Identifying Anaphoric and Non-Anaphoric Noun Phrases to Improve Coreference Resolution

Vincent Ng and Claire Cardie
Department of Computer Science
Cornell University

Plan for the Talk

- Noun phrase coreference resolution
  - general machine learning approach
  - baseline coreference resolution system
- Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  - why anaphoricity info can help coreference resolution
  - general machine learning approach
  - anaphoricity determination system
- Using anaphoricity information in coreference resolution

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

> Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

> Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

> Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

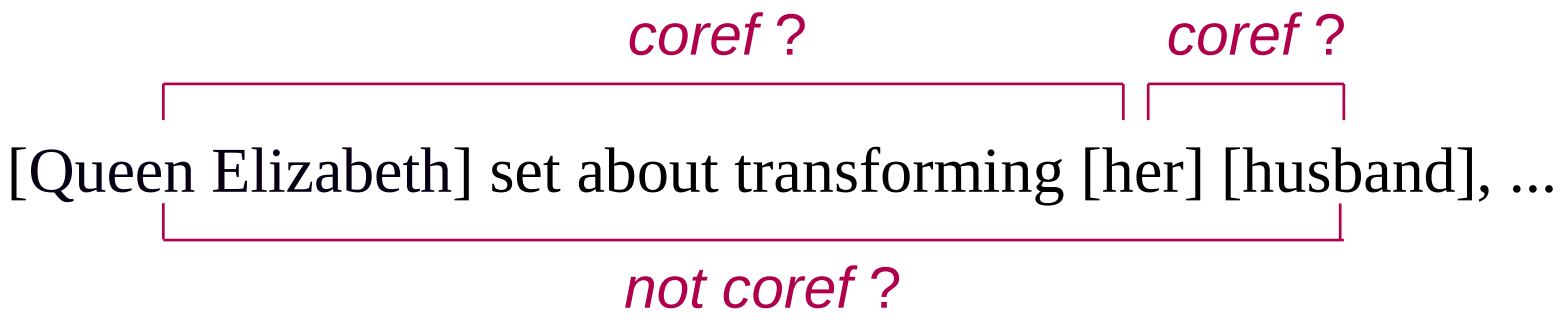
Noun Phrase Coreference

Identify all noun phrases that refer to the same entity

Queen Elizabeth set about transforming her husband, King George VI, into a viable monarch. Logue, a renowned speech therapist, was summoned to help the King overcome his speech impediment...

A Machine Learning Approach

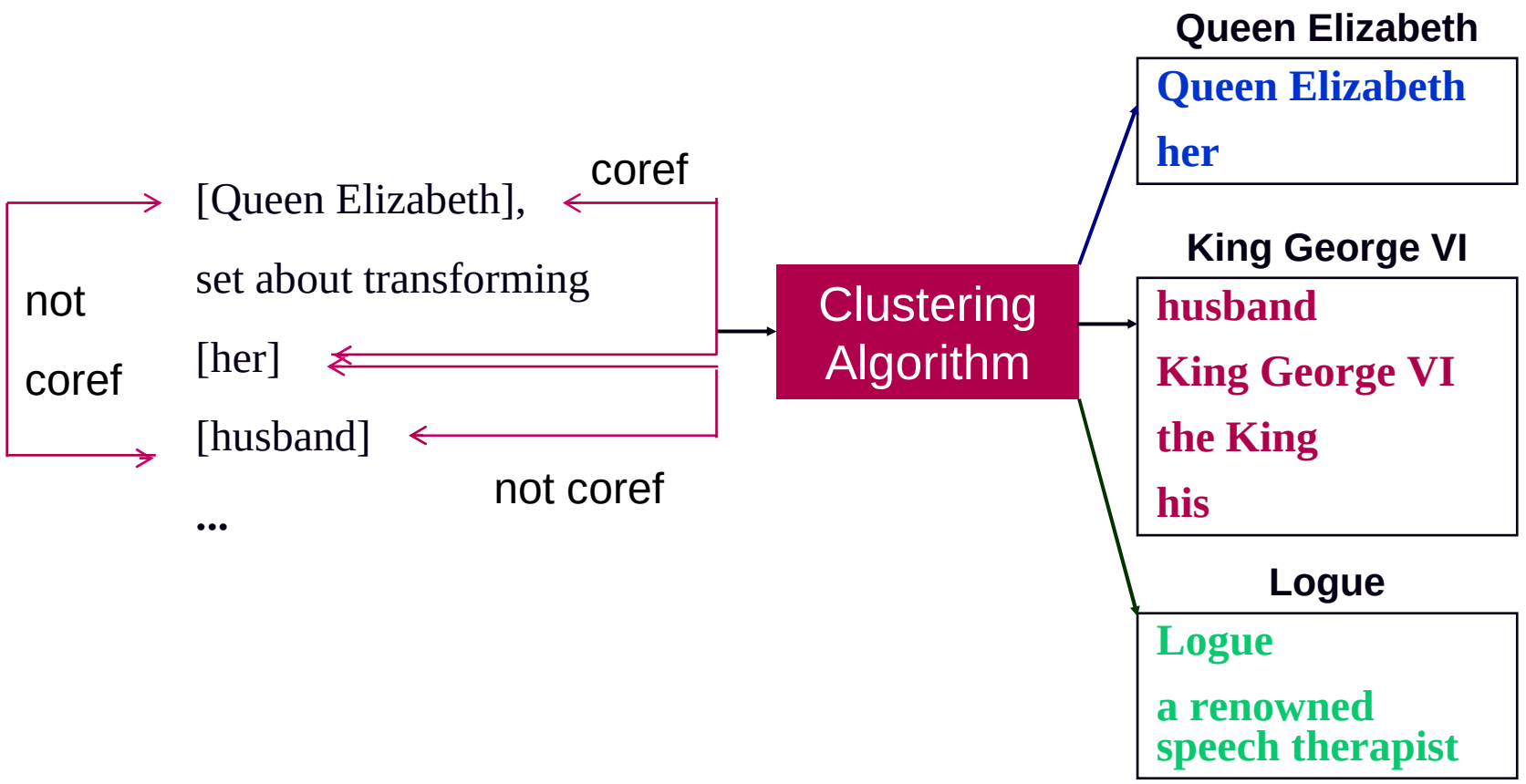
- Classification
  - given a description of two noun phrases, NP_i and NP_j, classify the pair as *coreferent* or *not coreferent*



Aone & Bennett [1995]; Connolly et al. [1994];
McCarthy & Lehnert [1995]; Soon, Ng & Lim [2001]

A Machine Learning Approach

- Clustering
  - coordinates pairwise coreference decisions



Machine Learning Issues

- Training data creation
- Instance representation
- Learning algorithm
- Clustering algorithm

[Ng and Cardie, ACL'02]

Baseline System: Training Data Creation

- Creating training instances
  - texts annotated with coreference information
  - one instance for each pair of noun phrases
    - feature vector: describes the two NPs and context
    - class value:

coref	pairs on the same coreference chain
not coref	otherwise

  - use sampling to deal with skewed class distributions

Baseline System: Instance Representation

- 53 features per instance

Lexical (9)	NP string matching operations
Semantic (6)	Semantic compatibility tests, aliasing
Positional (2)	Distance in terms of number of sentences/paragraphs
Knowledge-based (2)	Naïve pronoun resolution, rule-based coref resolution
Grammatical (34)	NP type Grammatical role Linguistic constraints Linguistic preferences Heuristics

Baseline System: Learning Algorithm

- C4.5 (Quinlan, 1993): decision tree induction
- Classifier outputs coreference likelihood

Baseline System: Clustering Algorithm

- Best-first single-link clustering algorithm
  - selects as antecedent the NP with the highest *coreference likelihood* from among preceding coreferent NPs for each noun phrase

Baseline System: Evaluation

- MUC-6 and MUC-7 coreference data sets
- documents annotated w.r.t. coreference
- MUC-6: 30 training texts + 30 test texts
- MUC-7: 30 training texts + 20 test texts
- MUC scoring program
  - recall, precision, F-measure

Baseline System: Results

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	70.3	58.3	**63.8**	65.5	58.2	**61.6**
Best MUC System	59	72	65	56.1	68.8	61.8
Worst MUC System	36	44	40	52.5	21.4	30.4

Baseline System: Results

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	70.3	58.3	**63.8**	65.5	58.2	**61.6**
Best MUC System	59	72	**65**	56.1	68.8	**61.8**
Worst MUC System	36	44	**40**	52.5	21.4	**30.4**

Plan for the Talk

- Noun phrase coreference resolution
  - general machine learning approach
  - baseline coreference resolution system
- Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  - why anaphoricity info can help coreference resolution
  - general machine learning approach
  - anaphoricity determination system
- Using anaphoricity information in coreference resolution

Motivation

- Baseline coreference system
  - single-link clustering algorithm attempts to find an antecedent for *each* noun phrase

Motivation

- Baseline coreference system
  - single-link clustering algorithm attempts to find an antecedent for *each* noun phrase

- What we *really* want
  - single-link clustering algorithm attempts to find an antecedent for *each* anaphoric noun phrase

Motivation

- Baseline coreference system
  - single-link clustering algorithm attempts to find an antecedent for *each* noun phrase
- What we *really* want
  - single-link clustering algorithm attempts to find an antecedent for *each* anaphoric noun phrase
- Availability of anaphoricity info can increase the precision of the coreference system

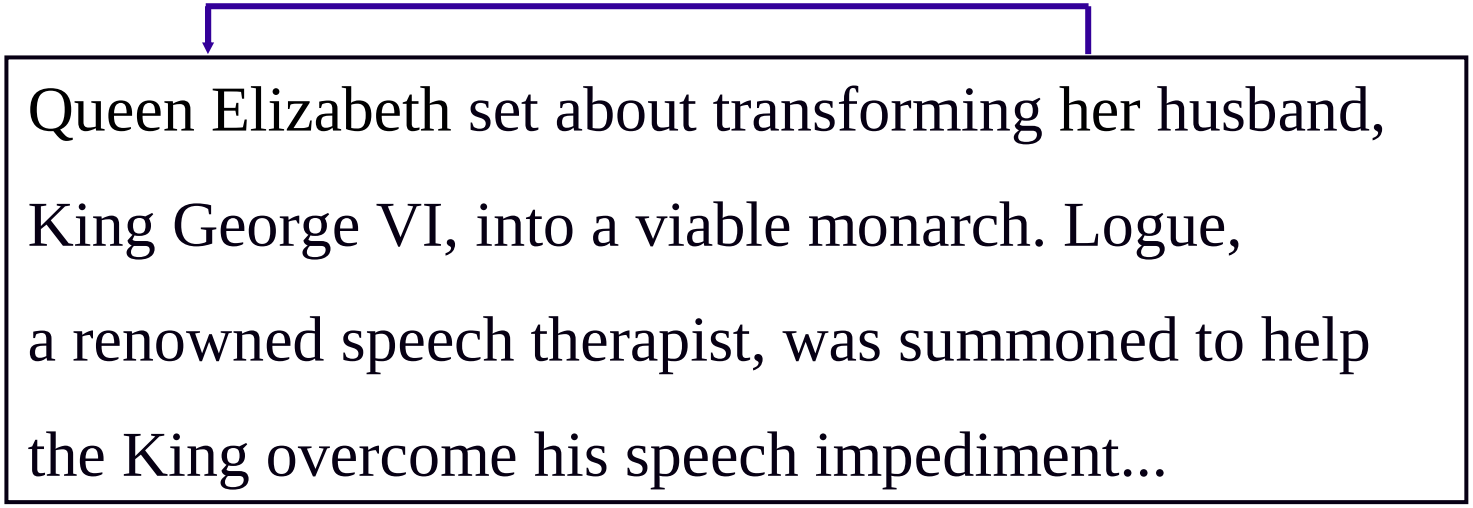
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

> Queen Elizabeth set about transforming her husband,
>
> King George VI, into a viable monarch. Logue,
>
> a renowned speech therapist, was summoned to help
>
> the King overcome his speech impediment...

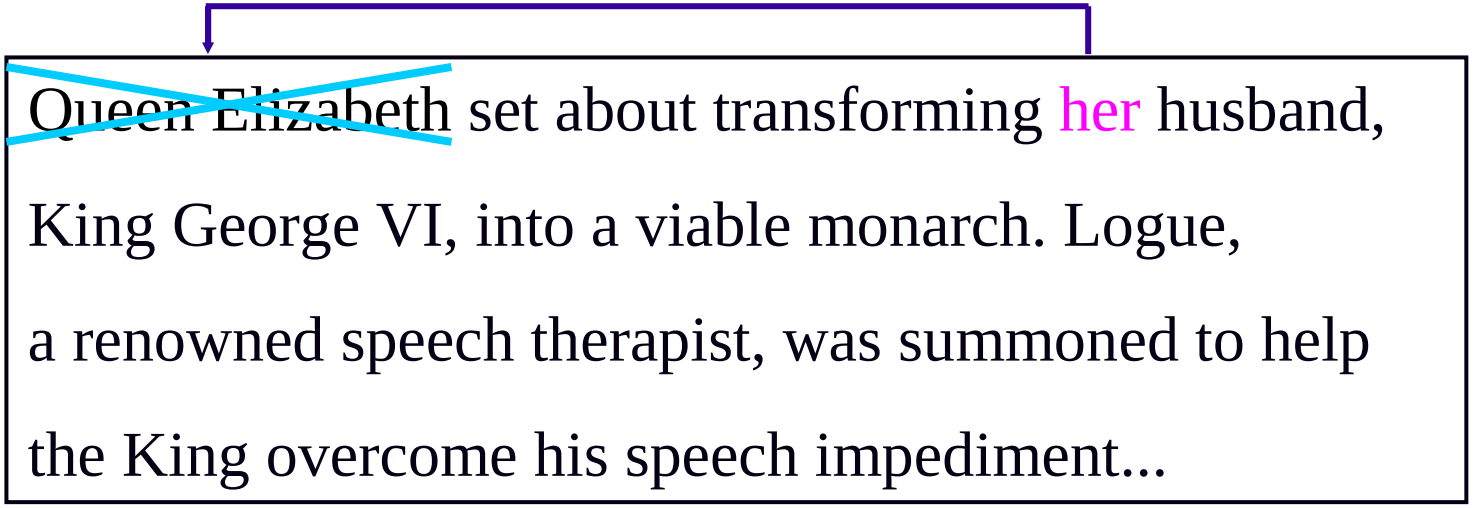
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.



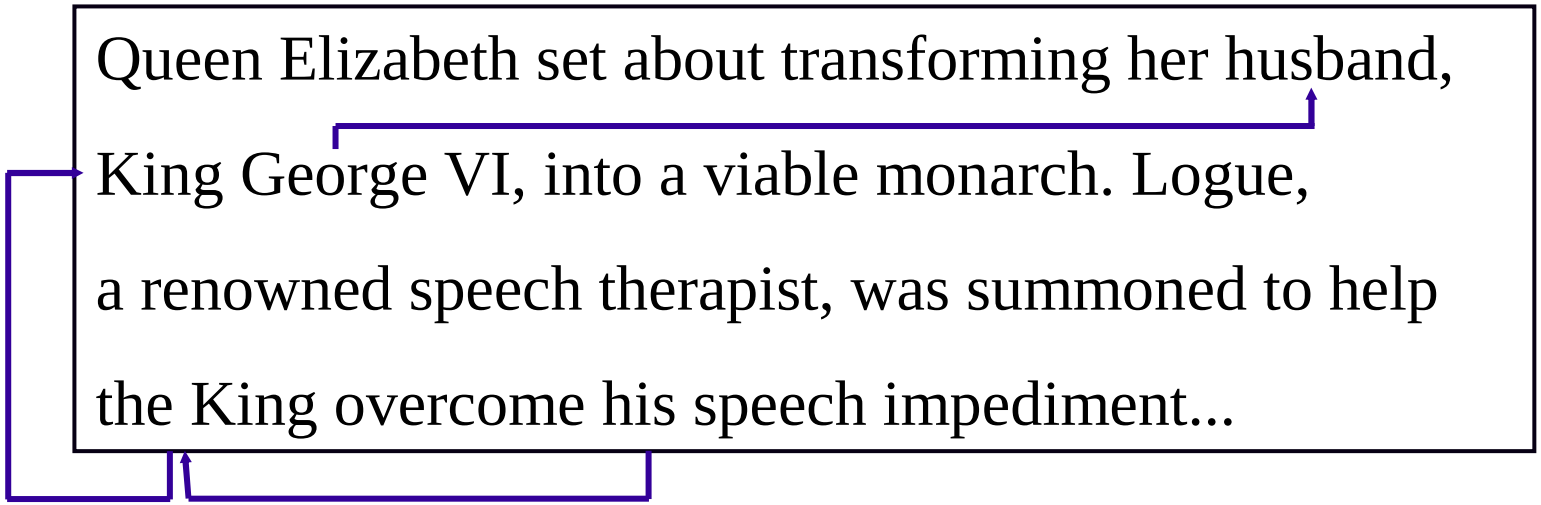
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.



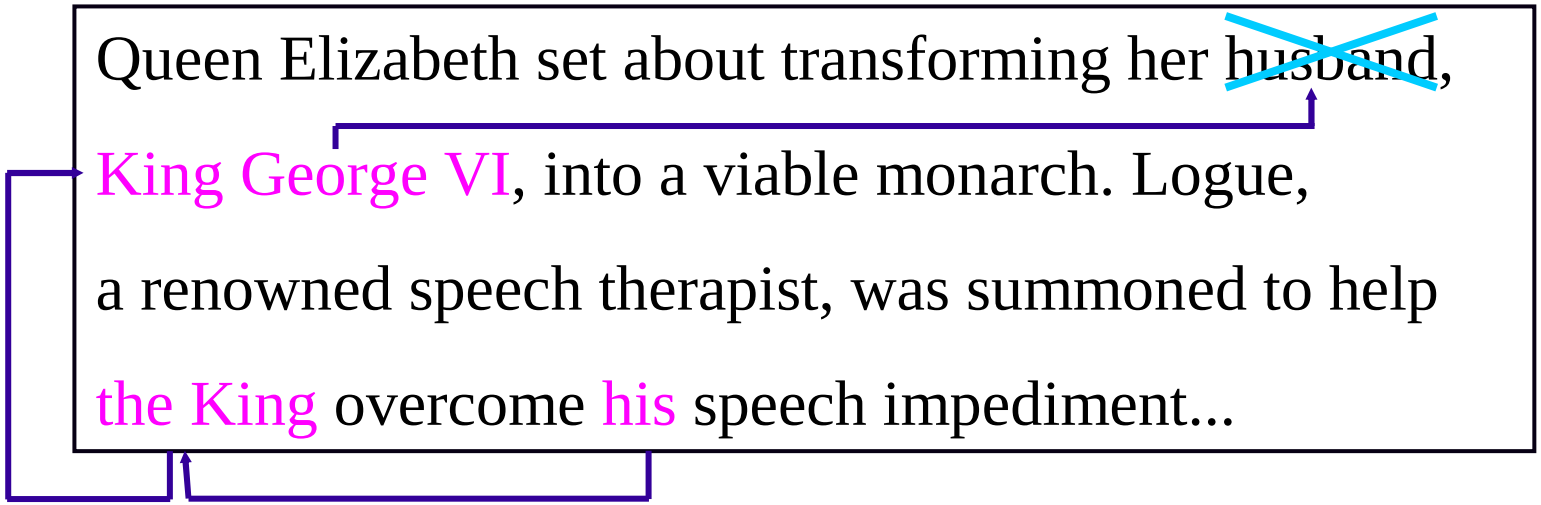
Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.



Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.



Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband,

King George VI, into a viable monarch. Logue,

a renowned speech therapist, was summoned to help

the King overcome his speech impediment...

Anaphoricity Determination

For each noun phrase in a text, determine whether it is part of a coreference chain but is not the head of the chain.

Queen Elizabeth set about transforming her husband,
King George VI, into a viable monarch. Logue,
a renowned speech therapist, was summoned to help
the King overcome his speech impediment...

A Machine Learning Approach

- Classification
  - given a description of a noun phrases, NP_i, classify NP_i as *anaphoric* or *not anaphoric*

non-anaphoric *anaphoric* *non-anaphoric*

[Queen Elizabeth] set about transforming [her] [husband], ...

Anaphoricity Determination System

- Training data creation
  - texts annotated with coreference information
  - one instance for each noun phrase

- Learning algorithm
  - C4.5

Anaphoricity Determination System

- Instance representation
  - 37 features per instance

Lexical (4)	case, string matching, head matching
Positional (3)	header, first sentence, first paragraph
Semantic (4)	title, aliasing, semantic compatibility
Grammatical (35)	NP type: definite, indefinite, bare plural NP property: pre-modified, post-modified, number Syntactic pattern: THE_N, THE_PN, THE_ADJ_N

Anaphoricity Determination System: Evaluation

- MUC-6 and MUC-7 coreference data sets

Corpus	Instances	% Negatives	Accuracy
MUC-6 test	4565	**66.3**	**86.1**
MUC-7 test	3558	**73.2**	**84.0**

Existing Approaches to Anaphoricity Determination

- Heuristic-based approaches
  - Paice and Husk (1987), Lappin and Leass (1994), Kennedy and Boguraev (1996), Denber (1998), Vieira and Poesio (2000)

- Machine learning approaches
  - **Unsupervised**: Bean and Riloff (1999)
  - **Supervised**: Evans (2001)

Comparison with Previous Work (I)

- Approaches to anaphoricity determination

Our Approach	Previous Approaches

Comparison with Previous Work (I)

- Approaches to anaphoricity determination

Our Approach	Previous Approaches
- focuses on common nouns	

Comparison with Previous Work (I)

- Approaches to anaphoricity determination

Our Approach	Previous Approaches
- focuses on common nouns - can operate on all types of noun phrases	

Comparison with Previous Work (I)

- Approaches to anaphoricity determination

Our Approach	Previous Approaches
focuses on common nouns	handle specific types of noun phrases only
can operate on all types of noun phrases	

Comparison with Previous Work (I)

- Existing anaphoricity determination algorithms address only specific types of NPs:
  - pleonastic pronouns
    - Paice and Husk (1987), Lappin and Leass (1994), Kennedy and Boguraev (1996), Denber (1998)
  - definite descriptions
    - Bean and Riloff (1999), Vieira and Peosio (2000)
  - anaphoric and non-anaphoric uses of *it*
    - Evans (2001)

Comparison with Previous Work (II)

- Using anaphoricity information in coreference resolution

Our Coref System	Previous Coref Systems

Comparison with Previous Work (II)

- Using anaphoricity information in coreference resolution

Our Coref System	Previous Coref Systems
- employs anaphoricity determination as a separate component	

Comparison with Previous Work (II)

- Using anaphoricity information in coreference resolution

Our Coref System	Previous Coref Systems
employs anaphoricity determination as a separate component	perform anaphoricity determination within the coreference system

Comparison with Previous Work (II)

- Most previous work performs anaphoricity determination *implicitly*
  - e.g. via a specific feature in the coreference system
  - One exception:
    - Harabagiu *et al*. (2001)
    - assumes perfect anaphoricity information
    - effectively employs a separate (manual) anaphoricity determination component

Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system

Our System	Previous Systems

Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system

Our System	Previous Systems
- evaluated as a standalone component	

Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system

Our System	Previous Systems
▪ evaluated as a standalone component	
▪ evaluated in the context of coreference resolution	

Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system

Our System	Previous Systems
evaluated as a standalone component	evaluated as a standalone component
evaluated in the context of coreference resolution	

Comparison with Previous Work (III)

- Evaluation of anaphoricity determination system

Our System	Previous Systems
evaluated as a standalone component	evaluated as a standalone component
evaluated in the context of coreference resolution	contribution to coreference resolution not evaluated

Comparison with Previous Work (III)

- Little previous work evaluates the effects of anaphoricity determination in anaphora/coreference resolution

Anaphoricity Determination System	Effects on Coref Resolution
Bean and Riloff (1999)	?
Denber (1998)	?
Evans (2001)	↓ **Mitkov *et al.* (2001)**
Kennedy and Boguraev (1996)	?
Lappin and Leass (1994)	?
Paice and Husk (1987)	?
Vieira and Poesio (2000)	↑

Plan for the Talk

- Noun phrase coreference resolution
  - general machine learning approach
  - baseline coreference resolution system
- Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  - why anaphoricity info can help coreference resolution
  - general machine learning approach
  - anaphoricity determination system
- Using anaphoricity information in coreference resolution

How can anaphoricity information be used?

- The clustering algorithm will only search for an antecedent for anaphoric noun phrases.

- Hypothesis
  - Anaphoricity information will improve precision

Anaphoricity Determination for Coref Resolution

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	70.3	**58.3**	63.8	65.5	**58.2**	61.6

- coreference system has fairly low precision

Results (Perfect Anaphoricity Information)

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	70.3	**58.3**	63.8	65.5	**58.2**	61.6
With perfect anaphoricity info	66.3	**81.4**	73.1	61.5	**83.2**	70.7

- perfect anaphoricity information can improve precision

Results (Learned Anaphoricity Information)

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	70.3	**58.3**	63.8	65.5	**58.2**	61.6
With learned anaphoricity info	57.4	**71.6**	63.7	47.0	**77.1**	58.4

- improvement in precision comes at the expense of significant loss in recall

What went wrong?

- Hypothesis 1
  - drop in recall and overall performance is caused by poor accuracy of anaphoricity classifier on *positive* instances

What went wrong?

- Hypothesis 1
  - drop in recall and overall performance is caused by poor accuracy of anaphoricity classifier on *positive* instances

- Accuracy of anaphoricity classifier
  - overall: 86.1% (MUC-6) and 84.0% (MUC-7)
  - positives only: 73.1% (MUC-6) and 66.2% (MUC-7)

- Anaphoricity classifier misclassifies 414 and 322 anaphoric entities as non-anaphoric for the MUC-6 and MUC-7 data sets, respectively

Need more accuracy?

- Hypothesis 1.1
  - accuracy levels of 66-73% on positive instances for anaphoricity determination are **not** adequate for improving coreference resolution

Need more accuracy?

- Hypothesis 1.1
  - accuracy levels of 66-73% on positive instances for anaphoricity determination are **not** adequate for improving coreference resolution

- Goal
  - improve the accuracy on positive instances

Need more accuracy?

- Hypothesis 1.1
  - accuracy levels of 66-73% on positive instances for anaphoricity determination are **not** adequate for improving coreference resolution
- Goal
  - improve the accuracy on positive instances
- How?

Improving Accuracy on Positive Instances

- Observations
  - **string matching** and **aliasing** are strong indicators of coreference

Improving Accuracy on Positive Instances

- Observations
  - **string matching** and **aliasing** are *strong* indicators of coreference
  - **string matching** and **aliasing** are *weaker* indicators of anaphoricity

Improving Accuracy on Positive Instances

- Observations
  - **string matching** and **aliasing** are *strong* indicators of coreference
  - **string matching** and **aliasing** are *weaker* indicators of anaphoricity

- Goal
  - ensure that anaphoric NPs involved in these two types of relations are correctly classified

Classification with Constraints

- Assume that an NP is anaphoric (and bypass the anaphoricity classifier) if anaphoricity is indicated by either the **string matching** or the **aliasing** constraint

- Accuracy on positive instances
  - no constraints: 73.1% (MUC-6) and 66.2% (MUC-7)
  - with constraints: 82.0% (MUC-6) and 80.8% (MUC-7)

Results (Classification with Constraints)

	MUC-6			MUC-7		
	R	P	F	R	P	F
Baseline	70.3	58.3	**63.8**	65.5	58.2	**61.6**
With anaphoricity (no constraints)	57.4	71.6	**63.7**	47.0	77.1	**58.4**
With anaphoricity (with constraints)	63.4	68.3	**65.8**	59.7	69.3	**64.2**

- large gains in precision and smaller drops in recall
- automatically acquired anaphoricity info can be used to improve the performance of coreference resolution

Results (Comparison with Best MUC Systems)

	MUC-6			MUC-7		
	R	P	F	R	P	F
With anaphoricity (with constraints)	63.4	68.3	**65.8**	59.7	69.3	**64.2**
Best MUC System	59	72	**65**	56.1	68.8	**61.8**

Results (Comparison with Perfect Anaphoricity)

	MUC-6			MUC-7		
	R	P	F	R	P	F
With anaphoricity (with constraints)	63.4	68.3	**65.8**	59.7	69.3	**64.2**
With perfect anaphoricity info	66.3	81.4	**73.1**	61.5	83.2	**70.7**

- substantial room for improvement in anaphoricity determination

Summary

- Presented a supervised learning approach for anaphoricity determination that can handle all types of NPs
- Investigated the use of anaphoricity information in coreference resolution
- Showed automatically acquired knowledge of anaphoricity can be used to improve the performance of a learning-based coreference system