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

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

(2) The reporter who the photographer sent to the editor hoped 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.
1 Introduction: framework & case study
2 ACT-R & left-corner parsing
3 Results
4 Conclusion
1 Introduction: framework & case study

2 ACT-R & left-corner parsing

3 Results

4 Conclusion
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  \[\Rightarrow\]  Goal> TASK: retrieving category  \[\Rightarrow\]  Retrieval> ISA: word FORM: car
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)
An eager left-corner parser in ACT-R

Rules:
S → NP VP
NP → Det N
VP → V

Visual input:
A boy sleeps.

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

Output
- Stack: N NP S (Goal)
- Structure:
  \[
  S \rightarrow NP \rightarrow Det \rightarrow a
  \]
An eager left-corner parser in ACT-R

Rules:
S → NP VP
NP → Det N
VP → V

Visual input:
A boy sleeps.

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

Output
- Stack: VP (Goal)
- Structure:

```
    S
   / \  /
  NP   VP
  /   /
Det N
  /   /
 a   boy
```
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:

```
   S
   /\  
  NP  VP
  /\   
 Det N V
   /\  /
  a boy
    
  sleeps
```
Flow chart of parsing process per word

1. **attend word**
2. **retrieve lex. information about word**
3. **retrieve syntactic parse if applicable (e.g., wh-word)**
4. **parse**
5. **press key**
6. **move visual attention**
Parameters – visual encoding (EMMA)

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

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

\[ \text{ACT-R}(k; r; F; f) \Rightarrow Latency \]

\[ \text{RT} \sim \text{normal} (\text{Latency}; 10) \]
1 Introduction: framework & case study

2 ACT-R & left-corner parsing

3 Results

4 Conclusion
Posterior predictions (Model 1)

Predicted RTs (95% CRIs) and observed RTs (ms)
Posterior predictions (Model 1)

Predicted RTs (95% CRIs) and observed RTs (ms)
Posterior predictions (Model 1)
Posterior predictions (Model 1)
Posterior predictions (Model 1)

Predicted RTs (95% CRIs) and observed RTs (ms)

- Subj
- Obj

- Reporter
- Who
- Sent / the
- The / photographer
- Photographer /
- Sent
- To
- The
- Editor
- Hoped
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

1. **Attend word**
2. **Retrieve lexical information about word**
   - **Retrieve syntactic parse** if applicable (e.g., wh-word)
3. **Parse**
4. **Press key**
5. **Move visual attention**

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 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

![Graph showing predicted and observed RTs (ms) for subj and obj with 95% CRIs.](image-url)
Model 3: spillover after object gap captured
WAIC-based model comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>WAIC$_1$</th>
<th>WAIC$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1 (subject gaps)</td>
<td>388</td>
<td>1469</td>
</tr>
<tr>
<td>Model 2 (no subject gaps)</td>
<td>433</td>
<td>1613</td>
</tr>
<tr>
<td>Model 3 (‘parallel’ reader)</td>
<td>390</td>
<td>553</td>
</tr>
</tbody>
</table>

- 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)
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.


Model 1: parameter estimates

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