Reference Resolution

Adam Meyers
New York University

Computational Linguistics
Lecture 8
2016

Outline

- What is Reference Resolution?
- Linguistic Analysis of Coreference
- Coreference Algorithms: Proper Nouns, Pronouns, Common Nouns
- Evaluation Issues
- Summary

Reference Resolution

- Reference Resolution:
  - Which words/phrases refer to some other word/phrase?
  - How are they related?
- Anaphora vs. Cataphora
  - Anaphora: an ***anaphor*** is a word/phrase that refers back to another phrase: the ***antecedent*** of the anaphor
    - ***Mary*** *thought that* ***she*** *lost* ***her*** *keys.*
  - Cataphora (less common): a ***cataphor*** is a word/phrase that refers forward to another phrase: its **precedent**.
    - ***She*** *was at NYU, when* ***Mary*** *realized that she lost her keys.*
  - ***Anaphora*** is often used as a synonym for ***Reference Resolution*** and the term ***antecedent*** is often used instead of ***precedent***.

Types of Anaphora I

- Coreference: Antecedent = Anaphor
  - Though **Big Blue** won the contract, this official is suspicious of **IBM**.
  - **Mary** could not believe what **she** heard.
- Similar to Coreference
  - Type Coreference (vs. Token)
    - AKA, identify of sense (vs. identify of reference)
    - John ate **a sandwich** and Mary ate **one** also.
  - Bound variable
    - Every **lioness** guards **its** cubs
    - $(\forall \text{lioness}\, L)(\text{L guards L's cubs})$
- Predication and Apposition: some (not all) specs label as coreference
  - ***Mary*** is ***a basketball player***
  - ***Mary***, ***a basketball player from NYU***

Types of Anaphora II

- **Bridging Anaphora:** links between "related" objects
  - ***The amusement park*** *is very dangerous.* ***The gate*** *has sharp edges.* ***The rides*** *have not been inspected for years.*
- Some IE relation instances can be viewed as bridging
  - *When* ***the baby*** *cried,* ***the parents*** *rushed into the room.*
  - ACE Relation: **Per-Social.family***(**the baby**,**the parents**)*
- **"Other" Anaphora:** words including *other* and *another* invoke an "other instance of type" relation
  - **This book** is valuable, but **the other book** is not.
- **Non-NP Anaphora**, e.g., events/propositions
  - ***Mary left the room****.* ***This*** *upset her parents.*
  - *John* ***read the dictionary****. Then Mary did* ***it*** *too.*

2 Models of NP Coreference

- **Chains of Coreference**: Which words/phrases co-refer with which other words/phrases, possibly forming a chain of the form:
  - $Np_n \leftarrow Np_{n-1} \leftarrow \ldots \leftarrow NP_2 \leftarrow NP_1$
  - IBM ← Big Blue ←… ← The company ← they
- **Mentions and Entities** (ACE)**:** Which phrases refer to the same object in the real world?

Entity: International Business Machines

NP_n NP_{n-1} … NP_2 NP_1

IBM Big Blue … The company they

Chain vs. Entity Model

- Entity model
  - Especially suited for fully spelled out names
  - Instances where coreference is based entirely by the discourse context and not limited by proximity
    - Instances that are many lines apart
    - Cross-document coreference
- Chain Model
  - Especially suited to pronouns and definite common nouns that refer back to antecedent NPs
  - Instances in which the anaphor abbreviates, or provides a less specific descriptor than the antecedent
  - Instances of coreference where proximity of anaphor and antecedent is a factor

Coreference with different types of Nouns

- Coreference between Proper Nouns (NEs), including abbreviations, nicknames and substrings
  - Focus of most NLP systems: high precision/recall, links most informative NPs, ...
- Coreference between common noun phrases (CNPs) and preceding NPs (NEs and CNPs)
  - Worst system performance, least studied
- Coreference between pronouns and other NPs
  - Focus of largest body of theoretical work
  - Moderate system performance

Coreference between Proper Nouns (NEs)

- Instances of the same name string in a document usually refers to the same entity
  - ***IBM***, ***IBM***, ***IBM***, ***IBM***, … → ***Entity_IBM***
  - ***George Bush, George Bush, …*** → ***Entity_GB***
- Abbreviations and Nicknames match full name (full name is often first)
  - Abbreviations: mostly rule based (acronyms, subsequences, etc), Nicknames need a lexicon
  - Examples:
    - ***International Business Machines***, ***IBM***, ***Big Blue***… → ***Entity_IBM***
    - ***St. Petersburg*** → ***Saint Petersburg***
    - ***George Bush, George Bush, W, …*** → ***Entity_GB***
    - ***New York Yankees*** ←***New York***, ~~***New York Times***~~ ←***New York*** (place names only match some orgs)
- **S**imple rules work, links most informative NPs, results in high 90s, very little literature
  - Important component of IE systems
- One interesting problem: Name disambiguation
  - Distinguishing multiple individuals with the same name
  - Usually, a problem across documents
    - Exception: *George Bush and his son George W were there*.
  - Abbreviation rules may allow two possible antecedents (*and then George said*)
  - Standardized abbreviations may not be unique,
    - ***AMEX*** → ***American Express*** or ***American Stock Exchange***

Pronouns in English

- **Definite Pronouns**: typically refer to specific NPs
  - 3rd person personal pronouns
    - ***he, him, his, she, her, hers, it, its, they, them, their, theirs***
  - 3rd person Reflexive pronouns
    - ***himself, herself, itself, themselves***
  - ***each other*** – reciprocal pronoun, similar to reflexives
  - *1st and 2nd person pronouns*
    - ***I, me, my, myself, mine, our, ours, ourselves, you, your, yours, yourself***
    - Dialogues between 2 people; or writer/speaker and audience
- **Indefinite Pronouns:**
  - ***one*** – can be used for type coreference
  - Other indefinites – no antecedents in text
    - ***something, someone, everything, everyone, ...***

3rd Def Prons: NonSyntactic Constraints/Preferences

- Usually have an antecedent
- Gender/number/person agreement (language specific)
  - ***Robert* ← *he*, ~~*Robert*~~ ← *she*, ~~*Robert*~~ ← *it*, ~~*Robert*~~ ← *they***
  - ***~~IBM~~* ← *he*, *~~IBM~~* ← *she*, *IBM* ← *it*, *IBM* ← *they***
  - ***~~I~~* ← *she*, *~~me~~* ← *her*, *~~you~~* ← *they***
- Selection Restrictions
  - ***Children*** *have many toys.* ***They*** *love to play.*
  - *Children have* ***many toys***. ***They*** *are always breaking.*
- Pragmatics
  - *Mary yelled at* ***Alice***. ***She*** *interrupted the phone call.*
  - ***Mary*** *yelled at Alice.* ***She*** *can be so mean sometimes.*
- Others: closer antecedents preferred, repeated NPs are more likely to be antecedents, etc. (J&M have several more examples)

Binding Theory Constraints

- An Antecedent of personal pronouns cannot be “too close” to the pronoun.
- An Antecedent of a reflexive/reciprocal pronoun cannot be “too far” from the pronoun.
- Definitions of “too close” and “too far”
  - Vary from language to language
  - Vary among different classes of pronouns/reflexives
  - Are defined using different primitive concepts within different linguistic theories
- Binding Theory Constraints are usually defined in terms of syntactic configurations

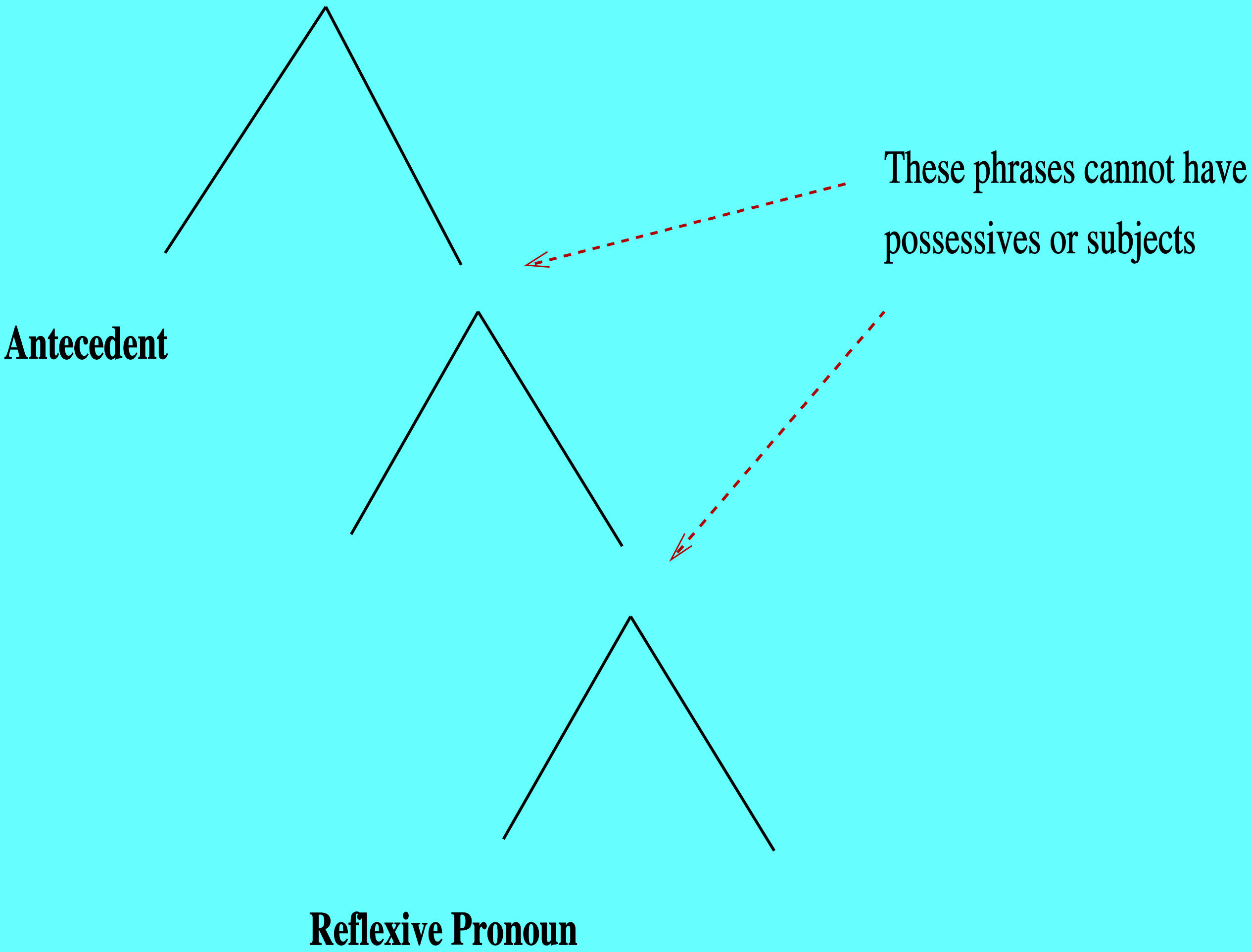
Binding Theory for English 3rd Pers Prons

- Case 1: If the pronoun ***p*** is inside an NP premodified by a possessive, the antecedent needs to be outside of this NP
  - ***John*** *likes Mary's drawing of* ***him***
  - ***John*** *likes* ***his*** *drawing of Mary*
- Case 2: Otherwise, the antecedent must be outside the immediate tensed clause containing the personal pronoun.
  - ***John*** *said that* ***he*** *liked pizza.*
  - ***John*** *wanted for* ***him*** *to like pizza.*
  - ***John*** *liked* ***him****.*
- Theories of binding vary about how these (and similar) constraints are encoded, but the differences in the final result (quality of system output) is minimal. While the above 2 rules cover most cases, there are also some exceptions:
  - ***John*** *always carries a slice of pizza with* ***him****.*

Binding Theory for English Reflexives/Reciprocals

- The antecedent of a reflexive/reciprocal **must be** the closest subject or possessive such that:
  - The antecedent precedes and “commands” the pronoun
    - **A** commands **B** if **A** is the sibling of a phrase that dominates **B**.
  - There is no possessive or subject for phrases in the path in the phrase structure tree between antecedent and pronoun
- Examples:
  - ***Mary*** *saw* ***herself*** vs. ****Mary*** *said that John would meet* ***herself*** ***soon***
  - ***Mary****'s picture of* ***herself*** vs. ****Mary*** *saw John's picture of* ***herself***
- These rules covers most cases.
  - Exception: *Pictures of* ***themselves*** *made* ***the actors*** *nervous.*

Reflexive Pronoun Constraint



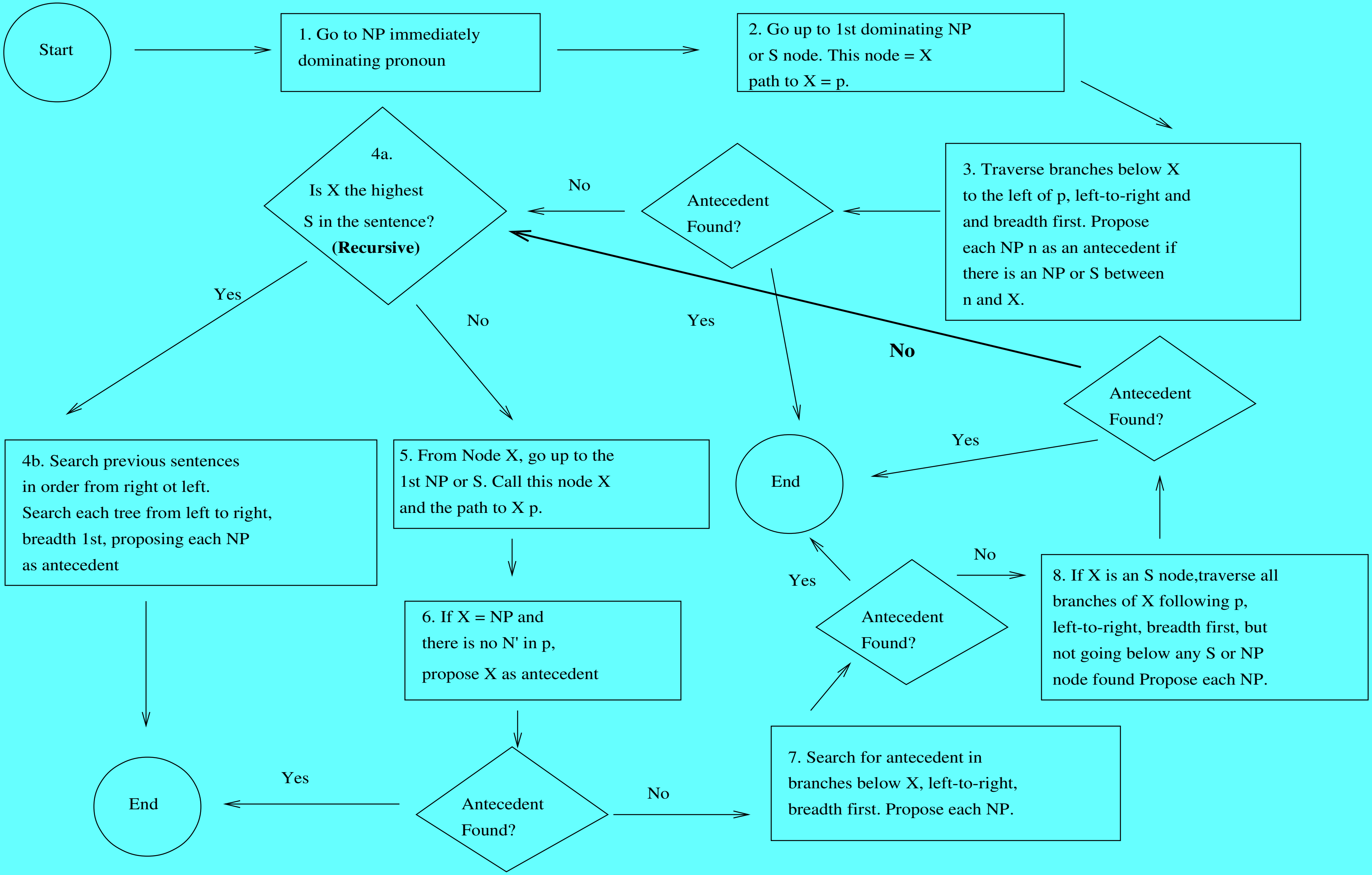
Binding Theory Details Described Above are English Specific

- *zìjǐ* – Chinese reflexive pronoun (example)
  - Ambiguous Example from Choi 1997
    - ***Zhangsan*** *renwei* ***Lisi*** *zhidao* ***Wangwu*** *xihuan* ***zìjǐ***
    - ***Zhangsan*** *thinks* ***Lisi*** *knows that* ***Wangwu*** *likes* ***self***
    - ***Zìjǐ*** can be coreferential with ***Zhangsan, Lisi*** or ***Wangwu***
  - In quasi-translated English, Wangwu would be the antecedent
    - ***Zhangsan*** *thinks* ***Lisi*** *knows that* ***Wangwu*** *likes* ***himself***
- Reflexive/Nonreflexive distinction holds across languages, but constraints on how close/far differ across languages: Icelandic, Chinese, etc.

Pronoun Resolution Methodology

- Hobbs search:
  - a simple system that provides a high baseline
  - Lappin and Leas (1994) report 82% F-score for Hobbs Search
- Sets a High Baseline for Pronoun Coreference
- Higher Scoring Systems Tend to be Much More Complex

Hobbs Search Algorithm to Find Antecedent of Anaphors

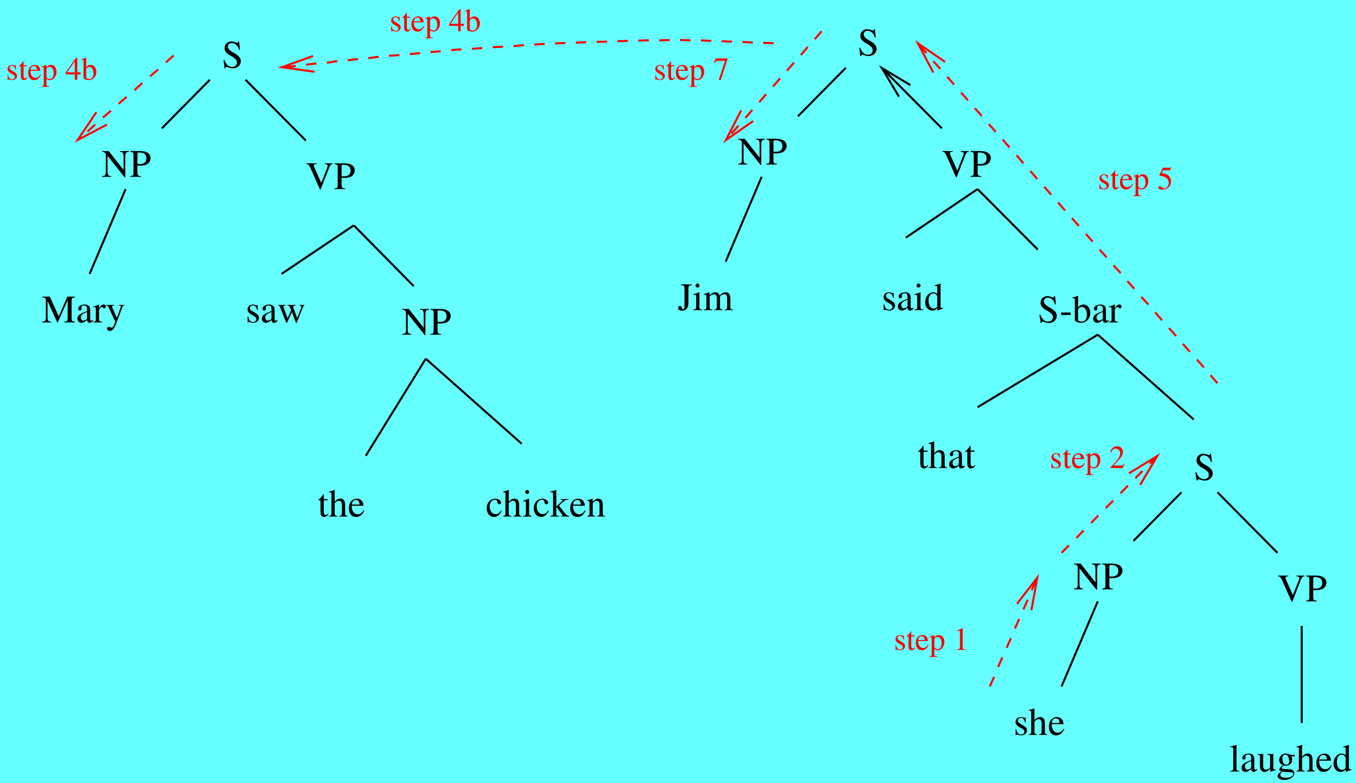


Computational Linguistics
Lecture 8
2016

Hobbs Search Example

1. Mary saw the chicken.

2. Jim said that she laughed.



Testing the Hobbs Algorithm

- Try Hobbs on instances of PRP in wsj_0003 from WSJ Penn Treebank
- How many cases does the Hobbs algorithm get correct?
- How many incorrect?
- Are there some tweaks that would give better results?
- Or would these tweaks hurt other cases?

No-Parse Hobbs-like Search

- Only Consider Nouns/NGs satisfying constraints
- Continue searching until antecedent or loop exits

1. Initialize sentence_counter to 0 and search current sentence from left to right, ending before pronoun.
2. Repeat the following step until an antecedent is found or sentence_counter reaches the maximum (e.g., 3)
   i. Search previous sentence from left to right
   ii. Increment sentence_counter by 1

More Pronoun Coreference Systems

- Lappin and Leass (1994): Hobbs-Search-like procedure, Morphological filter, Binding Theory, Pleonastic Pronoun Handler, preferences based on grammatical role hierarchy (subject > object > ind-object), preference for same grammatical role, frequency of noun, recency, decision procedure for finding pronoun coreference
  - 4% over Hobbs Search
- Other Systems Using Statistical Weights or Machine Learning Score a Little Bit Better, e.g., Dagan et. al. (1995) score another 3% better (89% vs. 82%).

Common Noun Coreference

- Definite Common Nouns
  - Poessio and Veira (2000) baseline:
    - A common noun phrase NP_1 with determiner “the” can be coreferential with a preceding NP_2 if:
      - NP_1 and NP_2 have the same head
      - And (ignoring determiners) NP_1 has a subset of the modifiers of NP_2
- There has been very little improvement on this baseline and very few systems that correctly identify the other cases with any large degree of accuracy
- Other factors:
  - Distance between NP_1 and NP_2
  - Other determiners, modifiers, possessives, etc.

Why is Common Noun Coreference Difficult?

- Only some common noun phrases are anaphoric
  - Definite vs. Generic
    - ***The officers*** vs. ***officers*** vs. ***an officer***
  - Limit to ***the*** phrases is a conservative decision
    - ***this***, ***that***, ***those***, possessives, ... improves recall, lowers precision
- When can a common noun corefer to another noun?
  - Limit to identical nouns is a conservative decision
  - Other choices improve recall, lower precision
    - My experience: a hand-crafted list of matches to NE classes
      - Ex: PERSON matches: man, human, person, individual, woman, .., officer, attorney, ...
      - Hurts approximately as much as it helps (paper wasn't accepted to conference)

Scoring Coreference 1

- Basics: $Precision = \frac{Correct}{\text{System Output}}$ $Recall = \frac{Correct}{\text{Answer Key}}$ $\text{F-Score} = \frac{2}{\frac{1}{precision} + \frac{1}{recall}}$
- Problem: How do you measure number of correct?
- MUC-6:
  - Coreference Chains = Partitions of NPs
  - Recall and Precision are based on mismatches (edit distance) between partitions: numbers of links added and/or subtracted to change incorrect partitions to correct ones
    - Given 7 NPs in a system output chain: A_1, A_2,A_3,A_4,A_5,B_1,B_2 such that:
      - The sets {A_1, A_2,A_3,A_4,A_5 } and {B_1,B_2 } belong to separate chains in the Answer Key
      - The system output contains: 5 correct links and 1 incorrect link
        - » Precision = 5/6 = 83%
      - The system has found all 5 correct links
        - » Recall = 5/5 = 100%
      - F-Score = 91%

Scoring Coreference 2

- B-Cubed (Bagga and Baldwin 1998)
  - Precision calculated for each system chain (and averaged)
    - Given 7 NPs in a system output chain: A_1, A_2,A_3,A_4,A_5,B_1,B_2 such that:
      - The sets {A_1, A_2,A_3,A_4,A_5 } and {B_1,B_2 } belong to separate chains in the Answer Key
      - The precision calculated for each item in chain and averaged:
      - $(5\,X(\frac{5}{7})+2\,X(\frac{2}{7}))*(\frac{1}{7})\approx .59$
  - Recall calculated for each answer key chain (and averaged)
    - $(5\,X(\frac{5}{5})+2\,X(\frac{2}{2}))*(\frac{1}{7})=1$
  - F-score = $\frac{2}{((\frac{1}{.59})+1)}=.74$
  - Difference with MUC Score:
    - penalizes incorrect links more for precision
    - gives credit for NPs that are not coreferential with other NPs
- ACE: complex weighted average designed to count names more than other types of NPs and Person names most of all.

Cross Document Coreference

- So far: coreference with a single discourse
  - within a single document
  - Names usually are unambiguous
  - Disambiguiation strategies for Exceptions, e.g., most recent instance
    - George Bush Sr. vs George W. Bush
      - Bush, George Bush, Mr. Bush, President Bush
    - New York City, New York State
      - New York
- Reference Independent of Invidual Documents
  - Same person name, abbreviation, organization name
  - How do we know when they have different referents?

Baseline Strategy For CrossDoc

- Do single document coreference in each document
- Entity = set of "mentions" that are coreferential
- Select only those Entities which include Name mentions
- Choose longest name string as representative label
  – (don't use abbreviations as label)
- Compare representative labels across documents
  – Merge if labels match exactly
  – Merge if labels match modulo minor modifications
    - Delete middle initial or match middle names
    - Possibly delete titles
  – Similar to name coreference, but more conservative

Hard Cases: Ambiguity and Aliases

- Same name, different middle initial, e.g., *George Bush*
- Ambiguous abbreviations
  - *AMEX: American Stock Exchange* or *American Express*
- People famous in specific domains
  - *Michael Jackson:* Musician, basketball player, football player, executive, …
- Places
  - *New York* (City vs State)
  - *Paris* in (France, Texas, Ontario, Denmark)
- Metonymy
  - *New York*: Rangers, Mets, Yankees, Giants, Jets, …
- Spelling Variation Across Documents (typos, transliteration, etc.)
  - *Osama bin Ladin, Usama ibn Ladin*
  - *(Moammar|Muammar) (Gadaffi|Gaddafi|Gathafi|Kadafi|Kaddafi|Khadafy|Qadhafi|Qathafi)*
- Name Changes over time
  - *Beijing, Beiping*
  - *Leningrad, Saint Petersburg*

Entity Linking Tasks

- TAC KBP Entity Linking Tasks
  - http://nlp.cs.rpi.edu/kbp/2014/
  - http://www.nist.gov/tac/2015/KBP/ColdStart/
  - http://nlp.cs.rpi.edu/kbp/2016/task.html
  - Do within document coreference
  - For each people, organization, GPE entity E, either
    - Link E to an entry in the existing wikipedia-based database OR
    - Link E with a cross-document cluster of entities that your system created
    - Or create a new cross-document entity
- Database created semi-automatically from Wikipedia
  - Database entries correspond to Wikipedia pages
    - Ex: there are several *Paris* pages, one for each "sense" of *Paris*

Strategies Researchers Use

- Machine Learning with lots of features
- Baseline strategies as described
- Contextual features: similar contexts/diff docs
  - Ngrams, relations, vocabulary distribution of whole document
- Extract from Wikipedia Info Boxes
- Other features of documents
  - News articles from the same date and similar location
  - Genre or topic of article

Summary

- Reference Resolution Covers a Wide Area
  - Most Studied Area is Coreference
    - Proper Noun Coreference
      - Easiest to find correct answer
      - Most important for many applications
    - Pronoun Coreference
      - Most thoroughly studied in linguistics
  - Opportunities for research:
    - common noun coreference, other types of reference resolution, connection with relation extraction
- Simple hard-to-beat baselines:
  - Hobbs
  - Poessio and Veira
- Evaluation is Non-Trivial

Readings

- J&M: Chapter 21:3-8, 21:9
- Lappin and Leas (1994)
  - http://www.aclweb.org/anthology/J94-4002
- I also can make available some coref corpora
  - the one used for MUC-6
  - Penn Treebank WSJ corpus with pronoun coref