The Same-head Heuristic for Coreference

Micha Elsner and Eugene Charniak

Department of Computer Science
Brown University

July 9, 2010
Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, ‘and what is the use of a book,’ thought Alice ‘without pictures or conversation?’
Same-head coreference

If two NPs have the same head, they are coreferent.

A natural starting point:

- Easy to code
- Can be very good in some experimental conditions
- Most work focuses on hard cases
  - Non-matching NPs
  - Pronouns

However, the heuristic doesn’t always work!
Unsupervised systems

Unsupervised work uses the same-head heuristic.

- (Haghighi+Klein ‘07): sparse prior on $p(\text{word}|\text{entity})$
- (Poon+Domingos ‘08): head-prediction clause
- (Haghighi+Klein ‘09): direct assumption

Why do they do this?
In this talk

Mention detection and scoring matter

Non-coreferent same-head pairs

Modeling
Gold mentions

- Anything marked by a MUC annotator
- Small subset of NPs
- Annotators don’t mark singleton NPs!

However, the Multiplication Table doesn’t signify: let’s try Geography. London is the capital of Paris, and Paris is the capital of Rome—no, THAT’S all wrong, I’m certain!

All NPs

However, the Multiplication Table doesn’t signify: let’s try Geography. London is the capital of Paris, and Paris is the capital of Rome—no, THAT’S all wrong, I’m certain!
What about metrics?

$b^3$ (Bagga+Baldwin ‘98)

- More important to get the big clusters right

CEAF (Luo ‘05)

- No precision/recall tradeoff
## Comparison

### Gold mentions/$b^3$

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect resolution</td>
<td>48.8</td>
<td>45.5</td>
</tr>
<tr>
<td>Same-head heuristic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3% gap looks unimportant

### NPs/CEAF

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect resolution</td>
<td>73.4</td>
<td>62.2</td>
</tr>
<tr>
<td>Same-head heuristic</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

10% gap looks substantial
Quick survey: the MUC data

Did some counting:

- MUC-6 dev
- 100 random pairs: same head, not coreferent
- Ad-hoc categories

<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two different entities</td>
<td>39</td>
</tr>
<tr>
<td>Time/measure phrase (&quot;three years&quot;)</td>
<td>24</td>
</tr>
<tr>
<td>Quantified and similar (&quot;most Senators&quot;)</td>
<td>12</td>
</tr>
<tr>
<td>Generics (&quot;during a campaign&quot;)</td>
<td>12</td>
</tr>
<tr>
<td>Others</td>
<td>12</td>
</tr>
</tbody>
</table>

Syntactic context and modifiers often disambiguate.
Modeling: coreference as alignment

Possible antecedent:

The slot for the new NP:

- Unsupervised
- Log-linear model
- Learned via EM
## Results

<table>
<thead>
<tr>
<th></th>
<th>Mentions</th>
<th>Linked</th>
<th>Mention CEAF</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perfect resolution</td>
<td>3993</td>
<td>864</td>
<td>73.4</td>
</tr>
<tr>
<td>Our model</td>
<td>3993</td>
<td>518</td>
<td>67.0</td>
</tr>
<tr>
<td>Heuristic</td>
<td>3993</td>
<td>1592</td>
<td>62.2</td>
</tr>
</tbody>
</table>

- System halves error in CEAF
- Fewer NPs linked
- However, $b^3$ declines
Conclusions from analysis

- Experimental setup matters:
  - Use realistic mention detector
  - Report multiple measures
- Modeling can help!

Come see the poster!

Thanks Google, BLLIP, Jean Carletta, Dan Jurafsky and Mark Johnson