Learning Noun Phrase Anaphoricity to Improve Coreference Resolution: Issues in Representation and Optimization

Vincent Ng

Department of Computer Science, Cornell University

Human Language Technology Research Institute, UT Dallas

Goal

Improve learning-based coreference systems using automatically acquired anaphoricity information

Plan for the Talk

- Noun phrase coreference resolution
  - standard machine learning approach
- Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  - why anaphoricity info can help coreference resolution
- Issues in computing and using anaphoricity information in coreference resolution

Noun Phrase Coreference

Identify all noun phrases (NPs) 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 (NPs) 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 (NPs) that refer to the same entity

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

Noun Phrase Coreference

Identify all noun phrases (NPs) 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...

Standard Machine Learning Approach

- **Classification**
  [Aone and Bennett (1995), McCarthy and Lehnert (1995), Soon et al., (2001), Ng and Cardie (2002), Strube et al. (2002)]
  - given a description of two noun phrases, NP_i and NP_j, classifies the pair as *coreferent* or *not coreferent*

Standard Machine Learning Approach

- **Classification**
  [Aone and Bennett (1995), McCarthy and Lehnert (1995), Soon et al., (2001), Ng and Cardie (2002), Strube et al. (2002)]
  - given a description of two noun phrases, NP_i and NP_j, classifies the pair as *coreferent* or *not coreferent*

- **Clustering**
  - coordinates pairwise classification decisions
  - single-link clustering algorithm commonly employed to find an antecedent for each NP

Plan for the Talk

- Noun phrase coreference resolution
  - standard machine learning approach
- Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  - why anaphoricity info can help coreference resolution
- Issues in computing and using anaphoricity information in coreference resolution

Motivation

- Currently
  - clustering algorithm attempts to resolve *each* NP

Motivation

- Currently
  - clustering algorithm attempts to resolve *each* NP
- What we really want
  - clustering algorithm attempts to resolve *each* anaphoric NP

Motivation

- Currently
  - clustering algorithm attempts to resolve *each* NP
- What we really want
  - clustering algorithm attempts to resolve *each* anaphoric NP
- Availability of anaphoricity info can potentially increase the precision of a coreference system

Previous Work on Anaphoricity Determination

- Focus on identifying specific types of noun phrases
  - pleonastic pronouns
    - Paice and Husk (1987), Lappin and Leass (1994), Kennedy and Boguraev (1996), Denber (1998)
  - definite descriptions
    - Bean and Riloff (1999), Vieira and Poesio (2000), Poesio et al. (2004)
  - anaphoric and non-anaphoric uses of *it*
    - Evans (2001) / Mitkov et al. (2002)

Anaphoricity Determination [Ng and Cardie, 2002; Uryupina, 2003]

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 (Cont')

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 (Cont')

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 (Cont')

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 (Cont')

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~~...

Plan for the Talk

- Noun phrase coreference resolution
  - standard machine learning approach
- Identification of anaphoric/non-anaphoric noun phrases (Anaphoricity determination)
  - why anaphoricity info can help coreference resolution
- Two issues in computing and using anaphoricity information in coreference resolution

Issue 1

- Representation of anaphoricity information for learning-based coreference systems
  - constraint-based representation
    - clustering algorithm only attempts to resolve anaphoric NPs
    - anaphoricity information serves as hard constraints
  - feature-based representation
    - anaphoricity information represented as a feature

Issue 1

- Representation of anaphoricity information for learning-based coreference systems
  - constraint-based representation
    - clustering algorithm only attempts to resolve anaphoric NPs
    - anaphoricity information serves as hard constraints
  - feature-based representation
    - anaphoricity information represented as a feature

Constraint-based or feature-based representation?

Issue 2

- Optimization of the anaphoricity determination procedure
  - local optimization
    - procedure developed independently of the coreference system
  - global optimization
    - procedure optimized for coreference performance

Issue 2

- Optimization of the anaphoricity determination procedure
  - local optimization
    - procedure developed independently of the coreference system
  - global optimization
    - procedure optimized for coreference performance

Local or global optimization?

Four Approaches to Anaphoricity Determination for Coreference Resolution

	Constraint-Based	Feature-Based
Locally-Optimized		
Globally-Optimized		

Four Approaches to Anaphoricity Determination for Coreference Resolution

	Constraint-Based	**Feature-Based**
Locally-Optimized	Ng and Cardie (2002)	
Globally-Optimized		

Four Approaches to Anaphoricity Determination for Coreference Resolution

	Constraint-Based	Feature-Based
Locally-Optimized	Ng and Cardie (2002)	?
Globally-Optimized	?	?

Goal

- Evaluate all four combinations of
  - **local** vs. **global** optimization and
  - **constraint-based** vs. **feature-based** representation

  of anaphoricity information in terms of their effectiveness in improving a learning-based coreference system

Goal

- Evaluate all four combinations of
  - **local** vs. **global** optimization and
  - **constraint-based** vs. **feature-based** representation

  of anaphoricity information in terms of their effectiveness in improving a learning-based coreference system

The Locally-Optimized Approach to Anaphoricity Determination

- Classification [Ng and Cardie, 2002]
  - 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], ...

The Locally-Optimized Approach to Anaphoricity Determination

- Classification [Ng and Cardie, 2002]
  - 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], ...

- Training data creation
  - texts annotated with coreference information
  - one instance for each noun phrase
    - positive if the noun phrase is anaphoric
    - negative otherwise

Potential Problem with Local Optimization

- During classifier training
  - learn a classifier that maximizes classification accuracy

Potential Problem with Local Optimization

- During classifier training
  - learn a classifier that maximizes classification accuracy
- Classifier may be sub-optimal w.r.t. improving the coreference system

Potential Problem with Local Optimization

- During classifier training
  - learn a classifier that maximizes classification accuracy
- Classifier may be sub-optimal w.r.t. improving the coreference system
  - too conservative in classifying an NP as anaphoric
  - too liberal in classifying an NP as anaphoric

Potential Problem with Local Optimization

- During classifier training
  - learn a classifier that maximizes classification accuracy
- Classifier may be sub-optimal w.r.t. improving the coreference system

  Given a **constraint-based** representation of anaphoricity info
  - too conservative in classifying an NP as anaphoric
  - too liberal in classifying an NP as anaphoric

Potential Problem with Local Optimization

- During classifier training
  - learn a classifier that maximizes classification accuracy
- Classifier may be sub-optimal w.r.t. improving the coreference system

  Given a **constraint-based** representation of anaphoricity info
  - too conservative in classifying an NP as anaphoric
    - clustering algorithm bypasses too many truly anaphoric NPs
  - too liberal in classifying an NP as anaphoric

Potential Problem with Local Optimization

- During classifier training
  - learn a classifier that maximizes classification accuracy
- Classifier may be sub-optimal w.r.t. improving the coreference system

  Given a **constraint-based** representation of anaphoricity info
  - too conservative in classifying an NP as anaphoric
    - clustering algorithm bypasses too many truly anaphoric NPs
  - too liberal in classifying an NP as anaphoric
    - anaphoricity info not effective for improving coreference

Potential Problem with Local Optimization

- During classifier training
  - learn a classifier that maximizes classification accuracy
- Classifier may be sub-optimal w.r.t. improving the coreference system

  Given a **constraint-based** representation of anaphoricity info
  - too conservative in classifying an NP as anaphoric
    - clustering algorithm bypasses too many truly anaphoric NPs
  - too liberal in classifying an NP as anaphoric
    - anaphoricity info not effective for improving coreference

Want a classifier with the right degree of conservativeness

Global Optimization for a Constraint-Based Representation

- Idea
  1. construct anaphoricity classifiers with different degrees of conservativeness
  2. pick the classifier that yields the best coreference performance on held-out data

Global Optimization for a Constraint-Based Representation

- Idea
  1. construct anaphoricity classifiers with different degrees of conservativeness
  2. pick the classifier that yields the best coreference performance on held-out data

How to implement step 1?

Constructing Classifiers with Different Degrees of Conservativeness

Method 1: Varying the cost ratio (*cr*)

$$cr = \frac{\text{Cost of misclassifying a positive instance}}{\text{Cost of misclassifying a negative instance}}$$

Constructing Classifiers with Different Degrees of Conservativeness

Method 1: Varying the cost ratio (*cr*)

$$cr = \frac{\text{Cost of misclassifying a positive instance}}{\text{Cost of misclassifying a negative instance}}$$

$$= \frac{\text{Cost of misclassifying an anaphoric NP}}{\text{Cost of misclassifying a non-anaphoric NP}}$$

Constructing Classifiers with Different Degrees of Conservativeness

Method 1: Varying the cost ratio (*cr*)

$$cr = \frac{\text{Cost of misclassifying a positive instance}}{\text{Cost of misclassifying a negative instance}}$$

$$= \frac{\text{Cost of misclassifying an anaphoric NP}}{\text{Cost of misclassifying a non-anaphoric NP}}$$

- *cr* ↑ ==> more liberal in classifying an NP as anaphoric

Constructing Classifiers with Different Degrees of Conservativeness

Method 1: Varying the cost ratio (*cr*)

$$cr = \frac{\text{Cost of misclassifying a positive instance}}{\text{Cost of misclassifying a negative instance}}$$

$$= \frac{\text{Cost of misclassifying an anaphoric NP}}{\text{Cost of misclassifying a non-anaphoric NP}}$$

- *cr* $\uparrow$ ==> more liberal in classifying an NP as anaphoric

Train classifiers with different values of *cr*

Constructing Classifiers with Different Degrees of Conservativeness

Method 1: Varying the cost ratio (*cr*)

$$cr = \frac{\text{Cost of misclassifying a positive instance}}{\text{Cost of misclassifying a negative instance}}$$

$$= \frac{\text{Cost of misclassifying an anaphoric NP}}{\text{Cost of misclassifying a non-anaphoric NP}}$$

- *cr* $\uparrow$ ==> more liberal in classifying an NP as anaphoric

Train classifiers with different values of *cr*
using RIPPER [Cohen, 1995]

Constructing Classifiers with Different Degrees of Conservativeness

Method 2: Varying the classification threshold

Constructing Classifiers with Different Degrees of Conservativeness

Method 2: Varying the classification threshold

1. Train a probabilistic model of anaphoricity $P_A(c \mid i)$
   - *i* is an instance representing an NP and
   - *c* is one of the two possible anaphoricity values

   using maximum entropy (ME)

Constructing Classifiers with Different Degrees of Conservativeness

<u>Method 2</u>: Varying the classification threshold

1. Train a probabilistic model of anaphoricity $P_A(c \mid i)$
   - i is an instance representing an NP and
   - c is one of the two possible anaphoricity values

   using maximum entropy (ME)

2. Construct an anaphoricity classifier M^t from P_A

   $M^t(i) = non\text{-}anaphoric$ iff $P_A(c = non\text{-}anaphoric \mid i) >= t$

Constructing Classifiers with Different Degrees of Conservativeness

Method 2: Varying the classification threshold

1. Train a probabilistic model of anaphoricity $P_A(c \mid i)$
   - i is an instance representing an NP and
   - c is one of the two possible anaphoricity values

   using maximum entropy (ME)

2. Construct an anaphoricity classifier M^t from P_A

   $M^t(i) = \textit{non-anaphoric}$ iff $P_A(c = \textit{non-anaphoric} \mid i) >= t$

- $t \uparrow$ ==> more liberal in classifying an NP as anaphoric

Constructing Classifiers with Different Degrees of Conservativeness

Method 2: Varying the classification threshold

1. Train a probabilistic model of anaphoricity $P_A(c \mid i)$
   - i is an instance representing an NP and
   - c is one of the two possible anaphoricity values

   using maximum entropy (ME)

2. Construct an anaphoricity classifier M^t from P_A

   $M^t(i) = non\text{-}anaphoric$ iff $P_A(c = non\text{-}anaphoric \mid i) >= t$

- $t \uparrow$ ==> more liberal in classifying an NP as anaphoric

Construct classifiers with different values of t

Global Optimization for a Constraint-Based Representation

- Idea
  1. construct anaphoricity classifiers with different degrees of conservativeness (by varying *cr* or *t*)
  2. pick the classifier that yields the best coreference performance on held-out data

Global Optimization for a Constraint-Based Representation

- Idea
  1. construct anaphoricity classifiers with different degrees of conservativeness (by varying *cr* or *t*)
  2. pick the classifier that yields the best coreference performance on held-out data

- *cr or t* ↑ ==> more liberal in classifying an NP as anaphoric

Global Optimization for a Constraint-Based Representation

- Idea
  1. construct anaphoricity classifiers with different degrees of conservativeness (by varying *cr* or *t*)
  2. pick the classifier that yields the best coreference performance on held-out data

- *cr or t* ↑ ==> more liberal in classifying an NP as anaphoric
  ==> ↑ recall and ↓ precision of coreference system

Global Optimization for a Feature-Based Representation

- Idea
  1. construct anaphoricity classifiers with different degrees of conservativeness (by varying *cr* or *t*)
  2. pick the classifier that yields the best coreference performance on held-out data

- *cr or t* ↑ ==> more liberal in classifying an NP as anaphoric
  ==> ↑ recall and ↓ precision of coreference system

Global Optimization for a Feature-Based Representation

- Idea ✓
  1. construct anaphoricity classifiers with different degrees of conservativeness (by varying *cr* or *t*)
  2. pick the classifier that yields the best coreference performance on held-out data

- *cr or t* ↑ ==> more liberal in classifying an NP as anaphoric
  ==> ↑ recall and ↓ precision of coreference system

Global Optimization for a Feature-Based Representation

- Idea
  1. construct anaphoricity classifiers with different degrees of conservativeness (by varying *cr* or *t*)
  2. pick the classifier that yields the best coreference performance on held-out data

- *cr or t* ↑ ==> more liberal in classifying an NP as anaphoric
  ==> ↑ recall and ↓ precision of coreference system

Global Optimization for a Feature-Based Representation

- Idea ✓
  1. construct anaphoricity classifiers with different degrees of conservativeness (by varying *cr* or *t*)
  2. pick the classifier that yields the best coreference performance on held-out data

- *cr or t* ↑ ==> more liberal in classifying an NP as anaphoric
  ==> ↑ recall and ↓ precision of coreference system

Relationship Between Local Optimization and Global Optimization

- The local approach is a special case of the global one
  - global approach: *cr* and *t* are tuned based on held-out data
  - local approach: default values of *cr* and *t* are used (*cr* is set to 1, *t* is set to 0.5)

What we've done so far …

	Constraint-Based	Feature-Based
Locally-Optimized		
Globally-Optimized		

What we've done so far …

	Constraint-Based	Feature-Based
Locally-Optimized		
Globally-Optimized		

What we've done so far …

	Constraint-Based	Feature-Based
Locally-Optimized		
Globally-Optimized		

What we've done so far …

	Constraint-Based	Feature-Based
Locally-Optimized	✓	✓
Globally-Optimized	✓	✓

Experimental Setup

- Coreference system [Ng and Cardie, ACL 2002]
  - implements the standard machine learning framework
- Features for anaphoricity determination [Ng and Cardie, COLING 2002]
  - 37 features per instance
- Learning algorithms
  - RIPPER and ME

Experimental Setup (Cont')

- The ACE coreference corpus
  - 3 data sets (Broadcast News, Newspaper, Newswire)
  - each data set comprises a training set and a test set
- NPs extracted automatically
- MUC scoring program
  - recall, precision, F-measure

Baseline System (No Anaphoricity): Results

		Broadcast News			Newspaper			Newswire		
	L	R	P	F	R	P	F	R	P	F
Baseline	RIP	57.4	55.3	**56.3**	60.0	63.6	**61.8**	53.2	50.3	**51.7**
	ME	60.9	52.1	**56.2**	65.4	58.6	**61.8**	54.9	46.7	**50.4**

Baseline System (No Anaphoricity): Results

		Broadcast News			Newspaper			Newswire		
	L	R	P	F	R	P	F	R	P	F
Baseline	RIP	57.4	55.3	**56.3**	60.0	63.6	**61.8**	53.2	50.3	**51.7**
	ME	60.9	52.1	**56.2**	65.4	58.6	**61.8**	54.9	46.7	**50.4**

Baseline System (No Anaphoricity): Results

		Broadcast News			Newspaper			Newswire		
	L	R	P	F	R	P	F	R	P	F
Baseline	RIP	57.4	55.3	**56.3**	60.0	63.6	**61.8**	53.2	50.3	**51.7**
	ME	60.9	52.1	**56.2**	65.4	58.6	**61.8**	54.9	46.7	**50.4**

Baseline System (No Anaphoricity): Results

		Broadcast News			Newspaper			Newswire		
	L	R	P	F	R	P	F	R	P	F
Baseline	RIP	57.4	55.3	**56.3**	60.0	63.6	**61.8**	53.2	50.3	**51.7**
	ME	60.9	52.1	**56.2**	65.4	58.6	**61.8**	54.9	46.7	**50.4**

Baseline System (No Anaphoricity): Results

		Broadcast News			Newspaper			Newswire		
	L	R	P	F	R	P	F	R	P	F
Baseline	RIP	57.4	55.3	**56.3**	60.0	63.6	**61.8**	53.2	50.3	**51.7**
	ME	60.9	52.1	**56.2**	65.4	58.6	**61.8**	54.9	46.7	**50.4**

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

- large gains in precision at the expense of recall

Results on the Constraint-Based, Locally-Optimized Approach (CBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBLO	RIP	42.5	77.2	**54.8**	*cr*=1	46.7	79.3	**58.8**	*cr*=1	42.1	64.2	**50.9**	*cr*=1
	RIP	45.4	72.8	**55.9**	*t*=.5	52.2	75.9	**61.9**	*t*=.5	36.9	61.5	**46.1**	*t*=.5
	ME	44.4	76.9	**56.3**	*cr*=1	50.1	75.7	**60.3**	*cr*=1	43.9	63.0	**51.7**	*cr*=1
	ME	47.3	70.8	**56.7**	*t*=.5	57.1	70.6	**63.1**	*t*=.5	38.1	60.0	**46.6**	*t*=.5

- large gains in precision at the expense of recall
- not very effective at improving the baseline

Results on the Feature-Based, Locally-Optimized Approach (FBLO)

	L	Broadcast News R	P	F	C	Newspaper R	P	F	C	Newswire R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
FBLO	RIP	53.5	61.3	**57.2**	*cr*=1	58.7	69.7	**63.7**	*cr*=1	54.2	46.8	**50.2**	*cr*=1
	RIP	58.3	58.3	**58.3**	*t*=.5	63.5	57.0	**60.1**	*t*=.5	63.4	35.3	**45.3**	*t*=.5
	ME	59.6	51.6	**55.3**	*cr*=1	65.6	57.9	**61.5**	*cr*=1	55.1	46.2	**50.3**	*cr*=1
	ME	59.6	51.6	**55.3**	*t*=.5	66.0	57.7	**61.6**	*t*=.5	54.9	46.7	**50.4**	*t*=.5

Results on the Feature-Based, Locally-Optimized Approach (FBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
FBLO	RIP	53.5	61.3	**57.2**	*cr*=1	58.7	69.7	**63.7**	*cr*=1	54.2	46.8	**50.2**	*cr*=1
	RIP	58.3	58.3	**58.3**	*t*=.5	63.5	57.0	**60.1**	*t*=.5	63.4	35.3	**45.3**	*t*=.5
	ME	59.6	51.6	**55.3**	*cr*=1	65.6	57.9	**61.5**	*cr*=1	55.1	46.2	**50.3**	*cr*=1
	ME	59.6	51.6	**55.3**	*t*=.5	66.0	57.7	**61.6**	*t*=.5	54.9	46.7	**50.4**	*t*=.5

- results using RIPPER are mixed

Results on the Feature-Based, Locally-Optimized Approach (FBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	**--**	60.0	63.6	**61.8**	**--**	53.2	50.3	**51.7**	**--**
	ME	60.9	52.1	**56.2**	**--**	65.4	58.6	**61.8**	**--**	54.9	46.7	**50.4**	**--**
FBLO	RIP	53.5	61.3	**57.2**	*cr*=1	58.7	69.7	**63.7**	*cr*=1	54.2	46.8	**50.2**	*cr*=1
	RIP	58.3	58.3	**58.3**	*t*=.5	63.5	57.0	**60.1**	*t*=.5	63.4	35.3	**45.3**	*t*=.5
	ME	59.6	51.6	**55.3**	*cr*=1	65.6	57.9	**61.5**	*cr*=1	55.1	46.2	**50.3**	*cr*=1
	ME	59.6	51.6	**55.3**	*t*=.5	66.0	57.7	**61.6**	*t*=.5	54.9	46.7	**50.4**	*t*=.5

- results using RIPPER are mixed; results using ME are poor

Results on the Feature-Based, Locally-Optimized Approach (FBLO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
FBLO	RIP	53.5	61.3	**57.2**	*cr*=1	58.7	69.7	**63.7**	*cr*=1	54.2	46.8	**50.2**	*cr*=1
	RIP	58.3	58.3	**58.3**	*t*=.5	63.5	57.0	**60.1**	*t*=.5	63.4	35.3	**45.3**	*t*=.5
	ME	59.6	51.6	**55.3**	*cr*=1	65.6	57.9	**61.5**	*cr*=1	55.1	46.2	**50.3**	*cr*=1
	ME	59.6	51.6	**55.3**	*t*=.5	66.0	57.7	**61.6**	*t*=.5	54.9	46.7	**50.4**	*t*=.5

- results using RIPPER are mixed; results using ME are poor
- not very effective at improving the baseline either

Results on the Constraint-Based, Globally-Optimized Approach (CBGO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBGO	RIP	54.5	68.6	**60.8**	*cr*=5	58.4	68.8	**63.2**	*cr*=4	50.5	56.7	**53.4**	*cr*=3
	RIP	54.1	67.1	**59.9**	*t*=.7	56.5	68.1	**61.7**	*t*=.65	50.3	53.8	**52.0**	*t*=.7
	ME	54.8	62.9	**58.5**	*cr*=5	62.4	65.6	**64.0**	*cr*=3	52.2	57.0	**54.5**	*cr*=3
	ME	54.1	60.6	**57.2**	*t*=.7	61.7	64.0	**62.8**	*t*=.7	52.0	52.8	**52.4**	*t*=.7

Results on the Constraint-Based, Globally-Optimized Approach (CBGO)

	L	Broadcast News R	P	F	C	Newspaper R	P	F	C	Newswire R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBGO	RIP	54.5	68.6	**60.8**	*cr*=5	58.4	68.8	**63.2**	*cr*=4	50.5	56.7	**53.4**	*cr*=3
	RIP	54.1	67.1	**59.9**	*t*=.7	56.5	68.1	**61.7**	*t*=.65	50.3	53.8	**52.0**	*t*=.7
	ME	54.8	62.9	**58.5**	*cr*=5	62.4	65.6	**64.0**	*cr*=3	52.2	57.0	**54.5**	*cr*=3
	ME	54.1	60.6	**57.2**	*t*=.7	61.7	64.0	**62.8**	*t*=.7	52.0	52.8	**52.4**	*t*=.7

- 2/3 of training texts for acquiring classifiers; 1/3 for development
- parameter tuning: 1,2, …, 10 and their reciprocals for *cr*
  0.05, 0.1, …, 1.0 for *t*

Results on the Constraint-Based, Globally-Optimized Approach (CBGO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBGO	RIP	54.5	68.6	**60.8**	*cr*=5	58.4	68.8	**63.2**	*cr*=4	50.5	56.7	**53.4**	*cr*=3
	RIP	54.1	67.1	**59.9**	*t*=.7	56.5	68.1	**61.7**	*t*=.65	50.3	53.8	**52.0**	*t*=.7
	ME	54.8	62.9	**58.5**	*cr*=5	62.4	65.6	**64.0**	*cr*=3	52.2	57.0	**54.5**	*cr*=3
	ME	54.1	60.6	**57.2**	*t*=.7	61.7	64.0	**62.8**	*t*=.7	52.0	52.8	**52.4**	*t*=.7

- no significantly worse results; 9 indicate significant improvements

Results on the Constraint-Based, Globally-Optimized Approach (CBGO)

	L	Broadcast News R	P	F	C	Newspaper R	P	F	C	Newswire R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBGO	RIP	54.5	68.6	**60.8**	*cr*=5	58.4	68.8	**63.2**	*cr*=4	50.5	56.7	**53.4**	*cr*=3
	RIP	54.1	67.1	**59.9**	*t*=.7	56.5	68.1	**61.7**	*t*=.65	50.3	53.8	**52.0**	*t*=.7
	ME	54.8	62.9	**58.5**	*cr*=5	62.4	65.6	**64.0**	*cr*=3	52.2	57.0	**54.5**	*cr*=3
	ME	54.1	60.6	**57.2**	*t*=.7	61.7	64.0	**62.8**	*t*=.7	52.0	52.8	**52.4**	*t*=.7

- no significantly worse results; 9 indicate significant improvements
- yields our best results on all three data sets

Results on the Constraint-Based, Globally-Optimized Approach (CBGO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
CBGO	RIP	54.5	68.6	**60.8**	*cr*=5	58.4	68.8	**63.2**	*cr*=4	50.5	56.7	**53.4**	*cr*=3
	RIP	54.1	67.1	**59.9**	*t*=.7	56.5	68.1	**61.7**	*t*=.65	50.3	53.8	**52.0**	*t*=.7
	ME	54.8	62.9	**58.5**	*cr*=5	62.4	65.6	**64.0**	*cr*=3	52.2	57.0	**54.5**	*cr*=3
	ME	54.1	60.6	**57.2**	*t*=.7	61.7	64.0	**62.8**	*t*=.7	52.0	52.8	**52.4**	*t*=.7

- no significantly worse results; 9 indicate significant improvements
- yields our best results on all three data sets
- locally-optimized classifiers are too conservative in classifying an NP as anaphoric

Results on the Feature-Based, Globally-Optimized Approach (FBGO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
FBGO	RIP	60.8	56.1	**58.4**	*cr*=8	62.2	61.3	**61.7**	*cr*=6	54.6	49.4	**51.9**	*cr*=8
	RIP	59.7	57.0	**58.3**	*t*=.6	63.6	59.1	**61.3**	*t*=.8	56.7	48.4	**52.3**	*t*=.7
	ME	59.9	51.0	**55.1**	*cr*=9	66.5	57.1	**61.4**	*cr*=1	56.3	46.9	**51.2**	*cr*=10
	ME	59.6	51.6	**55.3**	*t*=.95	65.9	57.5	**61.4**	*t*=.95	56.5	46.7	**51.1**	*t*=.5

Results on the Feature-Based, Globally-Optimized Approach (FBGO)

		Broadcast News				Newspaper				Newswire			
	L	R	P	F	C	R	P	F	C	R	P	F	C
Baseline	RIP	57.4	55.3	**56.3**	--	60.0	63.6	**61.8**	--	53.2	50.3	**51.7**	--
	ME	60.9	52.1	**56.2**	--	65.4	58.6	**61.8**	--	54.9	46.7	**50.4**	--
FBGO	RIP	60.8	56.1	**58.4**	*cr*=8	62.2	61.3	**61.7**	*cr*=6	54.6	49.4	**51.9**	*cr*=8
	RIP	59.7	57.0	**58.3**	*t*=.6	63.6	59.1	**61.3**	*t*=.8	56.7	48.4	**52.3**	*t*=.7
	ME	59.9	51.0	**55.1**	*cr*=9	66.5	57.1	**61.4**	*cr*=1	56.3	46.9	**51.2**	*cr*=10
	ME	59.6	51.6	**55.3**	*t*=.95	65.9	57.5	**61.4**	*t*=.95	56.5	46.7	**51.1**	*t*=.5

- not very effective at improving the baseline

Summary

- Evaluated four combinations of
  - local vs. global optimization and
  - constraint-based vs. feature-based representation

  of anaphoricity information in terms of their effectiveness in improving a learning-based coreference system
- Showed that the constraint-based, globally-optimized approach is the most effective

Future Work

- Investigate better features for anaphoricity determination [Poesio et al, 2004]
  - e.g., definite probability of an NP [Bean and Riloff, 1999; Uryupina, 2003]

Summary

- Evaluated four combinations of
  - local vs. global optimization and
  - constraint-based vs. feature-based representation

  of anaphoricity information in terms of their effectiveness in improving a learning-based coreference system

- Showed that the constraint-based, globally-optimized approach is the most effective

- Approach can be used in conjunction with
  - knowledge-based coreference systems
  - anaphora/coreference resolution systems for spoken dialogues [Strube and Műller, 2003]