Computational Psycholinguistics

Lecture 12: Constraint-based Models and the Ambiguity Advantage

Harm Brouwer
The Competition-Integration Model (CIM)
Processing a sentence

The cop [arrested by] the detective was guilty of taking bribes

McRae (1998, JML)
Processing a sentence

The cop arrested by [the detective] was guilty of taking bribes

> Weight mass is equally divided between old and new constraints

McRae (1998, JML)
The cop arrested by the detective [was guilty] of taking bribes

> Weight mass is equally divided between old and new constraints

McRae (1998, JML)
An eye-tracking experiment

I read that the bodyguard of the governor retiring after the troubles is very rich [ambiguous]

I read that the governor of the province retiring after the troubles is very rich [disambiguated: NP1/high-attachment]

I read that the province of the governor retiring after the troubles is very rich [disambiguated: NP2/low-attachment]

I read quite recently that the governor retiring after the troubles is very rich [unambiguous]
# The Ambiguity Advantage

## Experiment 2: means

<table>
<thead>
<tr>
<th></th>
<th>Disambiguating region</th>
<th>Post-disambiguation region</th>
<th>Final region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First-pass reading times (ms)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>378 (10)</td>
<td>552 (16)</td>
<td>851 (22)</td>
</tr>
<tr>
<td>High attachment</td>
<td>354 (11)</td>
<td>574 (19)</td>
<td>840 (25)</td>
</tr>
<tr>
<td>Low attachment</td>
<td>356 (9)</td>
<td>570 (17)</td>
<td>842 (23)</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>364 (11)</td>
<td>555 (17)</td>
<td>841 (26)</td>
</tr>
<tr>
<td><strong>First-pass regressions (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>12.1 (2.3)</td>
<td>13.6 (2.3)</td>
<td>63.4 (3.4)</td>
</tr>
<tr>
<td>High attachment</td>
<td>9.5 (2.1)</td>
<td>16.0 (2.5)</td>
<td>64.4 (3.4)</td>
</tr>
<tr>
<td>Low attachment</td>
<td>8.4 (2.0)</td>
<td>23.6 (2.9)</td>
<td>69.1 (3.2)</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>9.5 (2.1)</td>
<td>16.7 (2.6)</td>
<td>56.1 (3.6)</td>
</tr>
<tr>
<td><strong>Regression-path times (ms)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>441 (16)</td>
<td>723 (35)</td>
<td>2046 (116)</td>
</tr>
<tr>
<td>High attachment</td>
<td>420 (18)</td>
<td>754 (33)</td>
<td>2166 (122)</td>
</tr>
<tr>
<td>Low attachment</td>
<td>423 (19)</td>
<td>801 (34)</td>
<td>2330 (137)</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>436 (20)</td>
<td>708 (25)</td>
<td>1945 (108)</td>
</tr>
<tr>
<td><strong>Total times (ms)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ambiguous</td>
<td>542 (21)</td>
<td>797 (31)</td>
<td>1065 (35)</td>
</tr>
<tr>
<td>High attachment</td>
<td>578 (25)</td>
<td>880 (37)</td>
<td>1103 (34)</td>
</tr>
<tr>
<td>Low attachment</td>
<td>601 (27)</td>
<td>899 (33)</td>
<td>1073 (36)</td>
</tr>
<tr>
<td>Unambiguous</td>
<td>550 (25)</td>
<td>789 (27)</td>
<td>1019 (33)</td>
</tr>
</tbody>
</table>

*Notes.* The regions were as follows (delimited by brackets): I read that the bodyguard of the governor[retiring][after the troubles][is very rich.] Standard errors are in parentheses.
The Ambiguity Advantage (cont’d)

Hence, the one million dollar question is: Who is right?
The Ambiguity Advantage (cont’d)

Van Gompel et al. (2005, pg. 287):
“competition in the globally ambiguous sentences can never be weaker than in the disambiguated sentences, so the globally ambiguous sentences can never be easier to process”

Green and Mitchell (2006, pg. 10):
“the model predicts an ambiguity advantage for materials with a certain range of biases and the reverse in other cases”

> G&M’s argument is based on simulations
Q: Why are the disambiguated conditions averaged together?

“For purposes of presentation (to avoid an otherwise very cluttered graph), the values of (b) and (c) were then averaged for each inherited bias.” (G&M, 2006, pg. 9)
G&M - Simulation 3: Decomposed

Green and Mitchell (2006, JML)

Brouwer (2010), MSc thesis
Fitz, Brouwer & Hoeks (in prep.)
Interim Conclusions

CIM does not predict an ambiguity advantage on a per-item basis (and G&M are wrong)

However, ambiguity advantage is not found on a per-item basis, but by averaging over different items (as in common practice in psycholinguistic research)

> Hence, maybe the CIM does predict an ambiguity advantage if we average over different items?
G&M - Simulation 5

> Sample 24 random starting biases for the pre-critical region from $N(0.5,0.1)$

> Process the pre-critical region, thereby establishing a bias

> Inherit established bias, and process the critical region

Green and Mitchell (2006, JML)
G&M - Simulation 5: Results

G&M ran three simulations (= 3 x 24 items), and reported average cycles per condition for each of these

<table>
<thead>
<tr>
<th></th>
<th>Simulation 1</th>
<th>Simulation 2</th>
<th>Simulation 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambiguous</td>
<td>12.1</td>
<td>11.4</td>
<td>11.6</td>
</tr>
<tr>
<td>NP1-attachment</td>
<td>23.8</td>
<td>26.5</td>
<td>23.23</td>
</tr>
<tr>
<td>NP2-attachment</td>
<td>22.3</td>
<td>21.5</td>
<td>24.7</td>
</tr>
</tbody>
</table>

Contrasts between ambiguous and each of the disambiguated conditions yielded six (3x2) $F$ values ranging between $F(1,23) = 7.32$ and $F(1,23) = 23.33$. All $p$-values < .015.

> CIM does predict an ambiguity advantage when averaging over items (as in the VG et al. experiment)
Decomposing the results

Q1: What happens in the pre-critical region?

Starting biases for the pre-critical region are randomly sampled from an $N(0.5,0.1)$ distribution; assume an item with biases [0.51,0.49]

---

Model state after: 16 processing cycle(s)

Threshold: 0.880

Alternative [alternative1]: 0.928
Alternative [alternative2]: 0.072

Input node [cst: constraint1] [alt: alternative1] [wgt: 1.000]: 1.789
Input node [cst: constraint1] [alt: alternative2] [wgt: 1.000]: 0.077

Threshold [0.880] reached after: 16 processing cycle(s)

After the pre-critical region:

Alternative1 bias:
$\frac{1.789}{1.789 + 0.077} = 0.96$

Alternative2 bias:
$\frac{0.077}{1.789 + 0.077} = 0.04$

> Small imbalances are amplified during processing (strong imbalances even more so), and become strong biases for the next region
Decomposing the results (cont’d)

Q2: What happens in the critical region?

Given samples from $N(0.5,0.1)$ and the effect of bias amplification in the pre-critical region, we know that:

50% of the items fall in the far left of this graph, and 50% in the far right.

Hence, disambiguated items confirm these biases half of the time ($\rightarrow$ little competition), and disconfirm them the other half ($\rightarrow$ strong competition).
Decomposing the results (cont’d)

Ambiguous:
\[
\frac{12 \times \text{med} + 12 \times \text{med}}{24} \approx 12
\]

NP1 attachment:
\[
\frac{12 \times \text{high} + 12 \times \text{low}}{24} \approx 25
\]

NP2 attachment:
\[
\frac{12 \times \text{low} + 12 \times \text{high}}{24} \approx 25
\]

> Results rely on \( N(0.5, 0.1) \)
Balanced materials?

\[ N(0.5,0.1) \] implies that the materials in the pre-critical region are perfectly balanced regarding NP1- and NP2-attachment

**Q: Is this a fair assumption?**

Off-line questionnaires and completion tasks, as well as on-line studies suggest that there is a preference for NP2-attachment (e.g., Frazier & Clifton, 1996; Carreiras & Clifton, Fernandez, 2003) (but see also Traxler, Pickering, & Clifton, 1998)

> How does this affect the ambiguity advantage?
NP2-attachment preference

NP2-attachment preference can be modeled by sampling the starting biases for the pre-critical region from $N(0.75, 0.1)$

> For this sample, none of the items supports NP1-attachment
NP2-attachment preference (cont’d)

Ambiguous:
(24 x med) / 24 ≈ 12

NP1 attachment:
(24 x high) / 24 ≈ 55

NP2 attachment:
(24 x low) / 24 ≈ 2

> No ambiguity advantage, but an NP2-attachment advantage
Discussion

> Whether or not the CIM predicts an ambiguity advantage (on average) depends on modeling choices

> Hence, whether G&M or VG et al. are right, depends on what you believe to happen in the pre-critical region

> When modeling a specific effect, we should take into account that psycholinguistic effects are typically found in averages

> Even the simplest models (such as the CIM) often make unforeseen predictions; which is why we need modeling!
Conclusions

CIM does not predict an ambiguity advantage on a per-item basis (and G&M are still wrong in that respect)

CIM does predict an ambiguity advantage when averaging over items (and VG et al. are wrong in this respect)

... but only if there is no (strong) bias imbalance in the pre-critical region
Relevant References


