and Language* 99, pp. 1–34. DOI: https://doi.org/10.1016/j.jml.2017.08.004.

Reitter, David et al. (2011). "A computational cognitive model of syntactic priming". In: *Cognitive science* 35.4, pp. 587–637.

Resnik, Philip (1992). "Left-corner parsing and psychological plausibility". In: *Proceedings of the Fourteenth International Conference on Computational Linguistics*. Nantes, France.

Rij, Jacolien van (2012). *Pronoun processing: Computational, behavioral, and psychophysiological studies in children and adults*. Groningen.

References IV

- Salvucci, Dario D (2001). "An integrated model of eye movements and visual encoding". In: *Cognitive Systems Research* 1.4, pp. 201–220.
- Taatgen, Niels A and John R Anderson (2002). "Why do children learn to say "broke"? A model of learning the past tense without feedback". In: *Cognition* 86.2, pp. 123–155.
- Vasishth, Shravan et al. (2008). "Processing Polarity: How the Ungrammatical Intrudes on the Grammatical". In: *Cognitive Science* 32, pp. 685–712.
- Young, Richard M. and Richard L. Lewis (1999). "The Soar cognitive architecture and human working memory". In: ed. by Akira Miyake and Priti Shah, pp. 224–256.

Model 1: parameter estimates

- $k = 0.87, sd = 0.32$
- $F = 0.01, sd = 0.03$
- $f = 0.23, sd = 0.47$
- $r = 0.02, sd = 0.006$