Discourse: Coreference

Deep Processing Techniques for NLP
Ling 571
March 5, 2014

Roadmap

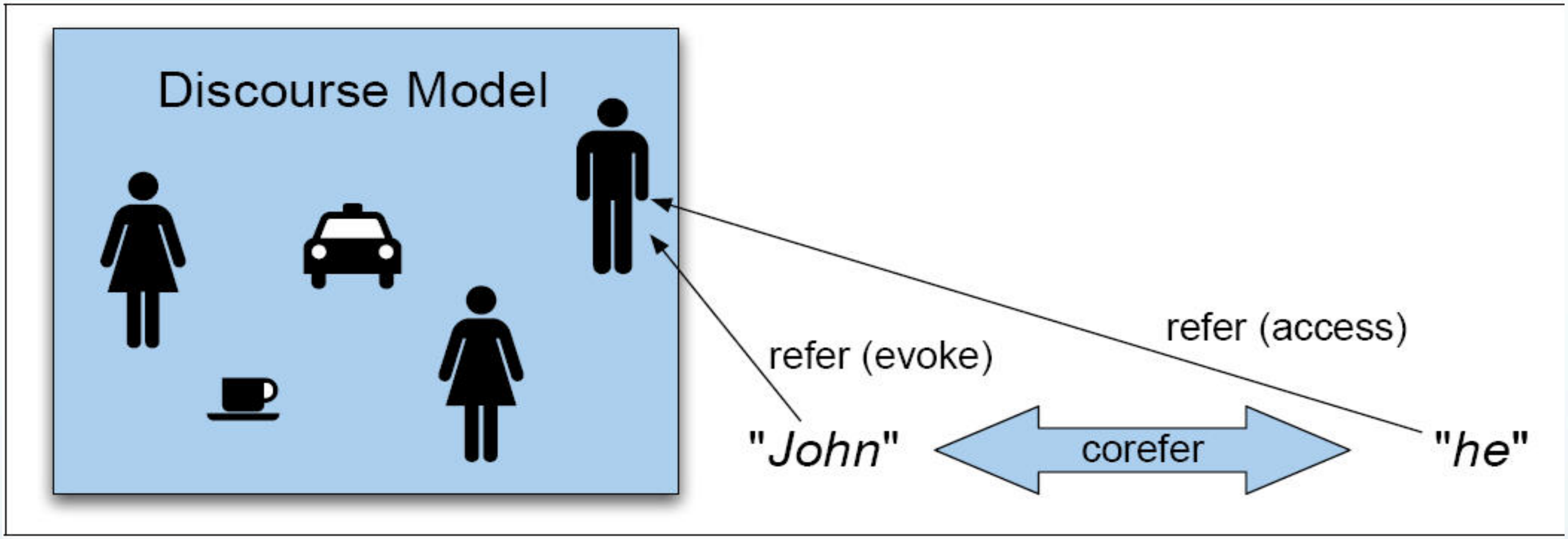
- Coreference
  - Referring expressions

  - Syntactic & semantic constraints
  - Syntactic & semantic preferences

  - Reference resolution:
    - Hobbs Algorithm: Baseline
    - Machine learning approaches
    - Sieve models

  - Challenges

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
- 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)

- 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.*
  - 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
- 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

- "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.

- 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. [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?

Hobbs' Resolution Algorithm

- Requires:
  - Syntactic parser
  - Gender and number checker
- Input:
  - Pronoun
  - Parse of current and previous sentences
- Captures:
  - Preferences: Recency, grammatical role
  - Constraints: binding theory, gender, person, number

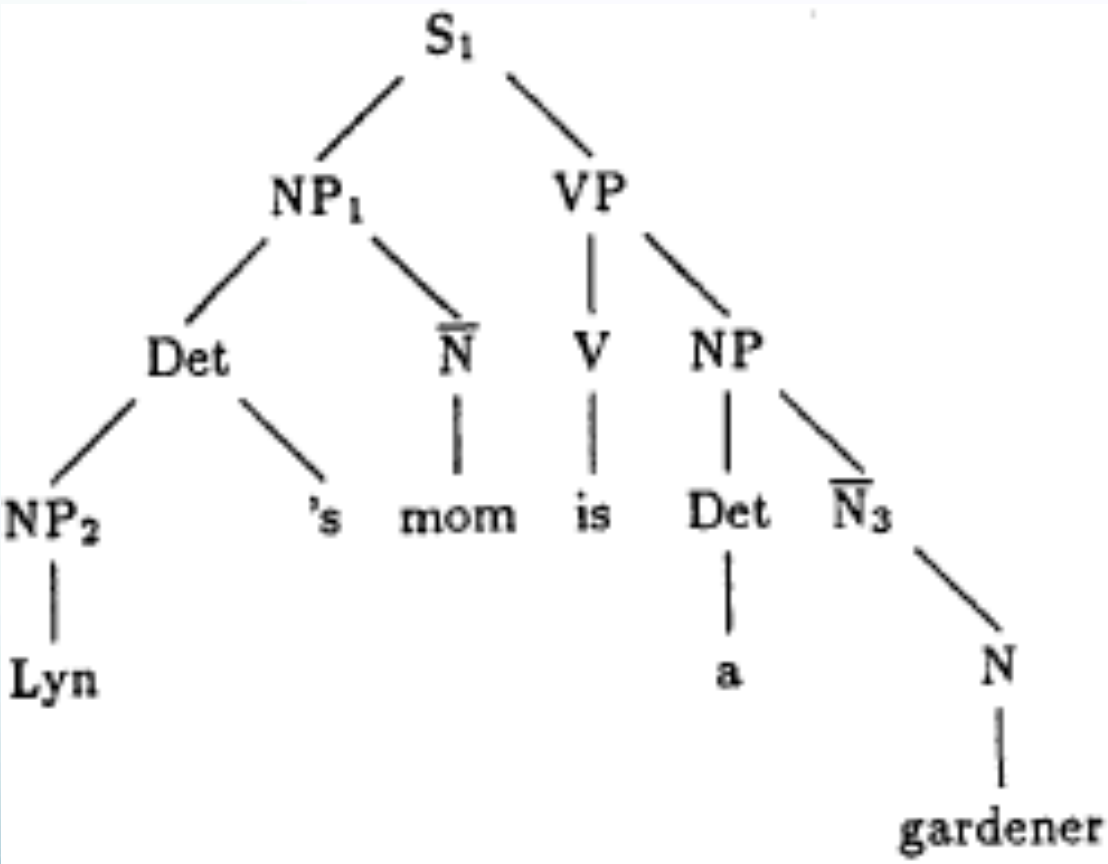
Hobbs Algorithm

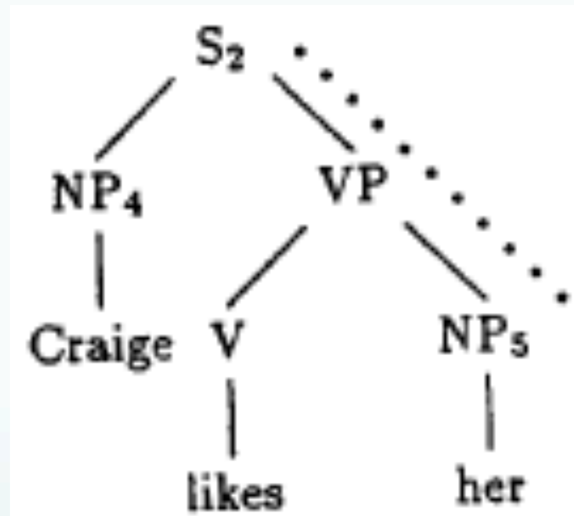
- Intuition:
  - Start with target pronoun
  - Climb parse tree to S root
  - For each NP or S
    - Do breadth-first, left-to-right search of children
      - Restricted to left of target
    - For each NP, check agreement with target
  - Repeat on earlier sentences until matching NP found

Hobbs Algorithm Detail

- Begin at NP immediately dominating pronoun
- Climb tree to NP or S: X=node, p = path
- Traverse branches below X, and left of p: BF, LR
  - If find NP, propose as antecedent
    - If separated from X by NP or S
- Loop: If X highest S in sentence, try previous sentences.
- If X not highest S, climb to next NP or S: X = node
- If X is NP, and p not through X's nominal, propose X
- Traverse branches below X, left of p: BF,LR
  - Propose any NP
- If X is S, traverse branches of X, right of p: BF, LR
    - Do not traverse NP or S; Propose any NP
  - Go to Loop

Hobbs Example





Lyn's mom is a gardener. Craige likes her.

Another Hobbs Example

- The castle in Camelot remained the residence of the King until 536 when he moved it to London.

- What is **it**?
  - residence

Another Hobbs Example

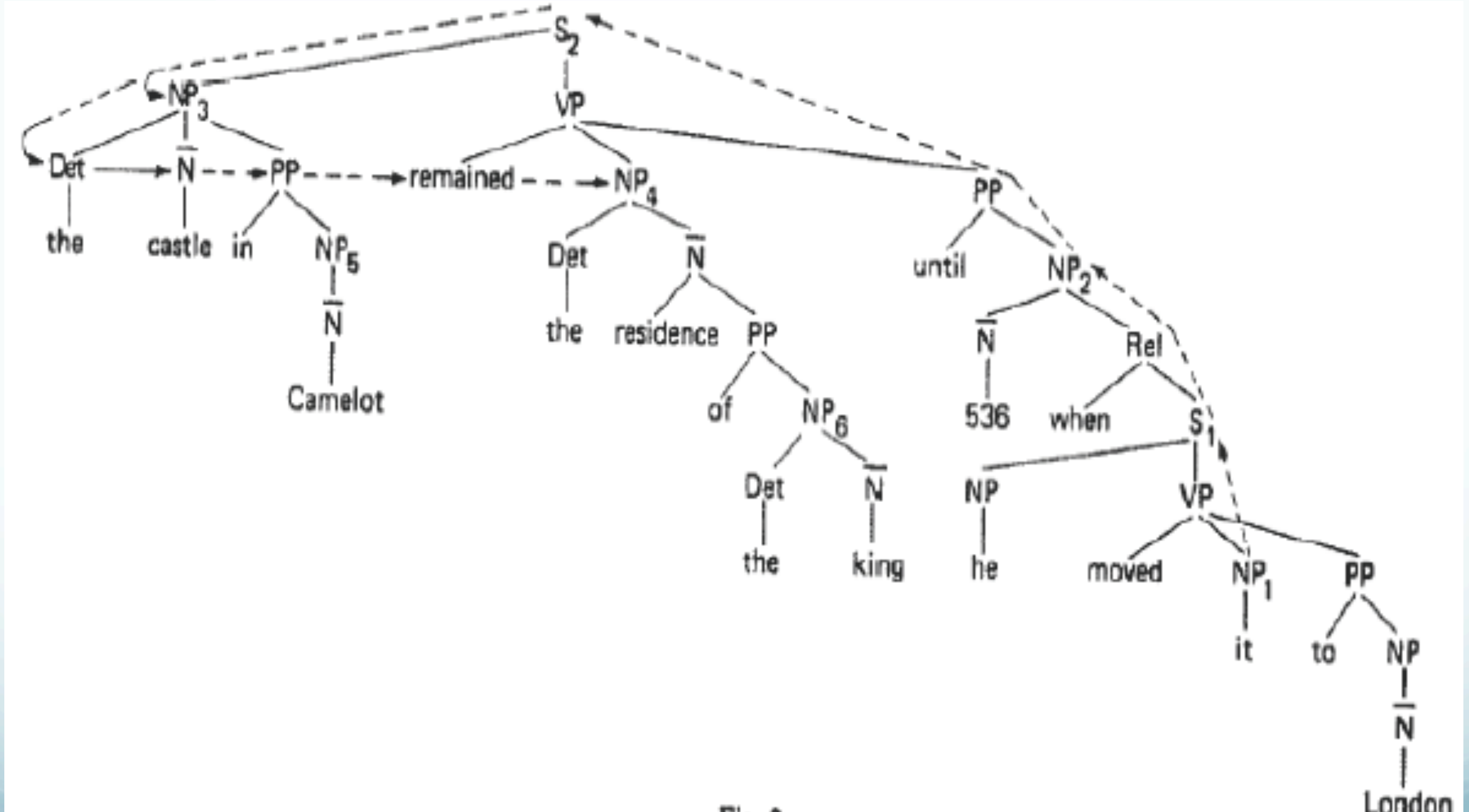


Fig. 2

Hobbs, 1978

Hobbs Algorithm

- Results: 88% accuracy ; 90+% intrasentential
  - On perfect, manually parsed sentences
- Useful baseline for evaluating pronominal anaphora
- Issues:
  - Parsing:
    - Not all languages have parsers
    - Parsers are not always accurate
  - Constraints/Preferences:
    - Captures: Binding theory, grammatical role, recency
    - But not: parallelism, repetition, verb semantics, selection

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
  - 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

- 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
  - 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

Clustering by Classification

- Mention-pair style system:

  - For each pair of NPs, classify +/- coreferent
    - Any classifier

  - Linked pairs form coreferential chains
    - Process candidate pairs from End to Start
    - All mentions of an entity appear in single chain

  - F-measure: MUC-6: 62-66%; MUC-7: 60-61%
    - Soon et. al, Cardie and Ng (2002)

Multi-pass Sieve Approach

- Raghunathan et al., 2010

- Key Issues:
  - Limitations of mention-pair classifier approach
    - Local decisions over large number of features
      - Not really transitive
      - Can't exploit global constraints
      - Low precision features may overwhelm less frequent, high precision ones

Multi-pass Sieve Strategy

- Basic approach:
  - Apply tiers of deterministic coreference modules
    - Ordered highest to lowest precision
  - Aggregate information across mentions in cluster
    - Share attributes based on prior tiers
  - Simple, extensible architecture
    - Outperforms many other (un-)supervised approaches

Pre-Processing and Mentions

- Pre-processing:
  - Gold mention boundaries given, parsed, NE tagged
- For each mention, each module can skip or pick best candidate antecedent
  - Antecedents ordered:
    - Same sentence: by Hobbs algorithm
    - Prev. sentence:
      - For Nominal: by right-to-left, breadth first: proximity/recency
      - For Pronoun: left-to-right: salience hierarchy
    - W/in cluster: aggregate attributes, order mentions
    - Prune indefinite mentions: can't have antecedents

Multi-pass Sieve Modules

- Pass 1: Exact match (N): P: 96%
- Pass 2: Precise constructs
  - Predicate nominative, (role) appositive, re;. pronoun, acronym, demonym
- Pass 3: Strict head matching
  - Matches cluster head noun AND all non-stop cluster wds AND modifiers AND non i-within-I (embedded NP)
- Pass 4 & 5: Variants of 3: drop one of above

Multi-pass Sieve Modules

- Pass 6: Relaxed head match
  - Head matches any word in cluster AND all non-stop cluster wds AND non i-within-I (embedded NP)
- Pass 7: Pronouns
  - Enforce constraints on gender, number, person, animacy, and NER labels

Multi-pass Effectiveness

Passes	MUC P	MUC R	MUC F1
{1}	95.9	31.8	47.8
{1,2}	95.4	43.7	59.9
{1,2,3}	92.1	51.3	65.9
{1,2,3,4}	91.7	51.9	66.3
{1,2,3,4,5}	91.1	52.6	66.7
{1,2,3,4,5,6}	89.5	53.6	67.1
{1,2,3,4,5,6,7}	83.7	74.1	78.6

Sieve Effectiveness

- ACE Newswire

This work (sieve)	**83.8**	73.2	**78.1**
This work (single pass)	82.2	71.5	76.5
Haghighi and Klein (2009) +S	77.0	**75.9**	76.5
Poon and Domingos (2008)	71.3	70.5	70.9
Finkel and Manning (2008) +G	78.7	58.5	67.1

Questions

- Good accuracies on (clean) text. What about…
  - Conversational speech?
    - Ill-formed, disfluent
  - Dialogue?
    - Multiple speakers introduce referents
  - Multimodal communication?
    - How else can entities be evoked?
    - Are all equally salient?

More Questions

- Good accuracies on (clean) (English) text: What about..
  - Other languages?
    - Salience hierarchies the same
      - Other factors
    - Syntactic constraints?
      - E.g. reflexives in Chinese, Korean,..
    - Zero anaphora?
      - How do you resolve a pronoun if you can't find it?

Reference Resolution Algorithms

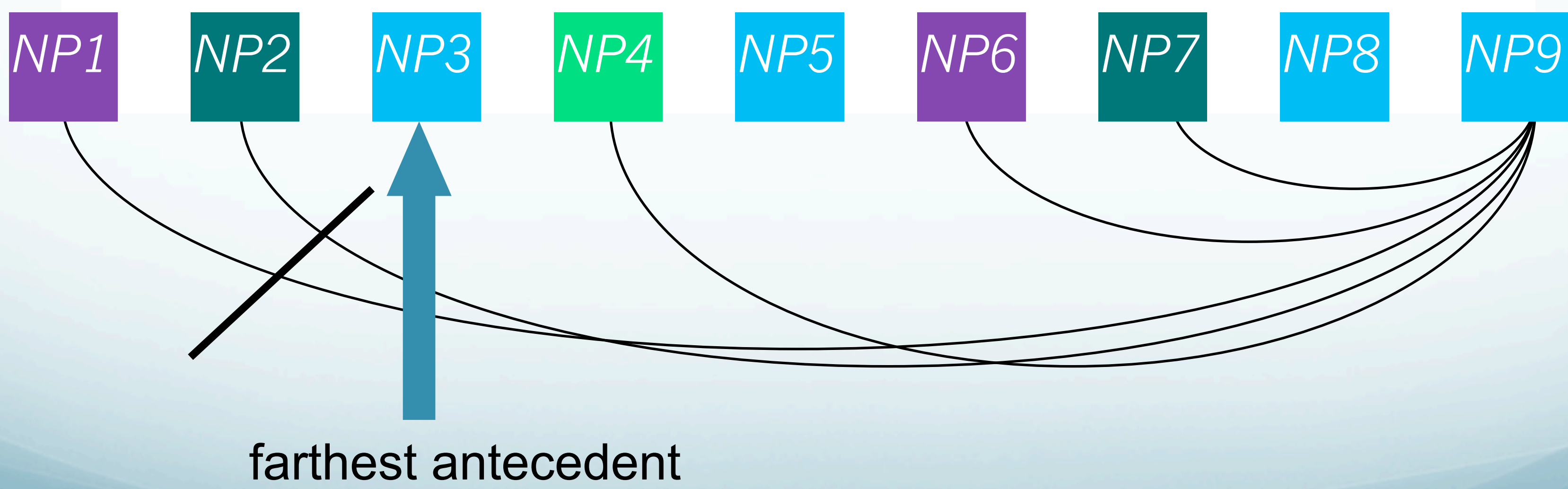
- Many other alternative strategies:
  - Linguistically informed, saliency hierarchy
    - Centering Theory
  - Machine learning approaches:
    - Supervised: Maxent
    - Unsupervised: Clustering
  - Heuristic, high precision:
    - Cogniac

Conclusions

- Co-reference establishes coherence
- Reference resolution depends on coherence
- Variety of approaches:
  - Syntactic constraints, Recency, Frequency,Role
- Similar effectiveness - different requirements
- Co-reference can enable summarization within and across documents (and languages!)

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