Reference

Ling575
Discourse and Dialogue
April 6, 2011

Roadmap

- Cohesion and Coreference
- Terminology and Referring Expressions
- Guiding coreference
  - Syntactic & Semantic Constraints & Preferences
- Heuristic approaches
- Machine Learning approaches
- Discussion

Holding Discourse Together

- Cohesion:
  - Necessary to make discourse a semantic unit
  - All utterances linked to some preceding utterance
  - Expresses continuity

  - Key: Enables hearers to interpret missing elements, through textual and environmental context links

Cohesive Ties

(Halliday & Hasan, 1972)

- “Reference”: e.g. “he”,”she”,”it”,”that”
  - Relate utterances by referring to same entities
- “Substitution”/”Ellipsis”:e.g. Jack fell. Jill did too.
  - Relate utterances by repeated partial structure w/contrast
- “Lexical Cohesion”: e.g. fell, fall, fall…,trip..
  - Relate utterances by repeated/related words
- “Conjunction”: e.g. and, or, then
  - Relate continuous text by logical, semantic, interpersonal relations. Interpretation of 2nd utterance depands on first

Entity-based Coherence

- *John went to his favorite music store to buy a piano.*
- *He had frequented the store for many years.*
- *He was excited that he could finally buy a piano.*

- VS
  - *John went to his favorite music store to buy a piano.*
  - *It was a store John had frequented for many years.*
  - *He was excited that he could finally buy a piano.*
  - *It was closing just as John arrived.*
- Which is better? Why?

Entity-based Coherence

  - *John went to his favorite music store to buy a piano.*
  - *He had frequented the store for many years.*
  - *He was excited that he could finally buy a piano.*
- VS
  - *John went to his favorite music store to buy a piano.*
  - *It was a store John had frequented for many years.*
  - *He was excited that he could finally buy a piano.*
  - *It was closing just as John arrived.*
- Which is better? Why?
  - ‘about’ one entity vs two, focuses on it for coherence

Reference

- 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...

Referring expression: (refexp)

Linguistic form that picks out entity in some model

Reference

- 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>...

Referring expression: (refexp)

Linguistic form that picks out entity in some model

That entity is the "referent"

Reference

- 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...

Referring expression: (refexp)

- Linguistic form that picks out entity in some model
- That entity is the "referent"
  - When introduces entity, "evokes" it

Reference

- 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...

Referring expression: (refexp)

- Linguistic form that picks out entity in some model
- That entity is the "referent"
  - When introduces entity, "evokes" it
  - Set up later reference, "antecedent"
- 2 refexps with same referent "co-refer"

Reference (terminology)

- 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...

- Anaphor:
  - Abbreviated linguistic form interpreted in context

Reference (terminology)

- 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...

- Anaphor:
  - Abbreviated linguistic form interpreted in context
    - Her, his, the King

Reference (terminology)

- 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...

- Anaphor:
  - Abbreviated linguistic form interpreted in context
    - Her, his, the King
  - Refers to previously introduced item (“accesses”)
    - Referring expression is then anaphoric

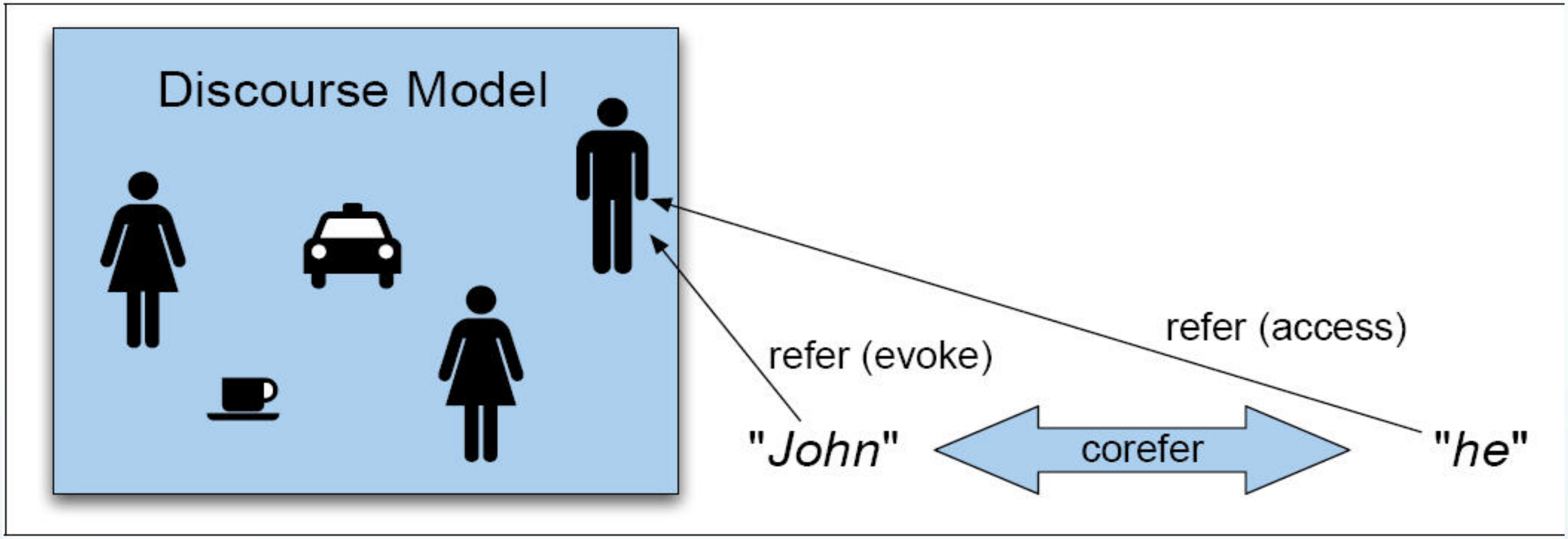
Referring Expressions

- Many alternatives:
  - Queen Elizabeth, she, her, the Queen, etc
  - Possible correct forms depend on discourse context
    - E.g. she, her presume prior mention, or presence in world

Referring Expressions

- Many alternatives:
  - Queen Elizabeth, she, her, the Queen, etc
  - Possible correct forms depend on discourse context
    - E.g. she, her presume prior mention, or presence in world
- Interpretation (and generation) requires:
  - Discourse Model with representations of:
    - Entities referred to in the discourse
    - Relationships of these entities
  - Need way to construct, update model
  - Need way to map refexp to hearer's beliefs

Reference and Model



Reference Resolution

- 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...

Coreference resolution:

Find all expressions referring to same entity, 'corefer'

Colors indicate coreferent sets

Pronominal anaphora resolution:

Find antecedent for given pronoun

Referring Expressions

- Indefinite noun phrases (NPs): e.g. “a cat”
  - Introduces new item to discourse context

Referring Expressions

- Indefinite noun phrases (NPs): e.g. "a cat"
  - Introduces new item to discourse context
- Definite NPs: e.g. "the cat"
  - Refers to item identifiable by hearer in context
    - By verbal, pointing, or environment availability; implicit

Referring Expressions

- Indefinite noun phrases (NPs): e.g. “a cat”
  - Introduces new item to discourse context
- Definite NPs: e.g. “the cat”
  - Refers to item identifiable by hearer in context
    - By verbal, pointing, or environment availability; implicit
- Pronouns: e.g. “he”,”she”, “it”
  - Refers to item, must be “salient”

Referring Expressions

- Indefinite noun phrases (NPs): e.g. “a cat”
  - Introduces new item to discourse context
- Definite NPs: e.g. “the cat”
  - Refers to item identifiable by hearer in context
    - By verbal, pointing, or environment availability; implicit
- Pronouns: e.g. “he”,”she”, “it”
  - Refers to item, must be “salient”
- Demonstratives: e.g. “this”, “that”
  - Refers to item, sense of distance (literal/figurative)

Referring Expressions

- Indefinite noun phrases (NPs): e.g. “a cat”
  - Introduces new item to discourse context
- Definite NPs: e.g. “the cat”
  - Refers to item identifiable by hearer in context
    - By verbal, pointing, or environment availability; implicit
- Pronouns: e.g. “he”,”she”, “it”
  - Refers to item, must be “salient”
- Demonstratives: e.g. “this”, “that”
  - Refers to item, sense of distance (literal/figurative)
- Names: e.g. “Miss Woodhouse”,”IBM”
  - New or old entities

Information Status

- Some expressions (e.g. indef NPs) introduce **new** info
- Others refer to old referents (e.g. pronouns)

Information Status

- Some expressions (e.g. indef NPs) introduce **new** info
- Others refer to old referents (e.g. pronouns)

- Theories link form of refexp to given/new status

Information Status

- Some expressions (e.g. indef NPs) introduce **new** info
- Others refer to old referents (e.g. pronouns)

- Theories link form of refexp to given/new status

The givenness hierarchy:

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
{it}	{ *that* *this* *this* N }	{that N}	{the N}	{indef. *this* N}	{*a* N}

Information Status

- Some expressions (e.g. indef NPs) introduce **new** info
- Others refer to old referents (e.g. pronouns)

- Theories link form of refexp to given/new status

The givenness hierarchy:

in focus >	activated >	familiar >	uniquely identifiable >	referential >	type identifiable
{it}	{*that*, *this*, *this* N}	{that N}	{the N}	{indef. *this* N}	{*a* N}

- Accessibility:
  - More salient elements easier to call up, can be shorter
    Correlates with length: more accessible, shorter refexp

Complicating Factors

- Inferrables:
  - Refexp refers to inferentially related entity
    - *I bought a car today, but the door had a dent, and the engine was noisy.*

Complicating Factors

- Inferrables:
  - Refexp refers to inferentially related entity
    - *I bought a car today, but the door had a dent, and the engine was noisy.*
  - E.g. car -> door, engine

Complicating Factors

- Inferrables:
  - Refexp refers to inferentially related entity
    - *I bought a car today, but the door had a dent, and the engine was noisy.*
    - E.g. car -> door, engine
- Generics:
    - *I want to buy a Mac. They are very stylish.*
  - General group evoked by instance.

Complicating Factors

- Inferrables:
  - Refexp refers to inferentially related entity
    - *I bought a car today, but the door had a dent, and the engine was noisy.*
  - E.g. car -> door, engine
- Generics:
  - *I want to buy a Mac. They are very stylish.*
  - General group evoked by instance.

Complicating Factors

- Inferrables:
  - Refexp refers to inferentially related entity
    - *I bought a car today, but the door had a dent, and the engine was noisy.*
    - E.g. car -> door, engine
- Generics:
    - *I want to buy a Mac. They are very stylish.*
  - General group evoked by instance.
- Non-referential cases:
    - *It's raining.*

Syntactic Constraints for Reference Resolution

- Some fairly rigid rules constrain possible referents

Syntactic Constraints for Reference Resolution

- Some fairly rigid rules constrain possible referents
- Agreement:
  - Number: Singular/Plural

Syntactic Constraints for Reference Resolution

- Some fairly rigid rules constrain possible referents
- Agreement:
  - Number: Singular/Plural
  - Person: 1st: I,we; 2nd: you; 3rd: he, she, it, they

Syntactic Constraints for Reference Resolution

- Some fairly rigid rules constrain possible referents
- Agreement:
  - Number: Singular/Plural
  - Person: 1st: I,we; 2nd: you; 3rd: he, she, it, they
  - Gender: he vs she vs it

Syntactic & Semantic Constraints

- Binding constraints:
  - Reflexive (x-self): corefers with subject of clause
  - Pronoun/Def. NP: can't corefer with subject of clause

Syntactic & Semantic Constraints

- Binding constraints:
  - Reflexive (x-self): corefers with subject of clause
  - Pronoun/Def. NP: can't corefer with subject of clause
- "Selectional restrictions":
  - "animate": The cows eat grass.
  - "human": The author wrote the book.
  - More general: drive: John drives a car....

Syntactic & Semantic Preferences

- Recency: Closer entities are more salient
  - The doctor found an old map in the chest. Jim found an even older map on the shelf. It described an island.

Syntactic & Semantic Preferences

- Recency: Closer entities are more salient
  - The doctor found an old map in the chest. Jim found an even older map on the shelf. It described an island.
- Grammatical role: Saliency hierarchy of roles

Syntactic & Semantic Preferences

- Recency: Closer entities are more salient
  - The doctor found an old map in the chest. Jim found an even older map on the shelf. It described an island.

- Grammatical role: Saliency hierarchy of roles
  - e.g. Subj > Object > I. Obj. > Oblique > AdvP

Syntactic & Semantic Preferences

- Recency: Closer entities are more salient
  - The doctor found an old map in the chest. Jim found an even older map on the shelf. It described an island.
- Grammatical role: Saliency hierarchy of roles
  - e.g. Subj > Object > I. Obj. > Oblique > AdvP
    - Billy Bones went to the bar with Jim Hawkins. He called for a glass of rum

Syntactic & Semantic Preferences

- Recency: Closer entities are more salient
  - The doctor found an old map in the chest. Jim found an even older map on the shelf. It described an island.
- Grammatical role: Saliency hierarchy of roles
  - e.g. Subj > Object > I. Obj. > Oblique > AdvP
    - Billy Bones went to the bar with Jim Hawkins. He called for a glass of rum. [he = Billy]
    - Jim Hawkins went to the bar with Billy Bones. He called for a glass of rum

Syntactic & Semantic Preferences

- Recency: Closer entities are more salient
  - The doctor found an old map in the chest. Jim found an even older map on the shelf. It described an island.

- Grammatical role: Saliency hierarchy of roles
  - e.g. Subj > Object > I. Obj. > Oblique > AdvP
    - Billy Bones went to the bar with Jim Hawkins. He called for a glass of rum. [he = Billy]
    - Jim Hawkins went to the bar with Billy Bones. He called for a glass of rum. [he = Jim]

Syntactic & Semantic Preferences

- Repeated reference: Pronouns more salient
  - Once focused, likely to continue to be focused
    - Billy Bones had been thinking of a glass of rum. He hobbled over to the bar. Jim Hawkins went with him. He called for a glass of rum.

Syntactic & Semantic Preferences

- Repeated reference: Pronouns more salient
  - Once focused, likely to continue to be focused
    - Billy Bones had been thinking of a glass of rum. He hobbled over to the bar. Jim Hawkins went with him. He called for a glass of rum. [he=Billy]
- Parallelism: Prefer entity in same role

Syntactic & Semantic Preferences

- Repeated reference: Pronouns more salient
  - Once focused, likely to continue to be focused
    - Billy Bones had been thinking of a glass of rum. He hobbled over to the bar. Jim Hawkins went with him. He called for a glass of rum. [he=Billy]
- Parallelism: Prefer entity in same role
    - Silver went with Jim to the bar. Billy Bones went with him to the inn.

Syntactic & Semantic Preferences

- Repeated reference: Pronouns more salient
  - Once focused, likely to continue to be focused
    - Billy Bones had been thinking of a glass of rum. He hobbled over to the bar. Jim Hawkins went with him. He called for a glass of rum. [he=Billy]
- Parallelism: Prefer entity in same role
    - Silver went with Jim to the bar. Billy Bones went with him to the inn. [him = Jim]
    - Overrides grammatical role
- Verb roles: "implicit causality", thematic role match,...
    - John telephoned Bill. He lost the laptop.

Syntactic & Semantic Preferences

- Repeated reference: Pronouns more salient
  - Once focused, likely to continue to be focused
    - Billy Bones had been thinking of a glass of rum. He hobbled over to the bar. Jim Hawkins went with him. He called for a glass of rum. [he=Billy]
- Parallelism: Prefer entity in same role
    - Silver went with Jim to the bar. Billy Bones went with him to the inn. [him = Jim]
    - Overrides grammatical role
- Verb roles: "implicit causality", thematic role match,...
    - John telephoned Bill. He lost the laptop. [He=John]
    - John criticized Bill. He lost the laptop.

Syntactic & Semantic Preferences

- Repeated reference: Pronouns more salient
  - Once focused, likely to continue to be focused
    - Billy Bones had been thinking of a glass of rum. He hobbled over to the bar. Jim Hawkins went with him. He called for a glass of rum. [he=Billy]
- Parallelism: Prefer entity in same role
    - Silver went with Jim to the bar. Billy Bones went with him to the inn. [him = Jim]
    - Overrides grammatical role
- Verb roles: "implicit causality", thematic role match,...
    - John telephoned Bill. He lost the laptop. [He=John]
    - John criticized Bill. He lost the laptop. [He=Bill]

Reference Resolution Approaches

- Common features
  - "Discourse Model"
    - Referents evoked in discourse, available for reference
    - Structure indicating relative salience
  - Syntactic & Semantic Constraints
  - Syntactic & Semantic Preferences
- Differences:
  - Which constraints/preferences? How combine? Rank?

A Resolution Algorithm (Lappin & Leass)

- Discourse model update:
  - Evoked entities:
    - Equivalence classes: Coreferent referring expressions
  - Salience value update:
    - Weighted sum of salience values:
      - Based on syntactic preferences

A Resolution Algorithm

- Pronoun resolution:
  - Collect potential referents (4 sent back)
  - Exclude referents that violate agreement constraints
  - Exclude referents that violate binding constraints
  - Compute salience by adding new weights to old
  - Select referent with highest salience value
    - Ties broken by distance (abs. value)

Salience Factors (Lappin & Leass 1994)

- Weights empirically derived from corpus
  - Recency: 100
  - Subject: 80
  - Existential: 70
  - Object: 50
  - Indirect Object/Oblique: 40
  - Non-adverb PP: 50
  - Head noun: 80
  - Parallelism: 35, Cataphora: -175
- Divide by 50% for each sentence distance

Example

- John saw a beautiful Acura Integra in the dealership.
- He showed it to Bob.
- He bought it.

Example

- John saw a beautiful Acura Integra in the dealership.

Rec	Subj	Exist	Obj	Ind-Obj	Non-Adv	Head N
100	80	70	50	40	50	80

Referent	Phrases	Value
John	{John}	310
Integra	{a beautiful Acura Integra}	280
Dealership	{the dealership}	230

Example

- He showed it to Bob.

Referent	Phrases	Value
John	{John, he1}	465
Integra	{a beautiful Acura Integra}	140
Dealership	{the dealership}	115

Referent	Phrases	Value
John	{John, he1}	465
Integra	{a beautiful Acura Integra,it1}	420
Dealership	{the dealership}	115

Example

- He showed it to Bob.

Referent	Phrases	Value
John	{John, he1}	465
Integra	{a beautiful Acura Integra}	420
Bob	{Bob}	270
Dealership	{the dealership}	115

Example

- He bought it.

Referent	Phrases	Value
John	{John, he1}	232.5
Integra	{a beautiful Acura Integra}	210
Bob	{Bob}	135
Dealership	{the dealership}	57.5

Referent	Phrases	Value
John	{John, he1}	542.5
Integra	{a beautiful Acura Integra}	490
Bob	{Bob}	135
Dealership	{the dealership}	57.5

Lapping & Leass Results

- Weights trained on corpus of computer training manuals
- Tested on held-out set in similar domains
- Accuracy: 86%

Reference Resolution Algorithms

- Many other alternative strategies:
  - Linguistically informed, saliency hierarchy
    - Centering Theory (Walker et al
  - Linguistically informed, tree based, recency, saliency
    - Hobbs algorithm
  - Shallow processing, simple heuristic, high precision:
    - Cogniac (Baldwin 2000)

Heuristic Reference Resolution: Agreements

- Knowledge-based
  - Deep analysis: full parsing, semantic analysis
  - Enforce syntactic/semantic constraints
  - Preferences:
    - Recency
    - Grammatical Role Parallelism (ex. Hobbs)
    - Role ranking
    - Frequency of mention
- Local reference resolution
- Little/No world knowledge
- Similar levels of effectiveness

Data-driven Reference Resolution

- Prior approaches: Knowledge-based, hand-crafted

Data-driven Reference Resolution

- Prior approaches: Knowledge-based, hand-crafted
- Data-driven machine learning approach
  - Coreference as classification, clustering, ranking problem

Data-driven Reference Resolution

- Prior approaches: Knowledge-based, hand-crafted
- Data-driven machine learning approach
  - Coreference as classification, clustering, ranking problem
    - Mention-pair model:
      - For each pair NPi,NPj, do they corefer?
      - Cluster to form equivalence classes

Data-driven Reference Resolution

- Prior approaches: Knowledge-based, hand-crafted
- Data-driven machine learning approach
  - Coreference as classification, clustering, ranking problem
    - Mention-pair model:
      - For each pair NPi,NPj, do they corefer?
      - Cluster to form equivalence classes
    - Entity-mention model
      - For each pair NP_k and cluster $C_{j,,}$ should the NP be in the cluster?

Data-driven Reference Resolution

- Prior approaches: Knowledge-based, hand-crafted
- Data-driven machine learning approach
  - Coreference as classification, clustering, ranking problem
    - Mention-pair model:
      - For each pair NPi,NPj, do they corefer?
      - Cluster to form equivalence classes
    - Entity-mention model
      - For each pair NP_k and cluster $C_{j,,}$ should the NP be in the cluster?
    - Ranking models
      - For each NP_k, and all candidate antecedents, which highest?

NP Coreference Examples

- Link all NPs refer to 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...

Example from Cardie&Ng 2004

Annotated Corpora

- Available shared task corpora
  - MUC-6, MUC-7 (Message Understanding Conference)
    - 60 documents each, newswire, English

Annotated Corpora

- Available shared task corpora
  - MUC-6, MUC-7 (Message Understanding Conference)
    - 60 documents each, newswire, English
  - ACE (Automatic Content Extraction)
    - Originally English newswite
    - Later include Chinese, Arabic; blog, CTS, usenet, etc

Annotated Corpora

- Available shared task corpora
  - MUC-6, MUC-7 (Message Understanding Conference)
    - 60 documents each, newswire, English
  - ACE (Automatic Content Extraction)
    - Originally English newswite
    - Later include Chinese, Arabic; blog, CTS, usenet, etc
- Treebanks
  - English Penn Treebank (Ontonotes)

Annotated Corpora

- Available shared task corpora
  - MUC-6, MUC-7 (Message Understanding Conference)
    - 60 documents each, newswire, English
  - ACE (Automatic Content Extraction)
    - Originally English newswite
    - Later include Chinese, Arabic; blog, CTS, usenet, etc
- Treebanks
  - English Penn Treebank (Ontonotes)
  - German, Czech, Japanese, Spanish, Catalan, Medline

Feature Engineering I

- Information similar to heuristics
  - Recency: distance between mentions
  - Grammatical salience: role ranking
  - Grammatical constraints: agreement features, binding

Feature Engineering I

- Information similar to heuristics
  - Recency: distance between mentions
  - Grammatical salience: role ranking
  - Grammatical constraints: agreement features, binding
- Heuristic techniques themselves:
  - Rank from Hobbs algorithm

Feature Engineering I

- Information similar to heuristics
  - Recency: distance between mentions
  - Grammatical salience: role ranking
  - Grammatical constraints: agreement features, binding
- Heuristic techniques themselves:
  - Rank from Hobbs algorithm
- Discourse segment boundaries

Feature Engineering (II)

- Other coreference (not pronominal) features

Feature Engineering (II)

- Other coreference (not pronominal) features
  - String-matching features:
    - Mrs. Clinton <->Clinton

Feature Engineering (II)

- Other coreference (not pronominal) features
  - String-matching features:
    - Mrs. Clinton <->Clinton
  - Semantic features:
    - Can candidate appear in same role w/same verb?
    - WordNet similarity
    - Wikipedia: broader coverage

Feature Engineering (II)

- Other coreference (not pronominal) features
  - String-matching features:
    - Mrs. Clinton <->Clinton
  - Semantic features:
    - Can candidate appear in same role w/same verb?
    - WordNet similarity
    - Wikipedia: broader coverage
  - Lexico-syntactic patterns:
    - E.g. X is a Y

Typical Feature Set

- 25 features per instance: 2NPs, features, class
  - lexical (3)
    - string matching for pronouns, proper names, common nouns
  - grammatical (18)
    - pronoun_1, pronoun_2, demonstrative_2, indefinite_2, …
    - number, gender, animacy
    - appositive, predicate nominative
    - binding constraints, simple contra-indexing constraints, …
    - span, maximalnp, …
  - semantic (2)
    - same WordNet class
    - alias
  - positional (1)
    - distance between the NPs in terms of # of sentences
  - knowledge-based (1)
    - naïve pronoun resolution algorithm

Coreference Evaluation

- Key issues:
  - Which NPs are evaluated?
    - Gold standard tagged or
    - Automatically extracted

Coreference Evaluation

- Key issues:
  - Which NPs are evaluated?
    - Gold standard tagged or
    - Automatically extracted
  - How good is the partition?
    - Any cluster-based evaluation could be used (e.g. Kappa)
    - MUC scorer:
      - Link-based: ignores singletons; penalizes large clusters
      - Other measures compensate

Classify & Cluster Coreference

- Classification:
  - For each pair of candidate coreferential NPs (NP_i,NP_j), classify as +/- coreferent

MUC-6 Data Set

```
ALIAS = C: +
ALIAS = I:
| SOON_STR_NONPRO = C:
| | ANIMACY = NA: -
| | ANIMACY = I: -
| | ANIMACY = C: +
| SOON_STR_NONPRO = I:
| | PRO_STR = C: +
| | PRO_STR = I:
| | | PRO_RESOLVE = C:
| | | | EMBEDDED_1 = Y: -
| | | | EMBEDDED_1 = N:
| | | | | PRONOUN_1 = Y:
| | | | | | ANIMACY = NA: -
| | | | | | ANIMACY = I: -
| | | | | | ANIMACY = C: +
| | | | | PRONOUN_1 = N:
| | | | | | MAXIMALNP = C: +
| | | | | | MAXIMALNP = I:
| | | | | | | WNCLASS = NA: -
| | | | | | | WNCLASS = I: +
| | | | | | | WNCLASS = C: +
| | | PRO_RESOLVE = I:
| | | | APPOSITIVE = I: -
| | | | APPOSITIVE = C:
| | | | | GENDER = NA: +
| | | | | GENDER = I: +
| | | | | GENDER = C: -
```

Unsupervised Approach to Coreference Resolution

- Cardie and Wagstaff
- Coreference as clustering:

Unsupervised Approach to Coreference Resolution

- Cardie and Wagstaff
- Coreference as clustering:
  - For a given text, partition all NP mentions
    - Cluster = Entity

Unsupervised Approach to Coreference Resolution

- Cardie and Wagstaff
- Coreference as clustering:
  - For a given text, partition all NP mentions
    - Cluster = Entity
  - Requires a distance metric
    - Coreferential NPs should be 'close'
    - Non-coreferential NPs should be farther apart

Unsupervised Approach to Coreference Resolution

- Cardie and Wagstaff
- Coreference as clustering:
  - For a given text, partition all NP mentions
    - Cluster = Entity
  - Requires a distance metric
    - Coreferential NPs should be 'close'
    - Non-coreferential NPs should be farther apart
  - Evaluate partition

Why Unsupervised Clustering?

- Unsupervised approach:

Why Unsupervised Clustering?

- Unsupervised approach:
  - Doesn’t rely on large, labeled training corpus
  - Less sensitive to label skew

Why Unsupervised Clustering?

- Unsupervised approach:
  - Doesn't rely on large, labeled training corpus
  - Less sensitive to label skew

- Clustering:

Why Unsupervised Clustering?

- Unsupervised approach:
  - Doesn’t rely on large, labeled training corpus
  - Less sensitive to label skew
- Clustering:
  - Fairly natural match to coreference problem
    - Group all mentions talking about the same thing

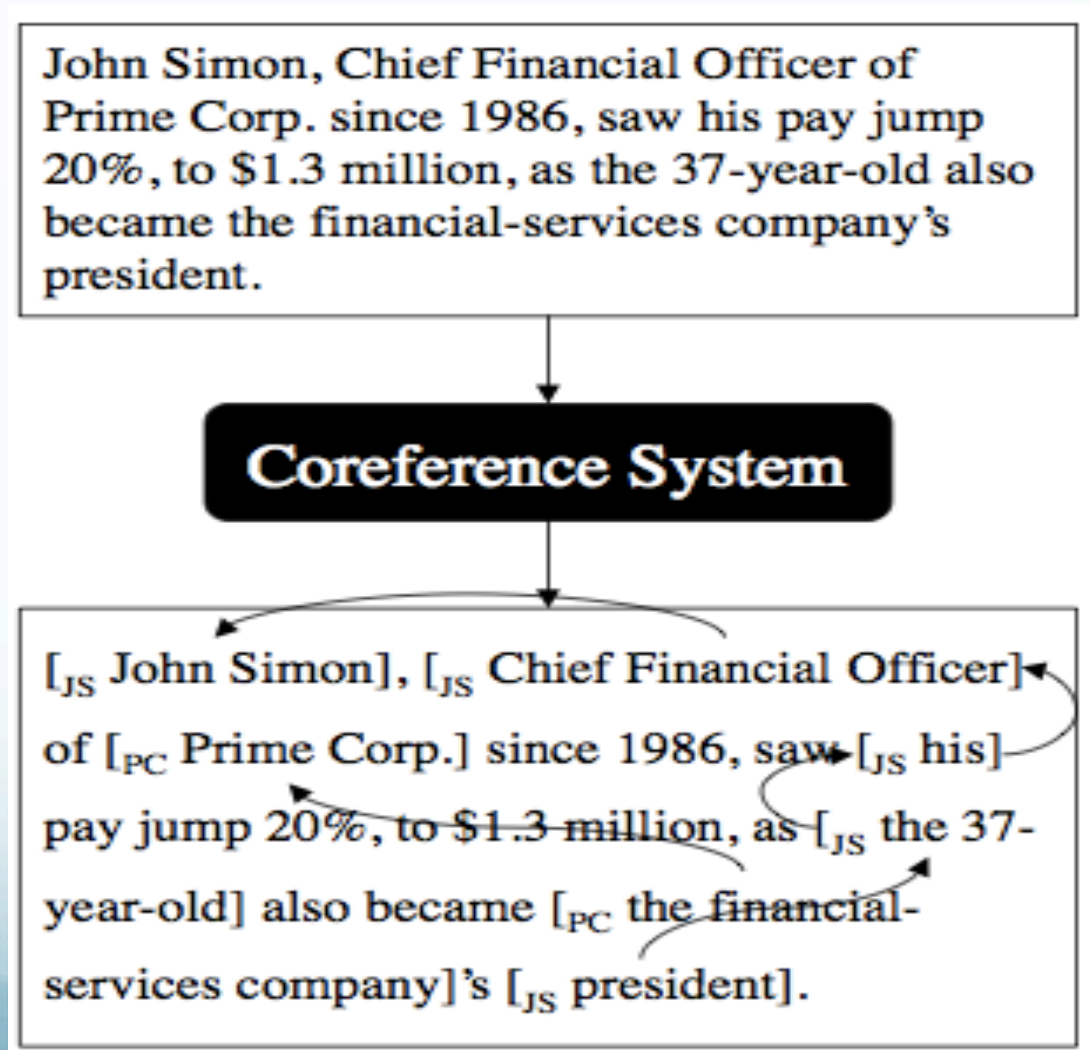
Why Unsupervised Clustering?

- Unsupervised approach:
  - Doesn't rely on large, labeled training corpus
  - Less sensitive to label skew
- Clustering:
  - Fairly natural match to coreference problem
    - Group all mentions talking about the same thing
  - Avoids some 'hard' classification decisions of other techniques
  - Can make global partition decisions

Instance Representation

- Automatically extracted base NPS
- 11 Features
  - Word in NP, head noun in NP
  - Position of NP (index) in text
  - Pronoun type (acc, nom, poss, none)
  - Article type (indef, def, none)
  - In Appositive phrase
  - Number, gender, animacy
  - Proper noun: Y/N
  - Semantic class

Example Text



Representation of Text

Words, Head Noun (in **bold**)	Position	Pronoun Type	Article	Appositive	Number	Proper Name	Semantic Class	Gender	Animacy
John **Simon**	1	NONE	NONE	NO	SING	YES	HUMAN	MASC	ANIM
Chief Financial **Officer**	2	NONE	NONE	NO	SING	NO	HUMAN	EITHER	ANIM
Prime **Corp.**	3	NONE	NONE	NO	SING	NO	COMPANY	NEUTER	INANIM
1986	4	NONE	NONE	NO	PLURAL	NO	NUMBER	NEUTER	INANIM
his	5	POSS	NONE	NO	SING	NO	HUMAN	MASC	ANIM
pay	6	NONE	NONE	NO	SING	NO	PAYMENT	NEUTER	INANIM
20%	7	NONE	NONE	NO	PLURAL	NO	PERCENT	NEUTER	INANIM
$1.3 **million**	8	NONE	NONE	NO	PLURAL	NO	MONEY	NEUTER	INANIM
the **37-year-old**	9	NONE	DEF	NO	SING	NO	HUMAN	EITHER	ANIM
the financial-services **company**	10	NONE	DEF	NO	SING	NO	COMPANY	NEUTER	INANIM
president	11	NONE	NONE	NO	SING	NO	HUMAN	EITHER	ANIM

Distance Measure

- Distance measure:
  - Weighted sum of 'incompatibility' features between NPs
    - Positive infinite weights: block clustering
    - Negative infinite weights: cluster, unless blocked
    - Weight = r: avoid coreference if incompatible
    - Others, heuristic

Distance Measure

- Distance measure:
  - Weighted sum of 'incompatibility' features between NPs
    - Positive infinite weights: block clustering
    - Negative infinite weights: cluster, unless blocked
    - Weight = r: avoid coreference if incompatible
    - Others, heuristic
- If distance > r (cluster radius), non-coref

Distance Weights

Feature f	Weight	Incompatibility function
Words	10.0	(# of mismatching words[a]) / (# of words in the longer NP)
Head Noun	1.0	1 if the head nouns differ; else 0
Position	5.0	(difference in position) / (maximum difference in document)
Pronoun	r	1 if NP_i is a pronoun and NP_j is not; else 0
Article	r	1 if NP_j is indefinite and not appositive; else 0
Words–Substring	$-\infty$	1 if NP_i subsumes (entirely includes as a substring) NP_j;
Appositive	$-\infty$	1 if NP_j is appositive and NP_i is its immediate predecessor; else 0
Number	∞	1 if they do not match in number; else 0
Proper Name	∞	1 if both are proper names, but mismatch on every word; else 0
Semantic Class	∞	1 if they do not match in class; else 0
Gender	∞	1 if they do not match in gender (allows EITHER to match MASC or FEM); else 0
Animacy	∞	1 if they do not match in animacy; else 0

Clustering

- Basic algorithm:
  - Initialize: Each NP is its own class
  - Working from End of text to Beginning
    - Compute the distance *d* between the two NPS
    - If $d < r$ AND no members of the classes are incompatible
      - Merge the classes

Clustering

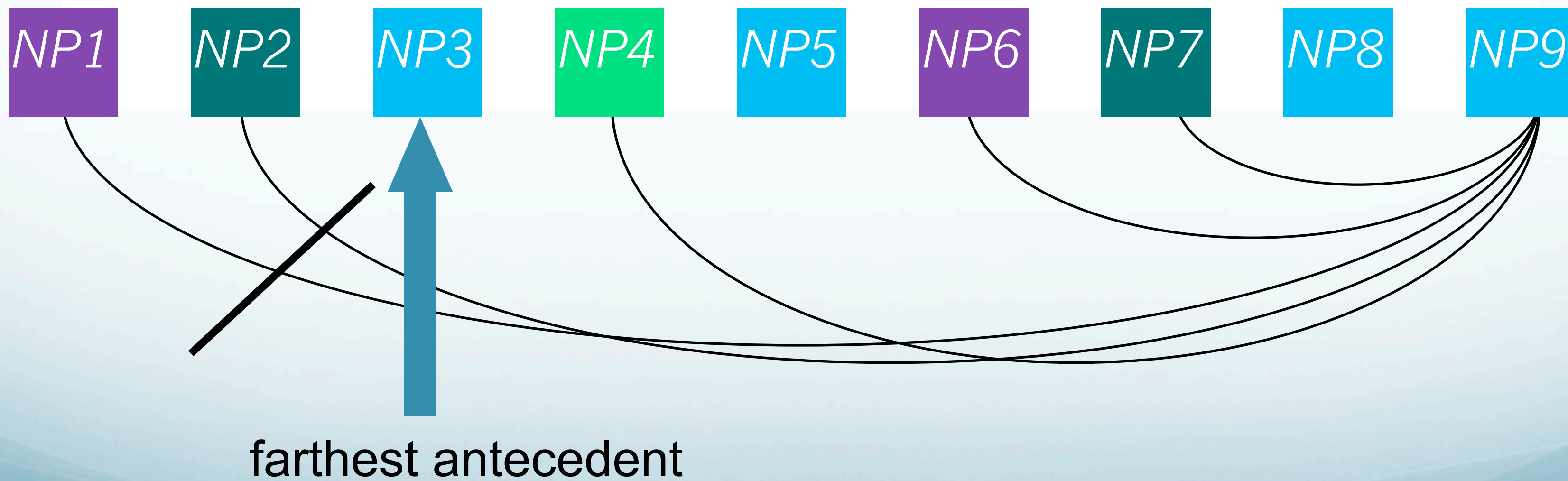
- Basic algorithm:
  - Initialize: Each NP is its own class
  - Working from End of text to Beginning
    - Compute the distance *d* between the two NPS
    - If $d < r$ AND no members of the classes are incompatible
      - Merge the classes
- F-measure: 0.53
  - Decent:
    - Limited by:
      - Automatic NP extraction: 0.67 if perfect
      - inaccurate features, non-ref. pronoun

Clustering by Classification

- Ng and Cardie (2002)
- Baseline mention-pair style system:
  - For each pair of NPs, classify +/- coreferent
  - Linked pairs form coreferential chains
    - Process candidate pairs from End to Start
    - All mentions of an entity appear in single chain
- Improve with
  - Better training set selection
  - Better clustering approach
  - Better feature set

Problem 1

- Coreference is a rare relation
  - skewed class distributions (2% positive instances)
  - *remove some negative instances*



Problem 2

- Coreference is a discourse-level problem
  - different solutions for different types of NPs
    - proper names: string matching and aliasing
  - inclusion of "hard" positive training instances
  - *positive example selection*: selects easy positive training instances (cf. Harabagiu *et al.* (2001))
    - Select most confident antecedent as positive instance

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

Problem 3

- Coreference is an equivalence relation
  - loss of transitivity
  - need to tighten the connection between classification and clustering
  - *prune learned rules w.r.t. the clustering-level coreference scoring function*

coref ? *coref* ?

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

not coref ?

Results Snapshot

System Variation	MUC-6 R	MUC-6 P	MUC-6 F	MUC-7 R	MUC-7 P	MUC-7 F
Original Soon et al.	58.6	67.3	62.6	56.1	65.5	60.4
Duplicated Soon Baseline	62.4	70.7	66.3	55.2	68.5	61.2
Learning Framework	62.4	73.5	67.5	56.3	71.5	63.0
String Match	60.4	74.4	66.7	54.3	72.1	62.0
Training Instance Selection	61.9	70.3	65.8	55.2	68.3	61.1
Clustering	62.4	70.8	66.3	56.5	69.6	62.3
All Features	70.3	58.3	63.8	65.5	58.2	61.6
Pronouns only	–	66.3	–	–	62.1	–
Proper Nouns only	–	84.2	–	–	77.7	–
Common Nouns only	–	40.1	–	–	45.2	–
Hand-selected Features	64.1	74.9	69.1	57.4	70.8	63.4
Pronouns only	–	67.4	–	–	54.4	–
Proper Nouns only	–	93.3	–	–	86.6	–
Common Nouns only	–	63.0	–	–	64.8	–

Classification & Clustering

- Classifiers:
  - C4.5 (Decision Trees)
  - RIPPER – automatic rule learner

Classification & Clustering

- Classifiers:
  - C4.5 (Decision Trees), RIPPER
- Cluster: Best-first, single link clustering
  - Each NP in own class
  - Test preceding NPs
  - Select highest confidence coreferent, merge classes

Baseline Feature Set

Feature Type	Feature
Lexical	SOON_STR
Grammatical	PRONOUN_1*
	PRONOUN_2*
	DEFINITE_2
	DEMONSTRATIVE_2
	NUMBER*
	GENDER*
	BOTH_PROPER_NOUNS*
	APPOSITIVE*
Semantic	WNCLASS*
	ALIAS*
Positional	SENTNUM*

Extended Feature Set

- Explore 41 additional features
  - More complex NP matching (7)
  - Detail NP type (4) – definite, embedded, pronoun,..
  - Syntactic Role (3)
  - Syntactic constraints (8) – binding, agreement, etc
  - Heuristics (9) – embedding, quoting, etc
  - Semantics (4) – WordNet distance, inheritance, etc
  - Distance (1) – in paragraphs
  - Pronoun resolution (2)
    - Based on simple or rule-based resolver

Feature Selection

- Too many added features
  - Hand select ones with good coverage/precision

Feature Selection

- Too many added features
  - Hand select ones with good coverage/precision

- Compare to automatically selected by learner
  - Useful features are:
    - Agreement
    - Animacy
    - Binding
    - Maximal NP
      - Reminiscent of Lappin & Leass

Feature Selection

- Too many added features
  - Hand select ones with good coverage/precision

- Compare to automatically selected by learner
  - Useful features are:
    - Agreement
    - Animacy
    - Binding
    - Maximal NP
      - Reminiscent of Lappin & Leass
- Still best results on MUC-7 dataset: 0.634

Weakly Supervised Learning

- Exploit small pool of labeled training data
  - Larger pool unlabeled

Weakly Supervised Learning

- Exploit small pool of labeled training data
  - Larger pool unlabeled
- Single-View Multi-Learner Co-training
  - 2 different learning algorithms, same feature set

Weakly Supervised Learning

- Exploit small pool of labeled training data
  - Larger pool unlabeled
- Single-View Multi-Learner Co-training
  - 2 different learning algorithms, same feature set
  - each classifier labels unlabeled instances for the other classifier
  - data pool is flushed after each iteration

Summary

- Constraints and preferences for reference resolution
- Resolution algorithms:
  - Heuristic approaches
  - Machine Learning approaches
    - Unsupervised, supervised semi-supervised
- Similar knowledge sources
  - Different implementations

Contrasts

- Heuristic pronominal resolution
- Vs
- Machine learning for coreference

Contrasts

- Heuristic pronominal resolution
- Vs
- Machine learning for coreference
- Questions:
  - How are these approaches influenced by differences in:
    - Data type:

Contrasts

- Heuristic pronominal resolution
- Vs
- Machine learning for coreference
- Questions:
  - How are these approaches influenced by differences in:
    - Data type:
      - Newswire text, Broadcast news
      - Conversational speech
        - Telephone, Face-to-face
      - Human-computer dialogue
      - Specific language

Projects

- Which elective?
- Collaboration?
- Broad areas:
  - Reference and resolution
  - Discourse structure
  - Dialogue modeling and understanding
  - Dialogue systems

Topic Ideas: Linguistic

- Analyze reference behavior in a:
  - Different language
  - Different register/style
    - E.g. patterns of pronominal reference in Chat/IM/...
- Investigate conversation style in SDS
  - Politeness, misunderstandings, vocabulary use,...
- Evaluate predictions for dialogue behavior
  - Amount of overlap and register/familiarity/language
- Analyze in depth a set of discourse structure models

Topic Ideas: Computational

- Implement a spoken language interface to...
- Implement/extend a discourse segmentation algorithm
- Develop an automatic recognition system for some aspect of speaking style – drunkenness?
- Improve dialogue act recognition by improving the modeling of dialogue history

Centering

- Identify the local "center" of attention
  - Pronominalization focuses attention, appropriate use establishes coherence
- Identify entities available for reference
- Describe shifts in what discourse is about
  - Prefer different types for coherence

Centering: Structures

- Each utterance (Un) has:
  - List of forward-looking centers: Cf(Un)
    - Entities realized/evoked in Un
    - Rank by likelihood of focus of future discourse
    - Highest ranked element: Cp(Un)

  - Backward looking center (focus): Cb(Un)

Centering: Transitions

	Cb(Un)=Cb(Un-1)	Cb(Un) != Cb(Un-1)
Cb(Un)=Cp(Un)	Continuing	Smooth Shift
Cb(Un)!=Cp(Un)	Retaining	Rough Shift

Centering: Constraints and Rules

- Constraints:
  - Exactly ONE backward -looking center
  - Everything in Cf(Un) realized in Un
  - Cb(Un): highest ranked item in Cf(Un) in Un-1
- Rules:
  - If any item in Cf(Un-1) realized as pronoun in Un, Cb(Un) must be realized as pronoun
  - Transitions are ranked:
    - Continuing > Retaining > Smooth Shift > Rough Shift

Centering: Example

- John saw a beautiful Acura Integra at the dealership
  - Cf: (John, Integra, dealership); No Cb
- He showed it to Bill.
  - Cf:(John/he, Integra/it*, Bill); Cb: John/he
- He bought it:
  - Cf: (John/he, Integra/it); Cb: John/he

CogNIAC

- Goal: Resolve with high precision
  - Identify where ambiguous, use no world knowledge, simple syntactic analysis
  - Precision: # correct labelings/# of labelings
  - Recall: # correct labelings/# of anaphors
- Uses simple set of ranked rules
  - Applied incrementally left-to-right
- Designed to work on newspaper articles
  - Tune/rank rules

CogNIAC: Rules

- Only resolve reference if unique antecedent
- 1) Unique in prior discourse
- 2) Reflexive: nearest legal in same sentence
- 3) Unique in current & prior:
- 4) Possessive Pro: single exact poss in prior
- 5) Unique in current
- 6) Unique subj/subj pronoun

CogNIAC: Example

- John saw a beautiful Acura Integra in the dealership.
- He showed it to Bill.
  - He= John : Rule 1; it -> ambiguous (Integra)
- He bought it.
  - He=John: Rule 6; it=Integra: Rule 3