Shallow Semantics for Coreference Resolution

Vincent Ng
Human Language Technology Research Institute
University of Texas at Dallas

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 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...

Standard Machine Learning Approach

- **Step 1: Classification**
  - given a description of two noun phrases, *NP_i* and *NP_j*, classifies the pair as *coreferent* or *not coreferent*

coref ? *coref ?*

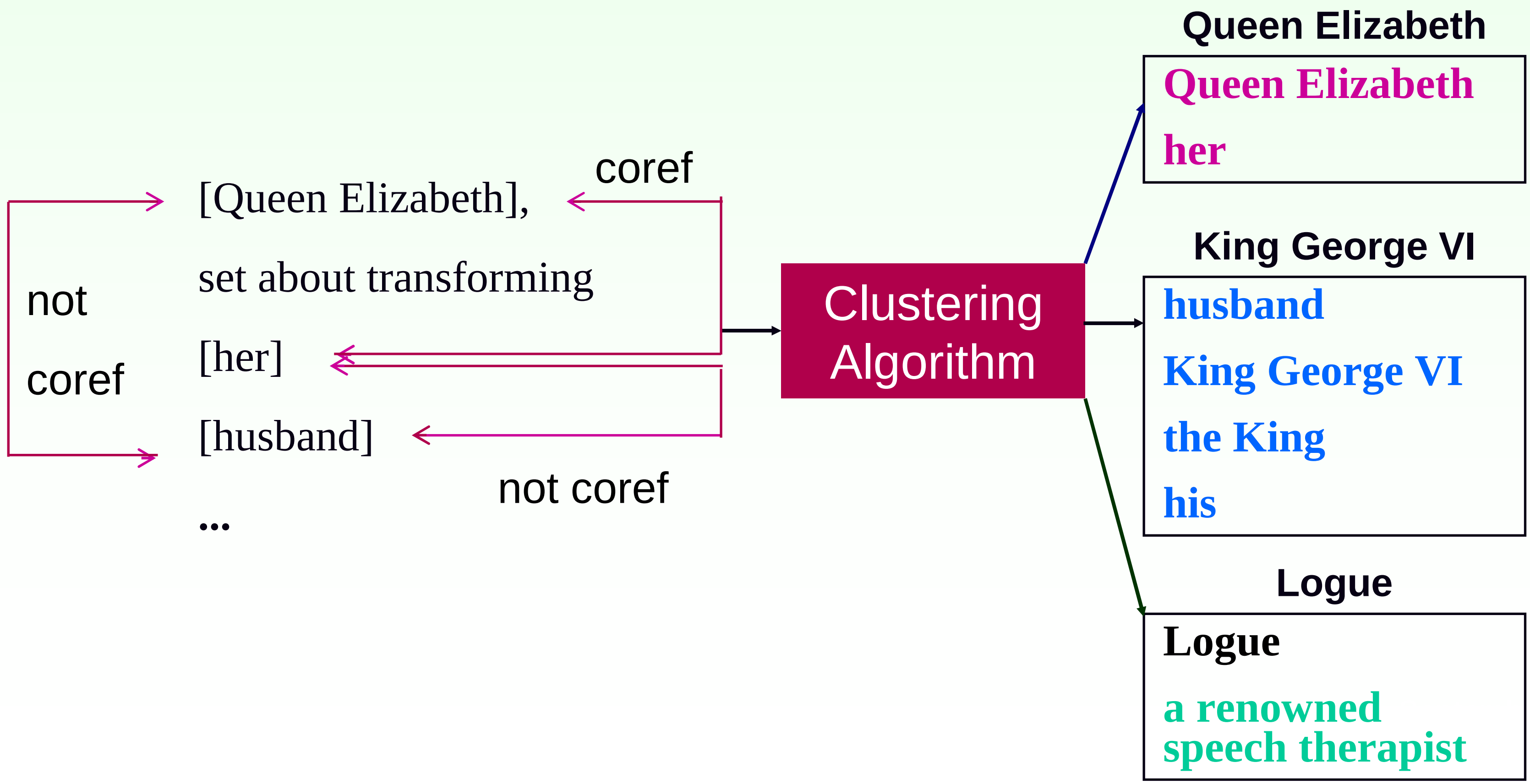
[Queen Elizabeth] set about transforming [her] [husband], ...

not coref ?

Aone & Bennett [1995]; Connolly et al. [1994]; McCarthy & Lehnert [1995]; Soon, Ng & Lim [2001]; Ng & Cardie [2002]

Standard Machine Learning Approach

- **Step 2: Clustering**
  - coordinates pairwise classification decisions



Machine Learning Issues

- Training instance creation

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm: C4.5

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm: C4.5
- Clustering algorithm

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm: C4.5
- Clustering algorithm: single-link clustering

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm: C4.5
- Clustering algorithm: single-link clustering
  - Select as the referent of each NP, NP_x, the closest preceding NP that is classified as coreferent with NP_x.

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm: C4.5
- Clustering algorithm: single-link clustering
  - Select as the referent of each NP, NP_x, the closest preceding NP that is classified as coreferent with NP_x.
  - If no such NP exists, no referent is selected for $NP_{x.}$

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm: C4.5
- Clustering algorithm: single-link clustering
  - Select as the referent of each NP, NP_x, the closest preceding NP that is classified as coreferent with NP_x.
  - If no such NP exists, no referent is selected for $NP_{x.}$
- Instance representation
  - how to describe the two NPs?

Machine Learning Issues

- Training instance creation
  - Pair each NP with each of its preceding NPs
  - Label an instance as positive iff the two NPs are coreferent
- Learning algorithm: C4.5
- Clustering algorithm: single-link clustering
  - Select as the referent of each NP, NP_x, the closest preceding NP that is classified as coreferent with NP_x.
  - If no such NP exists, no referent is selected for NP_x.
- Instance representation
  - how to describe the two NPs?

This Talk

- Development of linguistic features for coreference resolution

This Talk

- Development of linguistic features for coreference resolution

- Research in coreference resolution has largely adopted a knowledge-lean approach (Mitkov et al., 2001)

This Talk

- Development of linguistic features for coreference resolution

- Research in coreference resolution has largely adopted a knowledge-lean approach (Mitkov et al., 2001)
  - Resolvers operate by relying on morpho-syntactic cues
    - String matching, gender/number agreement, binding constraints

This Talk

- Development of linguistic features for coreference resolution

- Research in coreference resolution has largely adopted a knowledge-lean approach (Mitkov et al., 2001)
  - Resolvers operate by relying on morpho-syntactic cues
    - String matching, gender/number agreement, binding constraints
  - However, there are coreference relations that cannot be identified by using string-matching facilities and syntactic cues

This Talk

- Development of linguistic features for coreference resolution

- Research in coreference resolution has largely adopted a knowledge-lean approach (Mitkov et al., 2001)
  - Resolvers operate by relying on morpho-syntactic cues
    - String matching, gender/number agreement, binding constraints
  - However, there are coreference relations that cannot be identified by using string-matching facilities and syntactic cues
    - Coreference relations between two lexically dissimilar common nouns (e.g., *talks* and *negotiations*)

This Talk

- Development of linguistic features for coreference resolution

- Research in coreference resolution has largely adopted a knowledge-lean approach (Mitkov et al., 2001)
  - Resolvers operate by relying on morpho-syntactic cues
    - String matching, gender/number agreement, binding constraints
  - However, there are coreference relations that cannot be identified by using string-matching facilities and syntactic cues
    - Coreference relations between two lexically dissimilar common nouns (e.g., *talks* and *negotiations*)
    - Coreference relations between a proper NP and a common NP (e.g., *George W. Bush* and *the president*)

Goal

- Investigate features that encode semantic and other non-morpho-syntactic knowledge for improving the performance of a learning-based coreference system

Goal

- Investigate features that encode semantic and other non-morpho-syntactic knowledge for improving the performance of a learning-based coreference system

- Focus on inducing linguistic features
  - one feature exploits the fact that we are doing ACE coreference

Plan for the Talk

- Six linguistic features for coreference resolution
- The baseline feature set
- Evaluation

Plan for the Talk

- Six linguistic features for coreference resolution
- The baseline feature set
- Evaluation

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*
  - No for *Hyderabad* and *the president*

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*
  - No for *Hyderabad* and *the president*
- Need to compute the semantic classes of the two NPs

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*
  - No for *Hyderabad* and *the president*
- Need to compute the semantic classes of the two NPs
  - For a named entity, use a named entity recognizer

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*
  - No for *Hyderabad* and *the president*
- Need to compute the semantic classes of the two NPs
  - For a named entity, use a named entity recognizer
  - For a common noun, use WordNet (choose the first sense)

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*
  - No for *Hyderabad* and *the president*
- Need to compute the semantic classes of the two NPs
  - For a named entity, use a named entity recognizer
  - For a common noun, use WordNet (choose the first sense)
    - An overly simplistic heuristic

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*
  - No for *Hyderabad* and *the president*
- Need to compute the semantic classes of the two NPs
  - For a named entity, use a named entity recognizer
  - For a common noun, use WordNet (choose the first sense)
    - An overly simplistic heuristic
    - Feature not used by Soon et al.'s (2001) resolver for the MUC-6 dataset

1. The Semantic Class Agreement Feature

- Determines whether the semantic classes of two NPs agree
  - Yes for *Mahatma Ghandi* and *the president*
  - No for *Hyderabad* and *the president*
- Need to compute the semantic classes of the two NPs
  - For a named entity, use a named entity recognizer
  - For a common noun, use WordNet (choose the first sense)
    - An overly simplistic heuristic
    - Feature not used by Soon et al.'s (2001) resolver for the MUC-6 dataset
- Goal: improve computation of the semantic class of an NP

The Semantic Class Induction Algorithm

- Given a large, unannotated corpus
  - Extract appositive relations
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names
  - Infer the semantic class of a common nouns from the associated proper name

The Semantic Class Induction Algorithm

- Given a large, unannotated corpus ⟵ **BLLIP+Reuters**
  - Extract appositive relations
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names
  - Infer the semantic class of a common nouns from the associated proper name

The Semantic Class Induction Algorithm

- Given a large, unannotated corpus ⟵ **BLLIP+Reuters**
  - Extract appositive relations ⟵ **MINIPAR**
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names
  - Infer the semantic class of a common nouns from the associated proper name

The Semantic Class Induction Algorithm

- Given a large, unannotated corpus **BLLIP+Reuters**
  - Extract appositive relations **MINIPAR**
    - <Eastern Airlines, carrier>, <George Bush, president>, …
  - Use a named entity (NE) recognizer to find the semantic classes of the proper names **Identifinder (MUC-style NER)**
  - Infer the semantic class of a common nouns from the associated proper name

Potential Problem

- Identifinder is not perfect
  - Mislabels proper names
- MINIPAR is not perfect
  - Extracts NP pairs that are not in apposition

a common noun could be labeled with more than one semantic class

Potential Problem

- Identifinder is not perfect
  - Mislabels proper names
- MINIPAR is not perfect
  - Extracts NP pairs that are not in apposition
- Need a more robust method of inferring the semantic class of a common noun

a common noun could be labeled with more than one semantic class

Potential Problem

- Identifinder is not perfect
  - Mislabels proper names
- MINIPAR is not perfect
  - Extracts NP pairs that are not in apposition

a common noun could be labeled with more than one semantic class

- Need a more robust method of inferring the semantic class of a common noun
  1. Compute the probability that the common noun co-occurs with each of the named entity types
  2. If the most likely NE type has a probability above 0.7, label the common noun with the most likely NE type

Other Problems

- Common nouns that do not belong to one of the seven MUC NE types will remain unlabeled

Other Problems

- Common nouns that do not belong to one of the seven MUC NE types will remain unlabeled

- Common nouns that do not co-occur with any NE type with a probability above 0.7 will remain unlabeled

Other Problems

- Common nouns that do not belong to one of the seven MUC NE types will remain unlabeled
- Common nouns that do not co-occur with any NE type with a probability above 0.7 will remain unlabeled
- Solution: fall back on the first-sense heuristic

2. The ACE-Specific Semantic Agreement Feature

- Motivation
  - The SEM_CLASS feature was developed for use in a general-purpose coreference system
  - We may be able to improve performance on the ACE data if we develop an ACE-specific semantic agreement feature

2. The ACE-Specific Semantic Agreement Feature

- Motivation
  - The SEM_CLASS feature was developed for use in a general-purpose coreference system
  - We may be able to improve performance on the ACE data if we develop an ACE-specific semantic agreement feature
- ACE coreference
  - Resolve references to NPs that belong to one of the five ACE semantic classes (ASCs)
    - PERSON, ORGANIZATION, FACILITY, GSP, LOCATION

Definition of ACE Semantic Classes

- PERSON (human)
  - Mahatma Ghandi, the postman, …

Definition of ACE Semantic Classes

- PERSON (human)
  - Mahatma Ghandi, the postman, …
- ORGANIZATION (corporation, agency, government)
  - Indian Institute of Technology, the company, …

Definition of ACE Semantic Classes

- PERSON (human)
  - Mahatma Ghandi, the postman, …
- ORGANIZATION (corporation, agency, government)
  - Indian Institute of Technology, the company, …
- FACILITY (man-made structure)
  - Hyderabad International Convention Center, the building, …

Definition of ACE Semantic Classes

- PERSON (human)
  - Mahatma Ghandi, the postman, …
- ORGANIZATION (corporation, agency, government)
  - Indian Institute of Technology, the company, …
- FACILITY (man-made structure)
  - Hyderabad International Convention Center, the building, …
- GSP (geo-political region)
  - India, Hyderabad, the city, the province, …

Definition of ACE Semantic Classes

- PERSON (human)
  - Mahatma Ghandi, the postman, …
- ORGANIZATION (corporation, agency, government)
  - Indian Institute of Technology, the company, …
- FACILITY (man-made structure)
  - Hyderabad International Convention Center, the building, …
- GSP (geo-political region)
  - India, Hyderabad, the city, the province, …
- LOCATION (geographical area, landmass, body of water)
  - The Bay of Bengal, the Himalayas, the mountain, …

Definition of ACE Semantic Classes

- PERSON (human)
  - Mahatma Ghandi, the postman, …
- ORGANIZATION (corporation, agency, government)
  - Indian Institute of Technology, the company, …
- FACILITY (man-made structure)
  - Hyderabad International Convention Center, the building, …
- GSP (geo-political region)
  - India, Hyderabad, the city, the province, …
- LOCATION (geographical area, landmass, body of water)
  - The Bay of Bengal, the Himalayas, the mountain, …

- Goal: develop a feature that considers two NPs compatible if and only if the two NPs have a common ASC

Determining the ASC of an NP

- Based in part on the semantic class of the NP as computed by the SEM_CLASS feature

Determining the ASC of an NP

- Based in part on the semantic class of the NP as computed by the SEM_CLASS feature

- Rough correspondence between SEM_CLASS and ASC

SEM_CLASS		**ASC**
PERSON	→	PERSON

Determining the ASC of an NP

- Based in part on the semantic class of the NP as computed by the SEM_CLASS feature
- Rough correspondence between SEM_CLASS and ASC

SEM_CLASS	**ASC**
PERSON	PERSON
ORGANIZATION	ORGANIZATION
	FACILITY

Determining the ASC of an NP

- Based in part on the semantic class of the NP as computed by the SEM_CLASS feature

- Rough correspondence between SEM_CLASS and ASC

SEM_CLASS	**ASC**
PERSON	PERSON
ORGANIZATION	ORGANIZATION
	FACILITY
LOCATION	GSP
	LOCATION

The ASC Determination Algorithm

SEM_CLASS	ASC
PERSON	PERSON
ORG	ORG, FACILITY
LOCATION	GSP, LOCATION

The ASC Determination Algorithm

- If its SEM_CLASS is not PERSON, ORGANIZATION, or LOCATION, its ASC will be OTHERS

SEM_CLASS	ASC
PERSON	PERSON
ORG	ORG, FACILITY
LOCATION	GSP, LOCATION

The ASC Determination Algorithm

- If its SEM_CLASS is not PERSON, ORGANIZATION, or LOCATION, its ASC will be OTHERS

- If its SEM_CLASS is PERSON, its ASC will be PERSON

SEM_CLASS	ASC
PERSON	PERSON
ORG	ORG, FACILITY
LOCATION	GSP, LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION

SEM_CLASS	**ASC**
PERSON	PERSON
ORG	ORG, FACILITY
LOCATION	GSP, LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION
  - Check whether its head noun is a hypernym of a GSP-related word or a LOCATION-related word in WordNet

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION
  - Check whether its head noun is a hypernym of a GSP-related word or a LOCATION-related word in WordNet
    - GSP-related words: country, province, government, town, city, administration, society, island
    - LOCATION-related words: dry land, region, landmass, body of water, geographical area, geological formation

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION
  - Check whether its head noun is a hypernym of a GSP-related word or a LOCATION-related word in WordNet
    - GSP-related words: country, province, government, town, city, administration, society, island
    - LOCATION-related words: dry land, region, landmass, body of water, geographical area, geological formation
  - If head noun is a LOCATION word, its ASC is LOCATION
  - Else if head noun is a GSP word, its ASC is GSP
  - Otherwise, its ASC is both GSP and LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION
  - Check whether its head noun is a hypernym of a GSP-related word or a LOCATION-related word in WordNet
    - GSP-related words: country, province, government, town, city, administration, society, island
    - LOCATION-related words: dry land, region, landmass, body of water, geographical area, geological formation

Bay of Bengal

  - If head noun is a LOCATION word, its ASC is LOCATION
  - Else if head noun is a GSP word, its ASC is GSP
  - Otherwise, its ASC is both GSP and LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION
  - Check whether its head noun is a hypernym of a GSP-related word or a LOCATION-related word in WordNet
    - GSP-related words: country, province, government, town, city, administration, society, island
    - LOCATION-related words: dry land, region, landmass, body of water, geographical area, geological formation

Bay of Bengal

  - If head noun is a LOCATION word, its ASC is LOCATION
  - Else if head noun is a GSP word, its ASC is GSP
  - Otherwise, its ASC is both GSP and LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION
  - Check whether its head noun is a hypernym of a GSP-related word or a LOCATION-related word in WordNet
    - GSP-related words: country, province, government, town, city, administration, society, island
    - LOCATION-related words: dry land, region, landmass, body of water, geographical area, geological formation

Bay of Bengal

  - If head noun is a LOCATION word, its ASC is LOCATION
  - Else if head noun is a GSP word, its ASC is GSP
  - Otherwise, its ASC is both GSP and LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is LOCATION, need to determine whether its ASC is GSP or LOCATION
  - Check whether its head noun is a hypernym of a GSP-related word or a LOCATION-related word in WordNet
    - GSP-related words: country, province, government, town, city, administration, society, island
    - LOCATION-related words: dry land, region, landmass, body of water, geographical area, geological formation
  - If head noun is a LOCATION word, its ASC is LOCATION
  - Else if head noun is a GSP word, its ASC is GSP
  - Otherwise, its ASC is both GSP and LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is ORGANIZATION, need to determine whether its ASC is FACILITY or ORGANIZATION

SEM_CLASS	ASC
PERSON	PERSON
ORG	ORG, FACILITY
LOCATION	GSP, LOCATION

The ASC Determination Algorithm (Cont')

- If its SEM_CLASS is ORGANIZATION, need to determine whether its ASC is FACILITY or ORGANIZATION
  - Check whether its head noun is a hypernym of an ORGANIZATION-related word or a FACILITY-related word
    - ORGANIZATION-related words: social group
    - FACILITY-related words: establishment, construction, building, facility, workplace

SEM_CLASS	**ASC**
PERSON	PERSON
ORG	ORG, FACILITY
LOCATION	GSP, LOCATION

3. The Anaphoricity Feature

- Anaphoricity determination is the problem of determining whether an NP has an antecedent or not

3. The Anaphoricity Feature

- Anaphoricity determination is the problem of determining whether an NP has an antecedent or not
  - Knowledge of anaphoricity could improve system precision

3. The Anaphoricity Feature

- Anaphoricity determination is the problem of determining whether an NP has an antecedent or not
  - Knowledge of anaphoricity could improve system precision
- Previous approaches
  - Heuristic-based : Lappin and Leass (1994), Kennedy and Boguraev (1996), Vieira and Poesio (2000)
  - Unsupervised: Bean and Riloff (1999)
  - Supervised: Evans (2001), Ng and Cardie (2002)

3. The Anaphoricity Feature

- Anaphoricity determination is the problem of determining whether an NP has an antecedent or not
  - Knowledge of anaphoricity could improve system precision
- Previous approaches
  - Heuristic-based : Lappin and Leass (1994), Kennedy and Boguraev (1996), Vieira and Poesio (2000)
  - Unsupervised: Bean and Riloff (1999)
  - Supervised: Evans (2001), Ng and Cardie (2002)
- Goal: examine whether shallow anaphoricity information could benefit a learning-based coreference resolution

Computing the Anaphoricity Feature

- Given a corpus labeled with coreference information
  - Compute the anaphoricity of an NP as the probability that it has an antecedent in the corpus
    - If the NP never appears in the corpus, set its anaphoricity value to -1

Computing the Anaphoricity Feature

- Given a corpus labeled with coreference information
  - Compute the anaphoricity of an NP as the probability that it has an antecedent in the corpus
    - If the NP never appears in the corpus, set its anaphoricity value to -1

- Data sparseness is a problem, but the feature still captures some useful information
  - *it* is only moderately anaphoric
  - *the contrary* (from *on the contrary*) is never anaphoric

4. The Coreferentiality Feature

- Adapt the method for generating the anaphoricity feature to create a coreferentiality feature

- Feature encodes the probability that two NPs are coreferent
  - Estimate the probabilities from a coreference corpus
    - If one or both of the given NPs do not appear in the corpus, set the coreferentiality value to -1

The Remaining Features

5. The Semantic Similarity Feature
   - Determines the semantic similarity of two common NPs

The Remaining Features

5. The Semantic Similarity Feature
   - Determines the semantic similarity of two common NPs
     - Two semantically similar NPs are more likely to be coreferent than two semantically dissimilar NPs

The Remaining Features

5. The Semantic Similarity Feature
   - Determines the semantic similarity of two common NPs
     - Two semantically similar NPs are more likely to be coreferent than two semantically dissimilar NPs

6. The Pattern-Based Feature
   - Computed using information provided by an algorithm that learns patterns for extracting coreferent NP pairs

The Remaining Features

5. The Semantic Similarity Feature
   - Determines the semantic similarity of two common NPs
     - Two semantically similar NPs are more likely to be coreferent than two semantically dissimilar NPs

6. The Pattern-Based Feature
   - Computed using information provided by an algorithm that learns patterns for extracting coreferent NP pairs

- Employing this pattern-based feature does not yield significant improvement in coreference performance

Plan for the Talk

- Six linguistic features for coreference resolution
- The baseline feature set
- Evaluation

The Baseline Feature Set (34 Features)

- String-matching features
  - Exact string match, substring match, head noun match
- Grammatical features
  - Agreement w.r.t. gender, number, animacy, grammatical role
- Positional feature
  - Distance between the two NPs in sentences
- Semantic features
  - Alias, semantic class agreement

The Baseline Feature Set (34 Features)

- String-matching features
  - Exact string match, substring match, head noun match
- Grammatical features
  - Agreement w.r.t. gender, number, animacy, grammatical role
- Positional feature
  - Distance between the two NPs in sentences
- Semantic features
  - Alias, semantic class agreement

For a proper name, use a named entity finder
For a common noun, use WordNet + the first-sense heuristic

Plan for the Talk

- Six linguistic features for coreference resolution
- The baseline feature set
- Evaluation
  - How effective are the proposed features in improving the baseline coreference system?

Experimental Setup

- The 2003 ACE coreference corpus
  - comprises a training set and a test set
- Two coreference scoring programs
  - MUC scoring program (Vilain et al., 1995)
  - CEAF scoring program (Luo, 2005)
  - recall, precision, F-measure
- NPs extracted automatically

The Baseline Coreference System

- Feature set: the baseline feature set (34 features)
- Learning algorithm: C4.5
- Clustering: single-link clustering

Results (Baseline System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**

Results (Baseline System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**

Results (Baseline System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**

Results (Baseline System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**

Results (Baseline System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**

Results (Baseline System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**

How Strong are the Baseline Results?

- Replace the 34 baseline features with the 12 features employed by Soon et al.'s (2001) system
  - The first learning-based resolver that achieves performance comparable to the best MUC coreference systems

Results (Duplicated Soon et al. System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**

Results (Duplicated Soon et al. System)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**

Using the Expanded Feature Set

- Augment the baseline feature set with our six linguistic features
  - SEM_CLASS
  - ACE_SEMCLASS
  - SEM_SIM
  - PATTERN_BASED
  - ANAPHORICITY
  - COREFERENTIALITY

Using the Expanded Feature Set

- Augment the baseline feature set with our six linguistic features
  - SEM_CLASS
  - ACE_SEMCLASS
  - SEM_SIM
  - PATTERN_BASED
  - ANAPHORICITY
  - COREFERENTIALITY
- Remove the heuristic-based semantic class agreement feature from the feature set

Using the Expanded Feature Set

- Augment the baseline feature set with our six linguistic features
  - SEM_CLASS
  - ACE_SEMCLASS
  - SEM_SIM
  - PATTERN_BASED
  - ANAPHORICITY
  - COREFERENTIALITY

Requires an annotated corpus

- Remove the heuristic-based semantic class agreement feature from the feature set

Where does this annotated corpus come from?

- Partition the available training texts into two sets of roughly the same size: training subset and development subset

Where does this annotated corpus come from?

- Partition the available training texts into two sets of roughly the same size: training subset and development subset

↓

Training coreference classifier

Where does this annotated corpus come from?

- Partition the available training texts into two sets of roughly the same size: training subset and development subset

Training subset → Training coreference classifier

Development subset → Computing ANAPHORICITY and COREFERENTIALITY

Results (Expanded Feature Set)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**
Using the expanded feature set	54.7	77.8	**64.2**	56.7	69.0	**62.3**

Results (Expanded Feature Set)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**
Using the expanded feature set	54.7	77.8	**64.2**	56.7	69.0	**62.3**

Results (Expanded Feature Set)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**
Using the expanded feature set	54.7	77.8	**64.2**	56.7	69.0	**62.3**

- Performance difference is statistically significant compared to baseline: p=0.004 (MUC) and p=0.0016 (CEAF)

Results (Expanded Feature Set)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**
Using the expanded feature set	54.7	77.8	**64.2**	56.7	69.0	**62.3**
without SEM_CLASS	55.1	77.5	**64.4**	56.1	67.9	**61.4**
without ACE_SEM_CLASS	53.4	77.1	**63.1**	54.6	67.2	**60.2**
without SEM_SIM	54.7	77.6	**64.2**	56.4	68.1	**61.7**
without PATTERN_BASED	55.0	77.8	**64.5**	56.2	68.2	**61.6**
without ANAPHORICITY	53.7	77.8	**63.5**	55.0	67.9	**60.8**
without COREFERENTIALITY	53.7	78.3	**63.3**	55.0	68.5	**61.0**

Results (Expanded Feature Set)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**
Using the expanded feature set	54.7	77.8	**64.2**	56.7	69.0	**62.3**
without SEM_CLASS	55.1	77.5	**64.4**	56.1	67.9	**61.4**
without ACE_SEM_CLASS	53.4	77.1	**63.1**	54.6	67.2	**60.2**
without SEM_SIM	54.7	77.6	**64.2**	56.4	68.1	**61.7**
without PATTERN_BASED	55.0	77.8	**64.5**	56.2	68.2	**61.6**
without ANAPHORICITY	53.7	77.8	**63.5**	55.0	67.9	**60.8**
without COREFERENTIALITY	53.7	78.3	**63.3**	55.0	68.5	**61.0**

Results (Expanded Feature Set)

	MUC Scorer			CEAF Scorer		
	R	P	F	R	P	F
Using the Baseline features only	53.7	73.4	**62.0**	55.4	65.4	**60.0**
Using Soon et al.'s features only	46.2	73.2	**56.6**	49.8	64.9	**56.3**
Using the expanded feature set	54.7	77.8	**64.2**	56.7	69.0	**62.3**
without SEM_CLASS	55.1	77.5	**64.4**	56.1	67.9	**61.4**
without ACE_SEM_CLASS	53.4	77.1	**63.1**	54.6	67.2	**60.2**
without SEM_SIM	54.7	77.6	**64.2**	56.4	68.1	**61.7**
without PATTERN_BASED	55.0	77.8	**64.5**	56.2	68.2	**61.6**
without ANAPHORICITY	53.7	77.8	**63.5**	55.0	67.9	**60.8**
without COREFERENTIALITY	53.7	78.3	**63.3**	55.0	68.5	**61.0**

Summary

- Investigated the utility of six semantic and non-morpho-syntactic features for coreference resolution
- Showed improved performance on the ACE corpus
- Performance gains are limited in part by the difficulty in accurately computing these features