The Same-head Heuristic for Coreference

Micha Elsner and Eugene Charniak

Department of Computer Science
Brown University

July 9, 2010

Same-head coreference

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

Same-head heuristic

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(word|entity)
- (Poon+Domingos '08): head-prediction clause
- (Haghighi+Klein '09): direct assumption
- partial exception: (Ng '08)

Why do they do this?

In this talk

Mention detection and scoring matter

Non-coreferent same-head pairs

Modeling

Gold mentions

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

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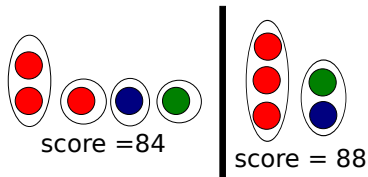
All NPs
However, the Multiplication Table doesn't signify: let's try Geography. London is the capital of Paris, and Paris is the capital of Rome, and Rome— no, THAT'S all wrong, I'm certain!
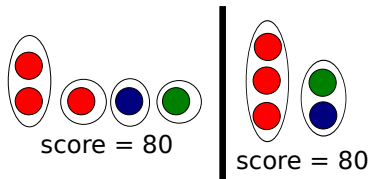
What about metrics?

b^3 (Bagga+Baldwin '98)

- More important to get the big clusters right

score =84 | score = 88

CEAF (Luo '05)

- No precision/recall tradeoff

score = 80 | score = 80

Comparison

Gold mentions/b^3

Perfect resolution for same-heads: **48.8**
Same-head heuristic: **45.5**

3% gap looks unimportant

NPs/CEAF

Perfect resolution for same-heads: **73.4**
Same-head heuristic: **62.2**

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

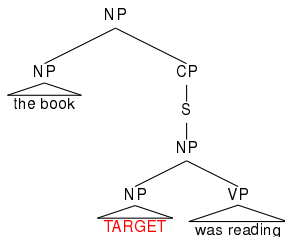|	
Two different entities	39
Time/measure phrase ("three years")	24
Quantified and similar ("most Senators")	12
Generics ("during a campaign")	12
Others	12

Syntactic context and modifiers often disambiguate.

Modeling: coreference as alignment

Possible antecedent:
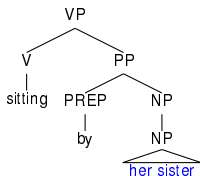
The slot for the new NP:

- ▶ Unsupervised
- ▶ Log-linear model
- ▶ Learned via EM

Results

	Mentions	Linked	Mention CEAF
NPs			
Perfect resolution	3993	864	73.4
Our model	3993	518	**67.0**
Heuristic	3993	1592	**62.2**

- ▶ 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