Bootstrapping Path-Based Pronoun Resolution

Shane Bergsma
University of Alberta

Dekang Lin
Google, Inc.

COLING-ACL 2006

The Path to Success

- Pronoun Resolution:
  - Determine which preceding noun each pronoun refers to
- Example:

  "John needs his friend"

  "John needs his support"
- Objective:
  - Learn likelihood of coreference along syntactic paths

The Path of the Talk

1. Pronoun resolution:
   - Traditional Approach and Limitations
2. Learning Path Coreference
3. Using Path Coreference
4. Evaluation and Results
5. Related Work
6. Summary

1. Pronoun Resolution: The Path Less Travelled

- "In **2005**, **Exxon Mobil** paid ***its*** **Chairman Lee Raymond $48.9 million"**
- Goal: determine the *antecedent* for the pronoun "its" – establish *coreference*
- Why? Information Retrieval, Document Summarization, Machine Translation
- "Who is the chairman of Exxon Mobil?"

Resolution Strategy

- "In **2005**, **Exxon Mobil** paid ***its*** **Chairman Lee Raymond $48.9 million**"

1. Parse the text to determine the noun phrases
2. Build a list of previous nouns as potential candidates
3. Filter candidates based on gender/number agreement, grammar violations, etc.
4. Select most likely candidate based on frequency, emphasis, etc.

Limitations

- Example:

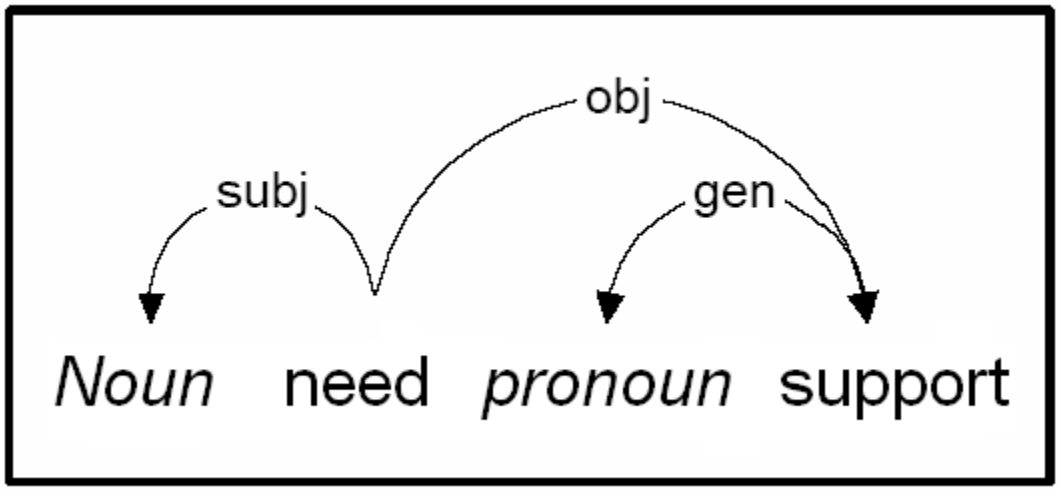
  "John needs his friend"

  "John needs his support"

- Need to know "world-knowledge":
  - "John" can be masculine
  - "John's friend" and "John's support" are compatible
  - People don't need their own support*

Dependency Path

- Sequence of dependency links between terminal entities



- Short form: *Noun* needs *pronoun's* support

2. Learning Path Coreference

- Goal: learn likelihood of coreference along a given dependency path
- Extract dependency paths from raw text
- For each path instance with terminals that are pronouns:
  - If they are from same pronoun class, coreference is likely.
  - If the terminals are from a different pronoun class, coreference is unlikely.

Counting Instances

I need *his* support	Disagree
They need *my* support	Disagree
We need *her* support	Disagree
She needs *her* back support	Agree
I really need *your* support	Disagree

Agreement = 20% ⇨ a *non-coreferent* path

Some cool ones:

- Example non-coreferent paths:
  - "*John* married *his* mother."
  - "*Sue* wrote *her* obituary."
- Example coreferent paths:
  - "*The newspaper* says *it* intends to…"
  - "*The revolutionaries* consolidated *their* power."
- Especially useful for phrases not covered by traditional syntactic constraints

3. Using Path Coreference

- Use it directly as a feature in your system
- Enhancing a semantic compatibility model
- Use it to bootstrap probabilistic noun gender/number information:
  - "*The newspaper* says *it* intends to…"
  - Assume coreference, count as instance of "newspaper" being neutral

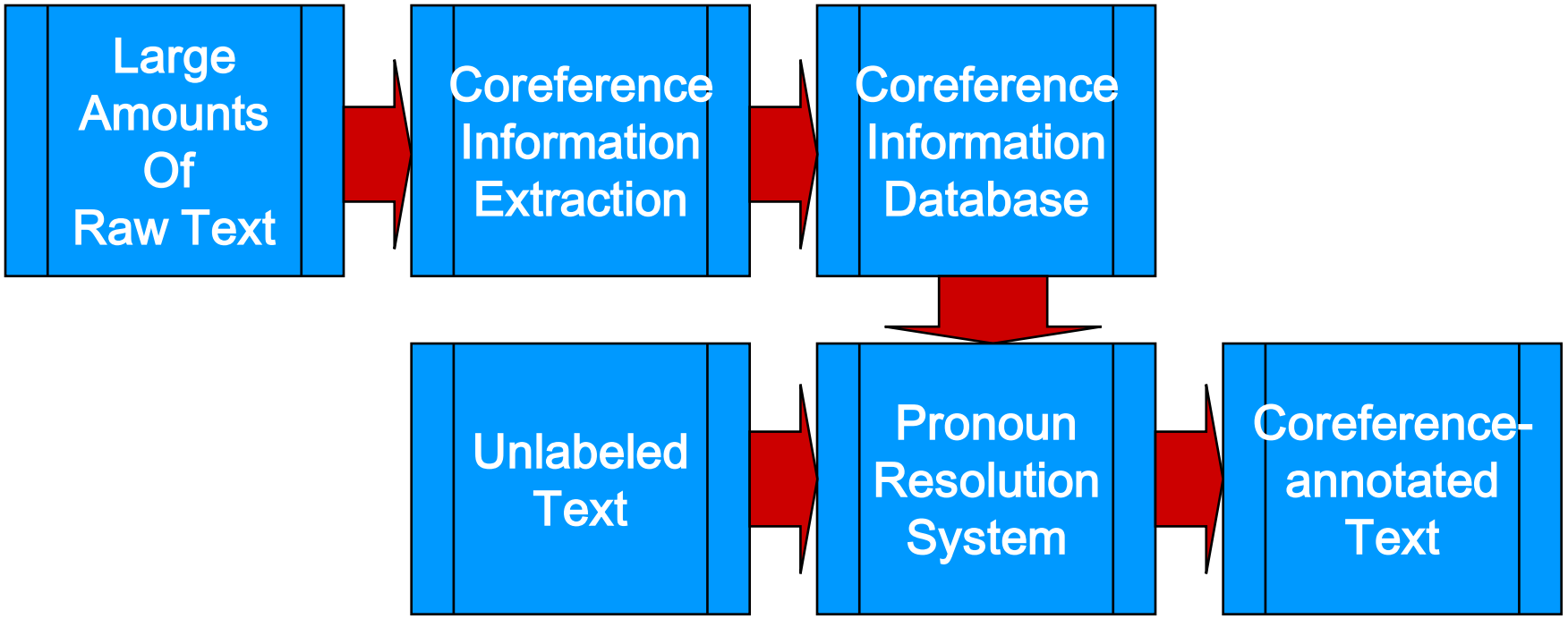
Probabilistic Gender Examples

- Produces very good, very clean data:

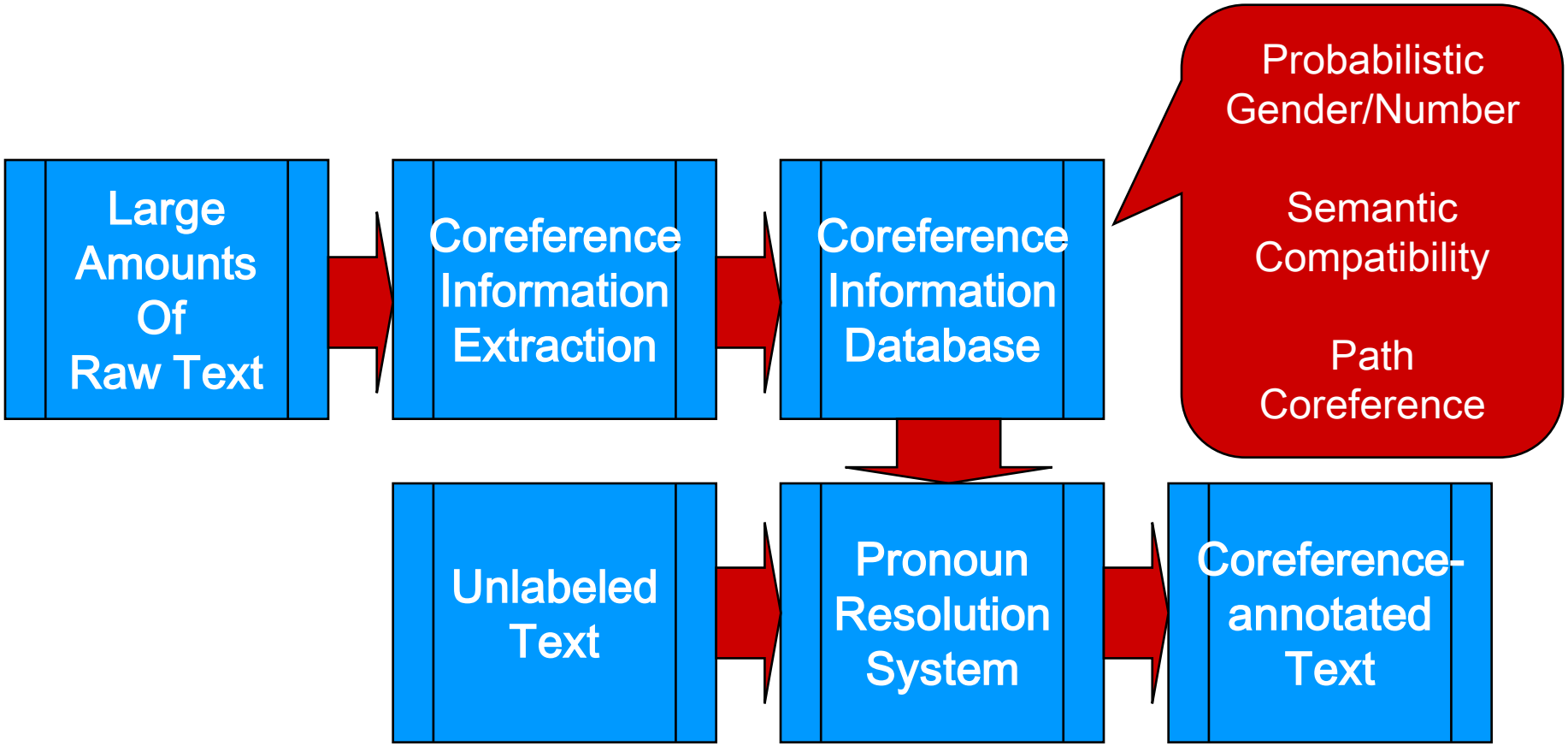
Word	*masc*	*fem*	*neut*	*plur*
company	0.6	0.1	98.1	1.2
condoleeza rice	4.0	92.7	0.0	3.2
pat	58.3	30.6	6.2	4.9
president	94.1	3.0	1.5	1.4
wife	9.9	83.3	0.8	6.1

- Download your own copy today!

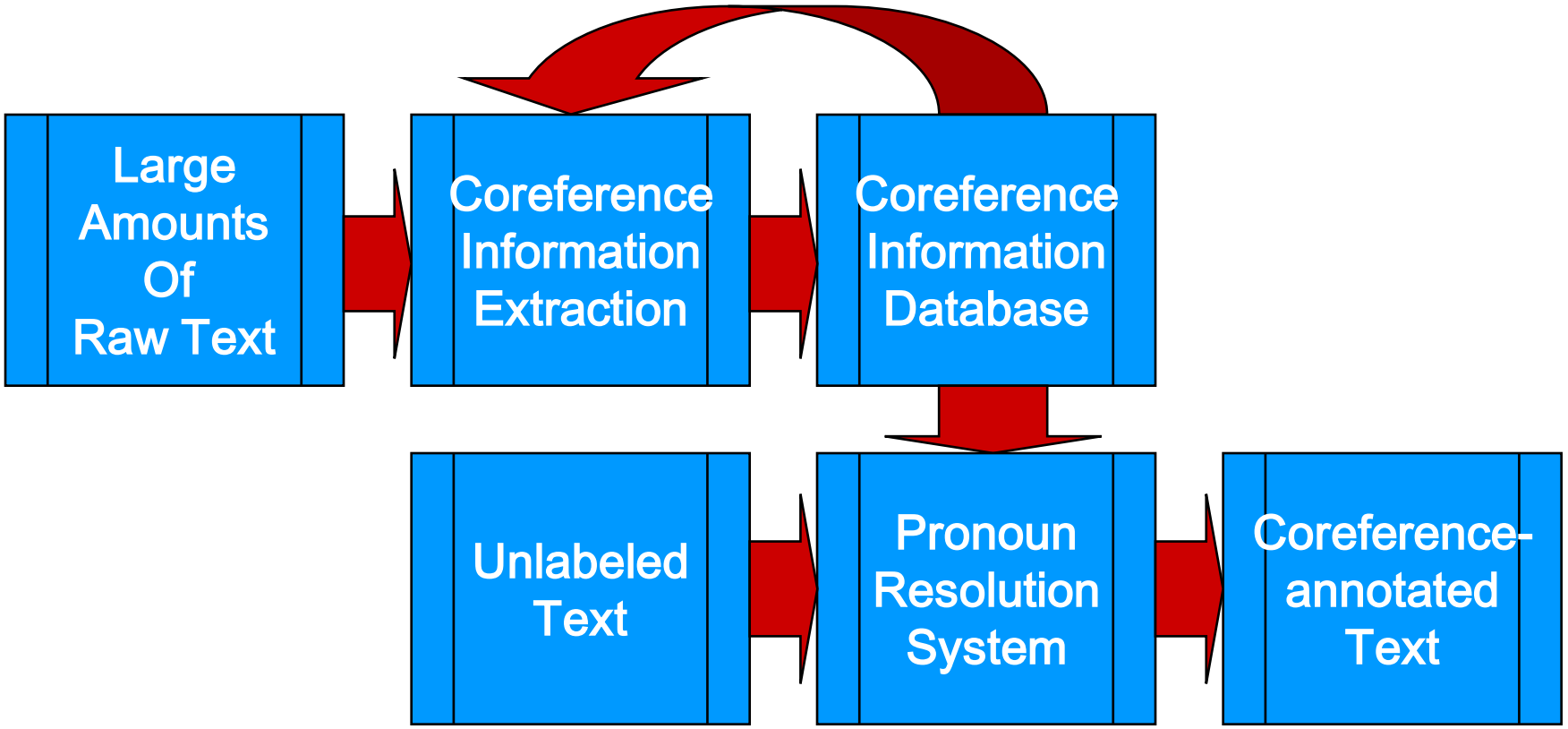
Bootstrapping Pronoun Resolution



Bootstrapping Pronoun Resolution



Bootstrapping Pronoun Resolution



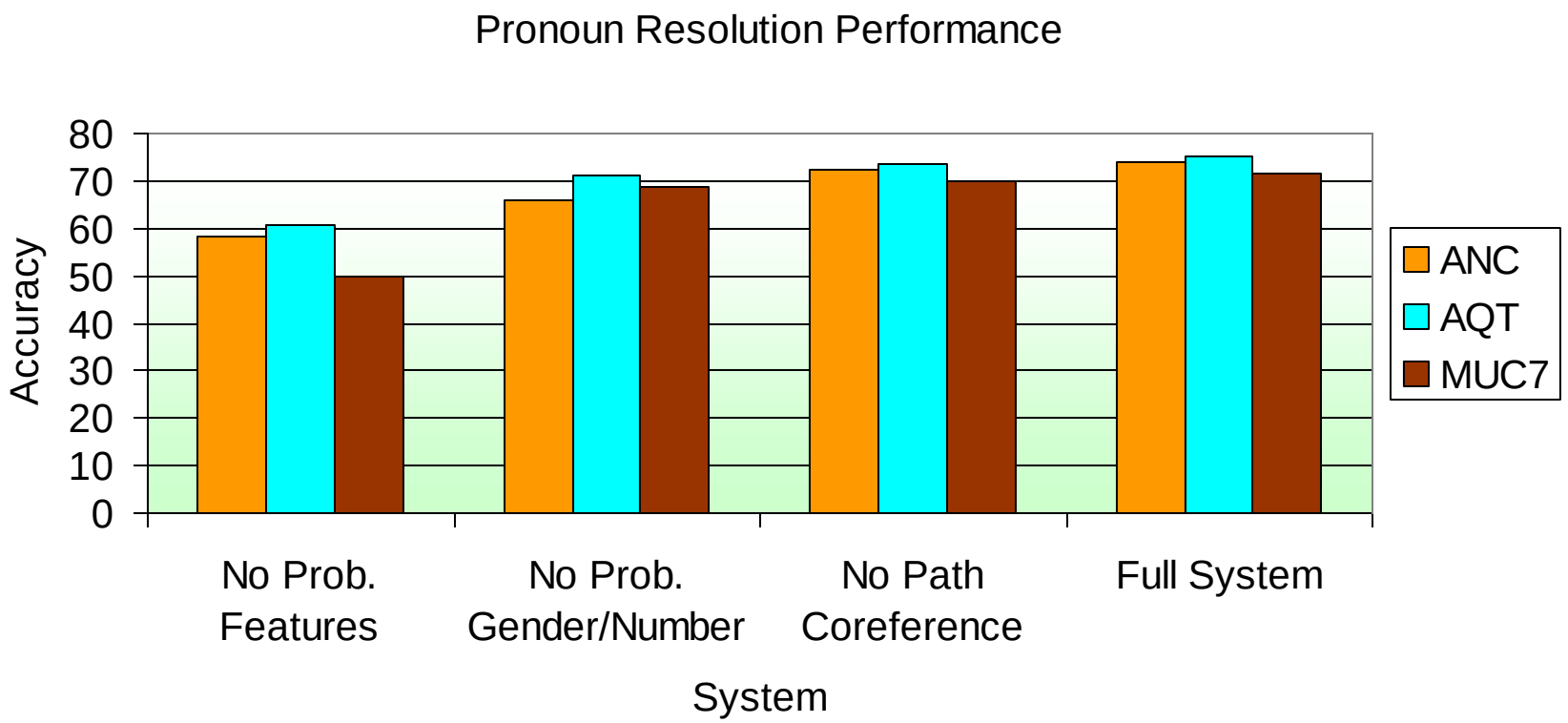
4. Evaluation

- Test system: machine-learned (SVM) pronoun resolution classifier
- Typical features: candidate distance, frequency, grammatical role, etc.
- World-knowledge (probabilistic) features:
  - path coreference, gender/number, mutual information of candidate with pronoun's parse tree parent

Evaluation

- Use paths with surface words (as shown above) *and* ones with grammatical tags:
  - *Noun* V *pronoun* "*John saw him.*"
  - *Noun* BE *pronoun*'s NP "Mary was her friend."
  - Noun V *pronoun*-self "It hurt itself."
- Probabilistic information acquired on a large amount of data:
  - 85 GB of news articles taken from the World Wide Web over a 2-year period

Results


Pronoun Resolution Performance
Accuracy
80
70
60
50
40
30
20
10
0
No Prob. Features
No Prob. Gender/Number
No Path Coreference
Full System
System
ANC
AQT
MUC7

5. Related Work

- Ge et al. (1998)
  - Gender information from unlabelled text
- Lin and Pantel (2001)
  - Paraphrases by comparing dependency paths
- Ng and Cardie (2002)
  - Features for Machine-Learned Coreference Resolution
- Yang et al. (2005)
  - Semantic compatibility (World-Knowledge)

6. Summary

- Path coreference:
  - a powerful feature for pronoun resolution
  - helps in situations not handled by previous resolution strategies
  - can be used to bootstrap very precise probabilistic gender/number information
- Check out the paper for more cool stuff

Thanks

- Thank you to:
  - Google Inc. for providing the internship
  - the very helpful COLING-ACL reviewers
  - the NLP group at the University of Alberta
  - our Funding bodies: