Coreference Resolution

CS224n
Christopher Manning
(borrows slides from Roger Levy, Altaf Rahman, Vincent Ng, Heeyoung Lee)

What is Coreference Resolution ?

– Identify all noun phrases (**mentions**) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

What is Coreference Resolution ?

– Identify all noun phrases (**mentions**) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

What is Coreference Resolution ?

– Identify all noun phrases (**mentions**) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his

secretary of state on Monday

had foreign affairs experience

What is Coreference Resolution ?

– Identify all noun phrases (**mentions**) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chose her because she had foreign affairs experience as a former First Lady.

What is Coreference Resolution ?

– Identify all noun phrases (**mentions**) that refer to the same real world entity

Barack Obama nominated Hillary Rodham Clinton as his secretary of state on Monday. He chos
had foreign affairs experience as a forr

A couple of years later, Vanaja met Akhila at the local park. Akhila's son Prajwal was just two months younger than her son Akash, and they went to the same school. For the pre-school play, Prajwal was chosen for the lead role of the naughty child Lord Krishna. Akash was to be a tree. She resigned herself to make Akash the best tree that anybody had ever seen. She bought him a brown T-shirt and brown trousers to represent the tree trunk. Then she made a large cardboard cutout of a tree's foliage, with a circular opening in the middle for Akash's face. She attached red balls to it to represent fruits. It truly was the nicest tree.

From The Star by Shruthi Rao, with some shortening.

Reference Resolution

- Noun phrases refer to entities in the world, many pairs of noun phrases co-refer, some nested inside others

John Smith, CFO of Prime Corp. since 1986,

saw his pay jump 20% to $1.3 million

as the 57-year-old also became

the financial services co.'s president.

Kinds of Reference

- Referring expressions
  - *John Smith*
  - *President Smith*
  - *the president*
  - *the company's new executive*

More common in newswire, generally harder in practice

- Free variables
  - Smith saw *his pay* increase
- Bound variables
  - The dancer hurt *herself.*

More interesting grammatical constraints, more linguistic theory, easier in practice

"anaphora resolution"

Not all NPs are referring!

- *Every dancer* twisted *her knee.*
- (*No dancer* twisted *her knee.*)

- There are three NPs in each of these sentences; because the first one is non-referential, the other two aren't either.

Coreference, anaphors, cataphors

- Coreference is when two mentions refer to the same entity in the world
- The relation of anaphora is when a term (anaphor) refers to another term (antecedent) and the interpretation of the anaphor is in some way determined by the interpretation of the antecedent

    … and traditionally the antecedent came first

Cataphora

"From the corner of the divan of Persian saddle-bags on which he was lying, smoking, as was his custom, innumerable cigarettes, Lord Henry Wotton could just catch the gleam of the honey-sweet and honey-coloured blossoms of a laburnum…"

(Oscar Wilde – The Picture of Dorian Gray)

Anaphora vs. coreference

- Not all anaphoric relations are coreferential

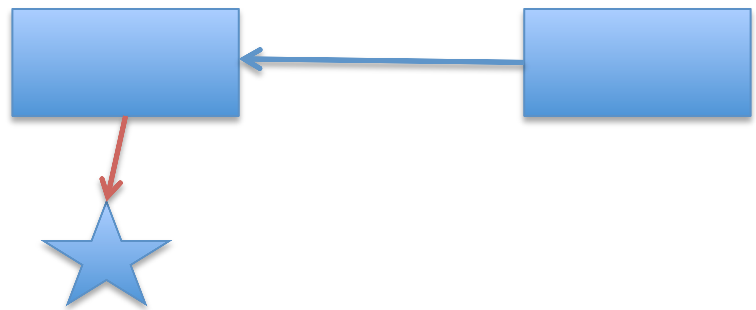
We went to see a concert last night. The tickets were really expensive.

- This is referred to as bridging anaphora.

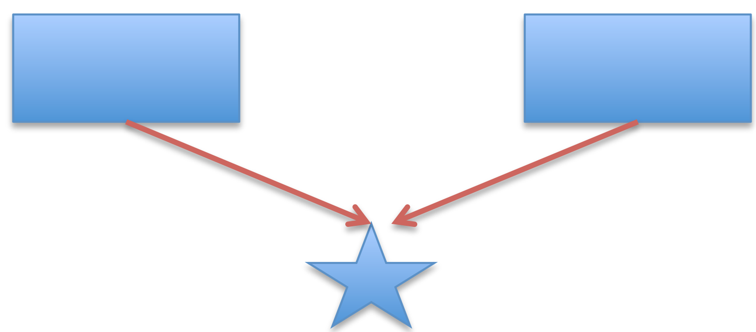
- Conversely, multiple identical full NP references are typically coreferential but not anaphoric.

Two different things…

- Anaphora
  - Text
  - World



- (Co)Reference
  - Text
  - World



Two different things…

- Something you might like to think about:
  - Do various models treat these two cases the same or differently?

  - Should we do more to treat them more differently?

Applications

- Full text understanding:
  - understanding an extended discourse
- Machine translation (if languages have different features of gender, number, etc.)
- Text summarization, including things like web snippets
- Tasks like information extraction and question answering, when some sentences have pronouns
  - *He married Claudia Ross in 1971.*

Traditional pronominal anaphora resolution: Hobbs' naive algorithm

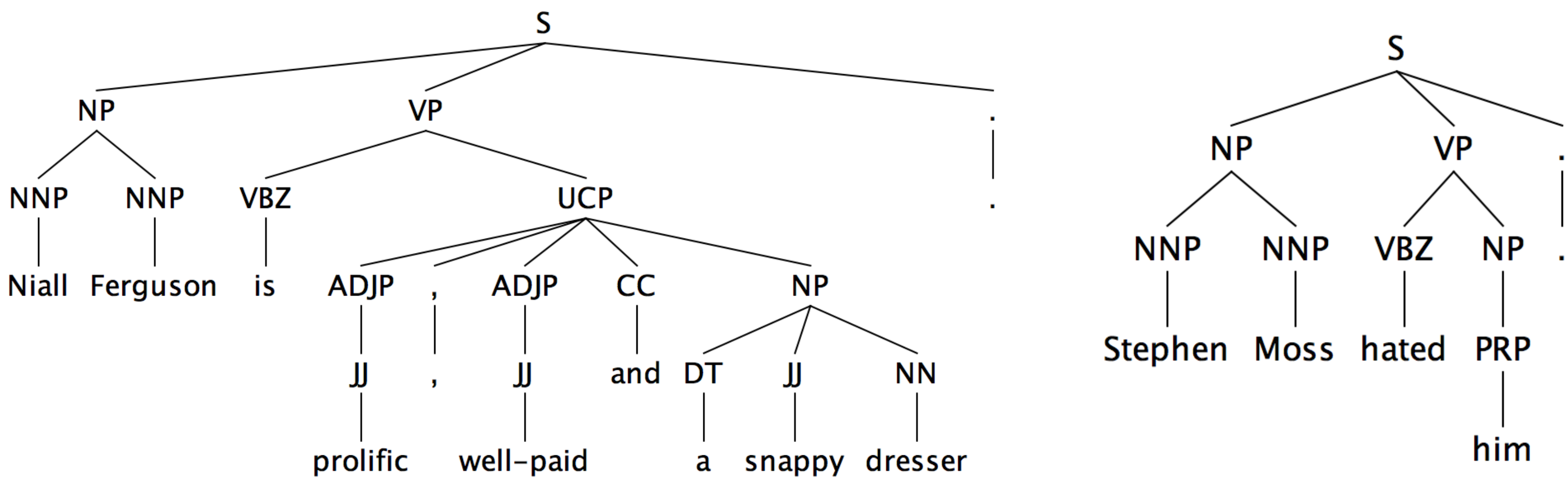
1. Begin at the NP immediately dominating the pronoun
2. Go up tree to first NP or S. Call this X, and the path p.
3. Traverse all branches below X to the left of p, left-to-right, breadth-first. Propose as antecedent any NP that has a NP or S between it and X
4. If X is the highest S in the sentence, traverse the parse trees of the previous sentences in the order of recency. Traverse each tree left-to-right, breadth first. When an NP is encountered, propose as antecedent. If X not the highest node, go to step 5.

Hobbs' naive algorithm

5. From node X, go up the tree to the first NP or S. Call it X, and the path p.
6. If X is an NP and the path p to X came from a non-head phrase of X (a specifier or adjunct, such as a possessive, PP, apposition, or relative clause), propose X as antecedent
   (The original said "did not pass through the N' that X immediately dominates", but the Penn Treebank grammar lacks N' nodes....)
7. Traverse all branches below X to the left of the path, in a left-to-right, breadth first manner. Propose any NP encountered as the antecedent
8. If X is an S node, traverse all branches of X to the right of the path but do not go below any NP or S encountered. Propose any NP as the antecedent.
9. Go to step 4

Actually still often used as a feature in ML systems!

Hobbs Algorithm Example



Supervised Machine Learning Pronominal Anaphora Resolution

- Given a pronoun and an entity mentioned earlier, classify whether the pronoun refers to that entity or not given the surrounding context (yes/no)

? ? ?

Mr. Obama visited the city. The president talked about Milwaukee 's economy. He mentioned new jobs.

- Usually first filter out pleonastic pronouns like "It is raining." (perhaps using hand-written rules)
- Use any classifier, obtain positive examples from training data, generate negative examples by pairing each pronoun with other (incorrect) entities
- This is naturally thought of as a binary classification (or ranking) task

Features for Pronominal Anaphora Resolution

- Constraints:
  - Number agreement
    - Singular pronouns (it/he/she/his/her/him) refer to singular entities and plural pronouns (we/they/us/them) refer to plural entities
  - Person agreement
    - He/she/they etc. must refer to a third person entity
  - Gender agreement
    - He → John; she → Mary; it → car
    - Jack gave Mary a gift. She was excited.
  - Certain syntactic constraints
    - John bought himself a new car. [himself → John]
    - John bought him a new car. [him can not be John]

Features for Pronominal Anaphora Resolution

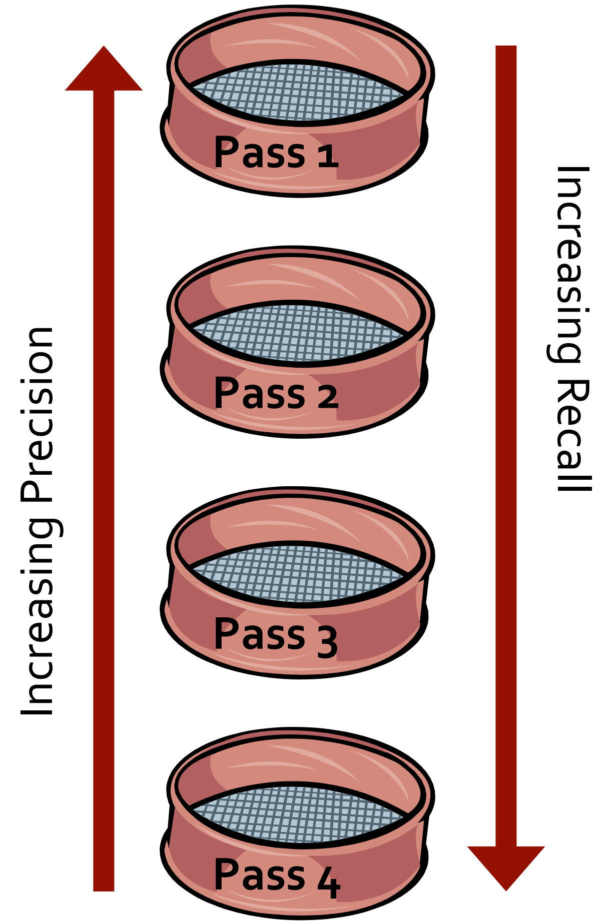
- Preferences:
  - Recency: More recently mentioned entities are more likely to be referred to
    - John went to a movie. Jack went as well. He was not busy.
  - Grammatical Role: Entities in the subject position is more likely to be referred to than entities in the object position
    - John went to a movie with Jack. He was not busy.
  - Parallelism:
    - John went with Jack to a movie. Joe went with him to a bar.

Features for Pronominal Anaphora Resolution

- Preferences:
  - Verb Semantics: Certain verbs seem to bias whether the subsequent pronouns should be referring to their subjects or objects
    - John telephoned Bill. He lost the laptop.
    - John criticized Bill. He lost the laptop.
  - Selectional Restrictions: Restrictions because of semantics
    - John parked his car in the garage after driving it around for hours.
- Encode all these and maybe more as features

Lee et al. (2010): Stanford deterministic coreference

- Cautious and incremental approach
- Multiple passes over text
- Precision of each pass is lesser than preceding ones
- Recall keeps increasing with each pass
- Decisions once made cannot be modified by later passes
- Rule-based ("unsupervised")



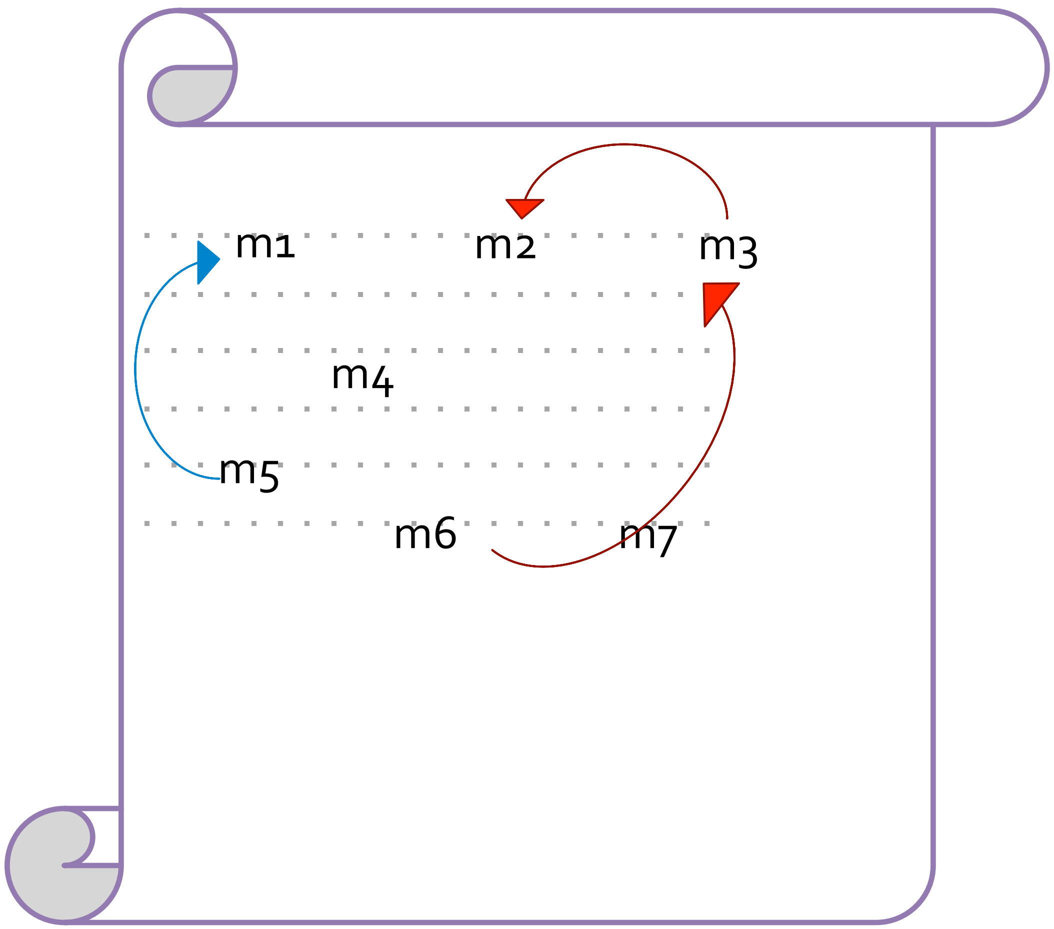
Approach: start with high precision clumpings

E.g.

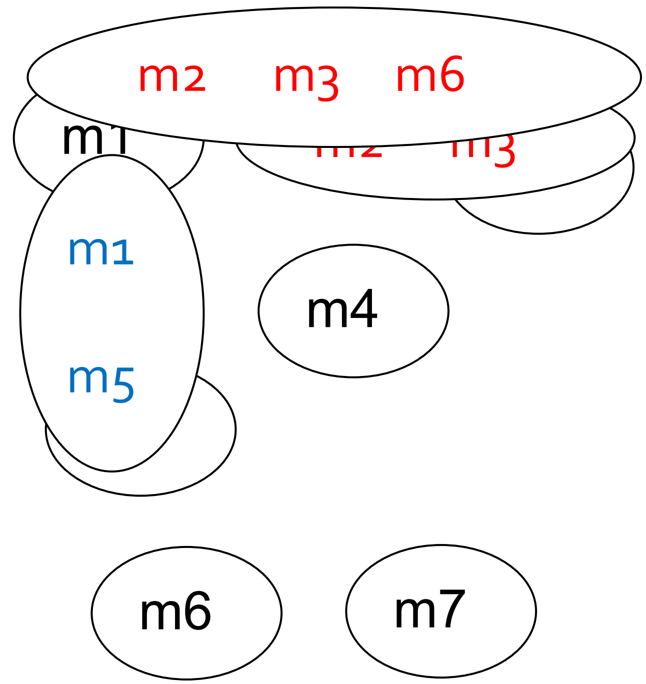
Pepsi hopes to take ***Quaker oats*** *to a whole new level. ... Pepsi says it expects to double Quaker's snack food growth rate. ... the deal gives Pepsi access to* ***Quaker oats'*** *Gatorade sport drink as well as ....*

Exact String Match: A high precision feature

Entity-mention model: Clusters instead of mentions



Clusters:



Detailed Architecture

The system consists of seven passes (or sieves):

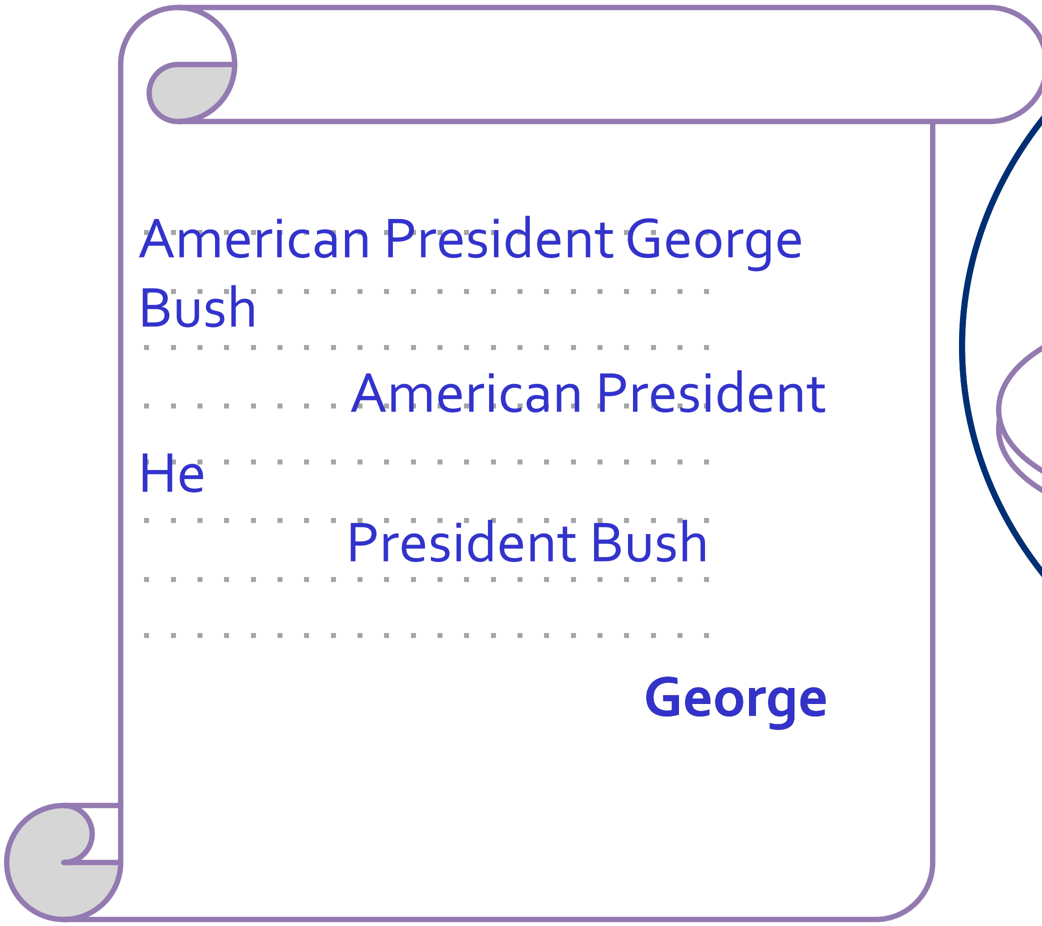
- Exact Match
- Precise Constructs (appositives, predicate nominatives, …)
- Strict Head Matching
- Strict Head Matching – Variant 1
- Strict Head Matching – Variant 2
- Relaxed Head Matching
- Pronouns

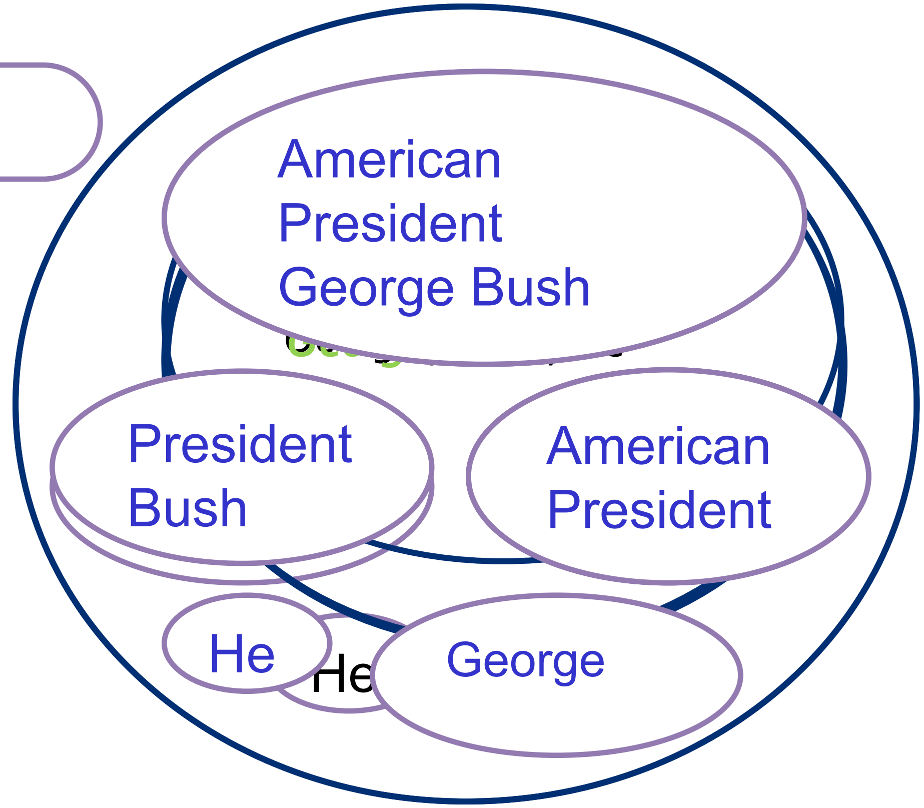
Passes 3 – 5: Examples

- **Pass 3**
  - Yes: "*the Florida Supreme Court*", "*the Florida court*"
  - No: "*researchers*", "*two Chinese researchers*"
- **Pass 4** `(-Compatible Modifiers)`
  - Yes: "*President Clinton*", *{American President, American President Bill Clinton, Clinton}*
- **Pass 5** `(-Word Inclusion)`
  - **Yes**: "*The Gridiron Club at the Greenbrier Hotel*", *{an organization of 60 Washington journalists, The Gridiron Club}*

Pass 6: Relaxed Head Matching

Relaxed Cluster Match



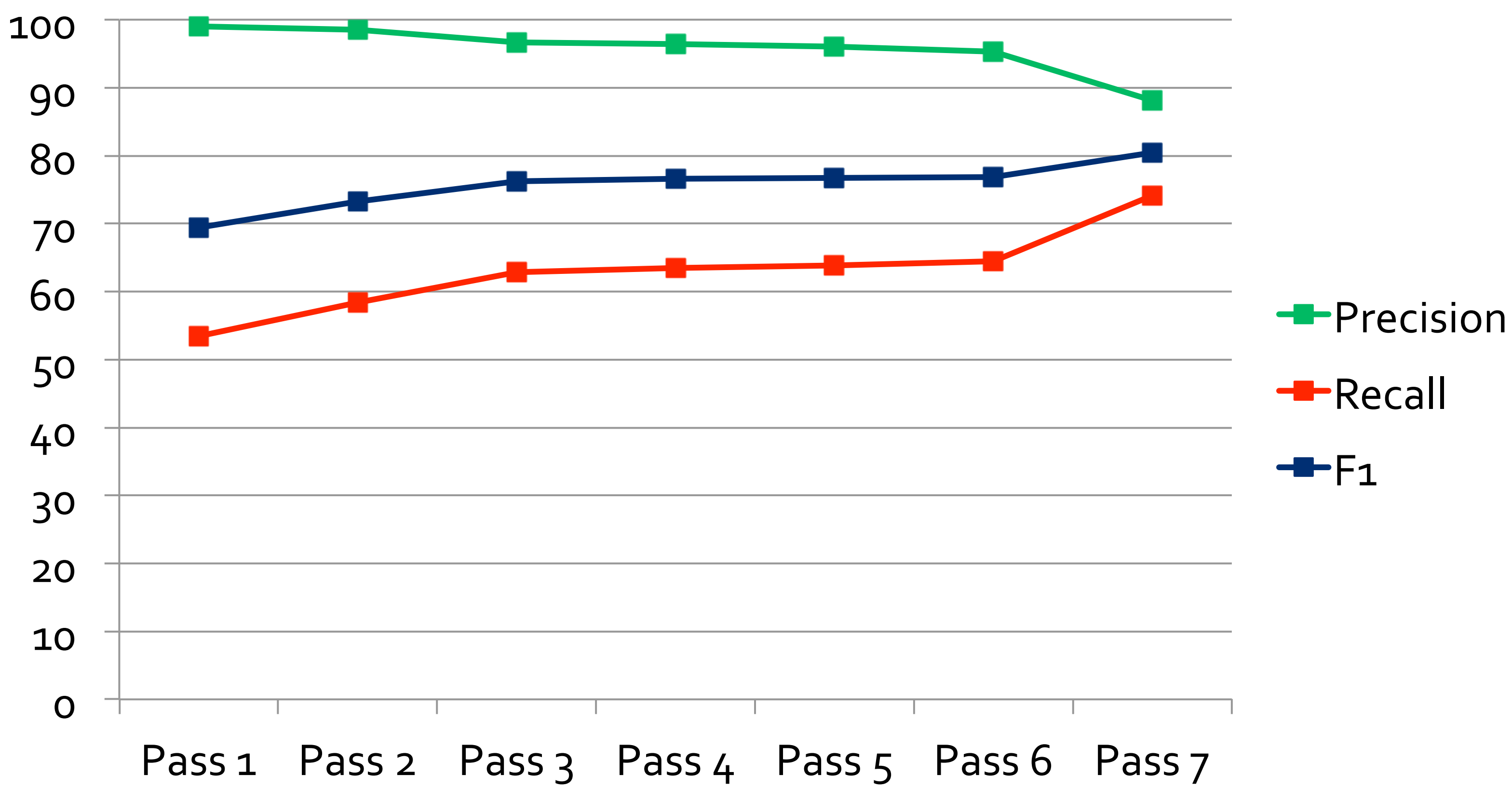


Both mentions have to be named entities of the same type

Pass 7 – Pronoun Resolution

- Attributes agree
  - Number
  - Gender
  - Person
  - Animacy
- Assigned using POS tags, NER labels, static list of assignments for pronouns
- Improved further using Gender and Animacy dictionaries of Bergsma and Lin (2006), and Ji and Lin (2009)

Cumulative performance of passes♪



Graph showing the system's B^3 Precision, Recall and F1 on ACE2004-DEV after each additional pass

CoNLL 2011 Shared task on coref

Official; Closed track; Predicted mentions

System	MD	MUC	B-CUBED	CEAF_m	CEAF_e	BLANC	Official
	F	F^1	F^2	F	F^3	F	$\frac{F^1+F^2+F^3}{3}$
lee	**70.70**	59.57	68.31	**56.37**	**45.48**	73.02	**57.79**
sapena	43.20	59.55	67.09	53.51	41.32	71.10	55.99
chang	64.28	57.15	**68.79**	54.40	41.94	**73.71**	55.96
nugues	68.96	58.61	65.46	51.45	39.52	71.11	54.53
santos	65.45	56.65	65.66	49.54	37.91	69.46	53.41
song	67.26	**59.95**	63.23	46.29	35.96	61.47	53.05
stoyanov	67.78	58.43	61.44	46.08	35.28	60.28	51.92
sobha	64.23	50.48	64.00	49.48	41.23	63.28	51.90
kobdani	61.03	53.49	65.25	42.70	33.79	62.61	51.04
zhou	62.31	48.96	64.07	47.53	39.74	64.72	50.92
charton	64.30	52.45	62.10	46.22	36.54	64.20	50.36
yang	63.93	52.31	62.32	46.55	35.33	64.63	49.99
hao	64.30	54.47	61.01	45.07	32.67	65.35	49.38
xinxin	61.92	46.62	61.93	44.75	36.23	64.27	48.46
zhang	61.13	47.28	61.14	44.46	35.19	65.21	48.07
kummerfeld	62.72	42.70	60.29	45.35	38.32	59.91	47.10
zhekova	48.29	24.08	61.46	40.43	35.75	53.77	40.43
irwin	26.67	19.98	50.46	31.68	25.21	51.12	31.28

Remarks

- This simple deterministic approach gives state of the art performance!
- Easy insertion of new features or models
- The idea of “easy first” model has also had some popularity in other (ML-based) NLP systems
  - Easy first POS tagging and parsing

- It’s a flexible architecture, not an argument that ML is wrong
  - Pronoun resolution pass would be easiest place to reinsert an ML model??

Machine learning models of coref

- Start with supervised data
    - positive examples that corefer
    - negative examples that don't corefer
  - Note that it's very skewed
    - The vast majority of mention pairs *don't* corefer
- Usually learn some sort of discriminative model of phrases/ clusters coreferring
  - Predict 1 for coreference, 0 for not coreferent
- But there is also work that builds clusters of coreferring expressions
  - E.g., generative models of clusters in (Haghighi & Klein 2007)

Evaluation

- B^3 (B-CUBED) algorithm for evaluation
  - Precision & recall for *entities in a reference chain*
  - Precision: % of elements in a hypothesized reference chain that are in the true reference chain
  - Recall: % of elements in a true reference chain that are in the hypothesized reference chain
  - Overall precision & recall are the (weighted) average of per-chain precision & recall
  - Optimizing chain-chain pairings is a hard problem
    - In the computational NP-hard sense
  - Greedy matching is done in practice for evaluation

Evaluation

- B-CUBED algorithm for evaluation

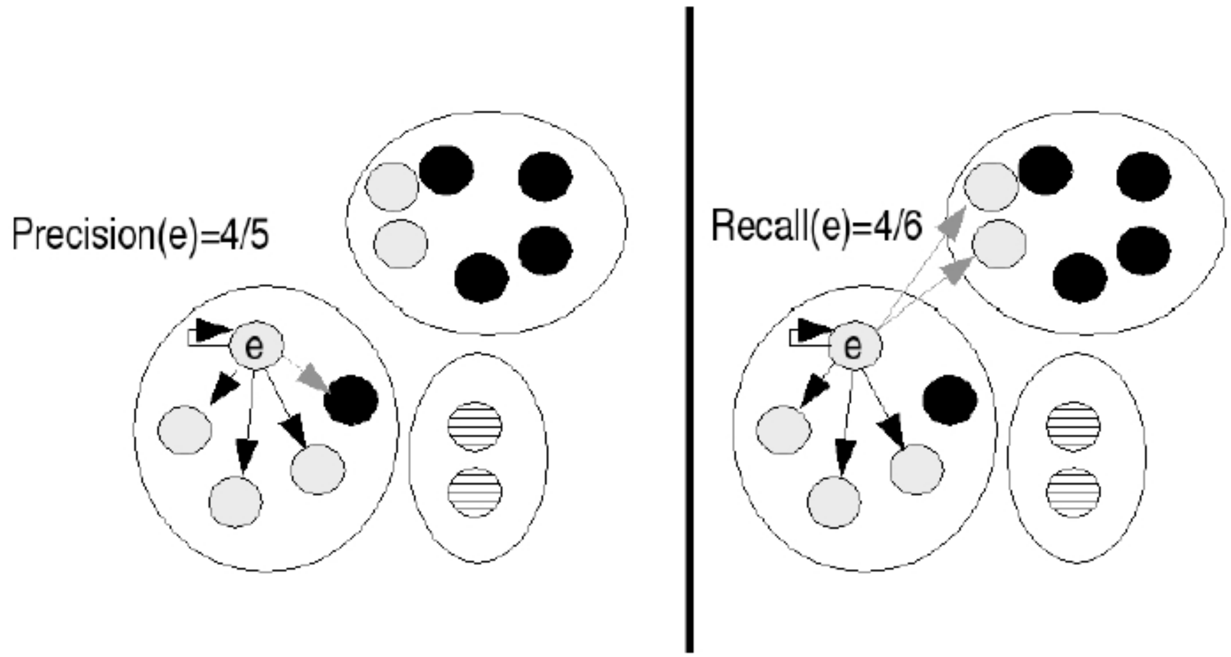


Figure from Amigo et al 2009

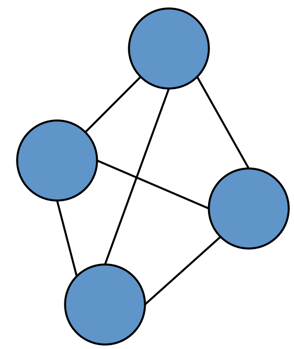
Evaluation metrics

- MUC Score (Vilain et al., 1995)
  - Link based: Counts the number of common links and computes f-measure
- CEAF (Luo 2005); entity based
- BLANC (Recasens and Hovy 2011) Cluster RAND-index
- …

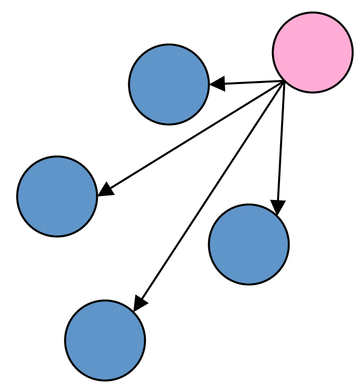
- All of them are sort of evaluating getting coreference links/ clusters right and wrong, but the differences can be important
  - Look at it in PA3

Kinds of Models

- Mention Pair models
  - Treat coreference chains as a collection of pairwise links
  - Make independent pairwise decisions and reconcile them in some way (e.g. clustering or greedy partitioning)



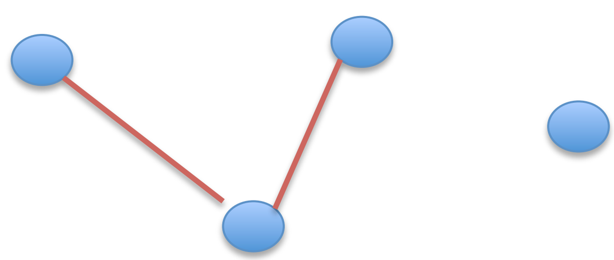
- Mention ranking models
  - Explicitly rank all candidate antecedents for a mention



- Entity-Mention models
  - A cleaner, but less studied, approach
  - Posit single underlying entities
  - Each mention links to a discourse entity [Pasula et al. 03], [Luo et al. 04]

Mention Pair Models

- Most common machine learning approach
- Build a classifier over pairs of NPs
  - For each NP, pick a preceding NP or NEW
  - Or, for each NP, choose link or no-link
- Clean up non-transitivity with clustering or graph partitioning algorithms
  - E.g.: [Soon et al. 01], [Ng and Cardie 02]
  - Some work has done the classification and clustering jointly [McCallum and Wellner 03]
- Failures are mostly because of insufficient knowledge or features for hard common noun cases



Features: Grammatical Constraints

- Apposition
  - Nefertiti, Amenomfis the IVth's wife, was born in …
- Predicatives/equatives
  - Sue is the best student in the class
  - It's questionable whether predicatives cases should be counted, but they generally are.

Features: Soft Discourse Constraints

- Recency
- Salience
- Focus
- Centering Theory [Grosz et al. 86]
- Coherence Relations

Other coreference features

- Additional features to incorporate aliases, variations in names etc., e.g. Mr. Obama, Barack Obama; Megabucks, Megabucks Inc.

- Semantic Compatibility
  - Smith had bought *a used car* that morning.
    - *The dealership* assured him it was in good condition.
    - *The machine* needed a little love, but the engine was in good condition.

But it’s complicated … so weight features

- Common nouns can differ in number but be coreferent:
  - a patrol … the soldiers
- Common nouns can refer to proper nouns
  - George Bush … the leader of the free world
- Split antecedence
  - John waited for Sasha. And then they went out.

Pairwise Features

1. **strict gender [true or false]**. True if there is a strict match in gender (e.g. male pronoun Pro_i with male antecedent NP_j).
2. **compatible gender [true or false]**. True if Pro_i and NP_j are merely compatible (e.g. male pronoun Pro_i with antecedent NP_j of unknown gender).
3. **strict number [true or false]** True if there is a strict match in number (e.g. singular pronoun with singular antecedent)
4. **compatible number [true or false]**. True if Pro_i and NP_j are merely compatible (e.g. singular pronoun Pro_i with antecedent NP_j of unknown number).
5. **sentence distance [0, 1, 2, 3,...]**. The number of sentences between pronoun and potential antecedent.
6. **Hobbs distance [0, 1, 2, 3,...]**. The number of noun groups that the Hobbs algorithm has to skip, starting backwards from the pronoun Pro_i, before the potential antecedent NP_j is found.
7. **grammatical role [subject, object, PP]**. Whether the potential antecedent is a syntactic subject, direct object, or is embedded in a PP.
8. **linguistic form [proper, definite, indefinite, pronoun]**. Whether the potential antecedent NP_j is a proper name, definite description, indefinite NP, or a pronoun.

Pairwise Features

Category	Features	Remark
Lexical	exact_strm	1 if two mentions have the same spelling; 0 otherwise
	left_subsm	1 if one mention is a left substring of the other; 0 otherwise
	right_subsm	1 if one mention is a right substring of the other; 0 otherwise
	acronym	1 if one mention is an acronym of the other; 0 otherwise
	edit_dist	quantized editing distance between two mention strings
	spell	pair of actual mention strings
	ncd	number of different capitalized words in two mentions
Distance	token_dist	how many tokens two mentions are apart (quantized)
	sent_dist	how many sentences two mentions are apart (quantized)
	gap_dist	how many mentions in between the two mentions in question (quantized)
Syntax	POS_pair	POS-pair of two mention heads
	apposition	1 if two mentions are appositive; 0 otherwise
Count	count	pair of (quantized) numbers, each counting how many times a mention string is seen
Pronoun	gender	pair of attributes of {female, male, neutral, unknown }
	number	pair of attributes of {singular, plural, unknown}
	possessive	1 if a pronoun is possessive; 0 otherwise
	reflexive	1 if a pronoun is reflexive; 0 otherwise

[Luo et al. 04]

Mention-Pair (MP) Model

- Soon et al. 2001 ; Ng and Cardie 2002
- Classifies whether **two mentions** are coreferent or not.
- Weaknesses
  - Insufficient information to make an informed coreference decision.

Mention-Pair (MP) Model

- Soon et al. 2001 ; Ng and Cardie 2002
- Classifies whether **two mentions** are coreferent or not.
- Weaknesses
  - Insufficient information to make an informed coreferenced decision.

Barack Obama ………………Hillary Rodham Clinton …….his

……….. **secretary of state** ……………………..He ….……..her

Mention-Pair (MP) Model

- Soon et al. 2001 ; Ng and Cardie 2002
- Classifies whether **two mentions** are coreferent or not.
- Weaknesses
  - Insufficient information to make an informed coreference decision.
  - Each candidate antecedent is considered independently of the others.

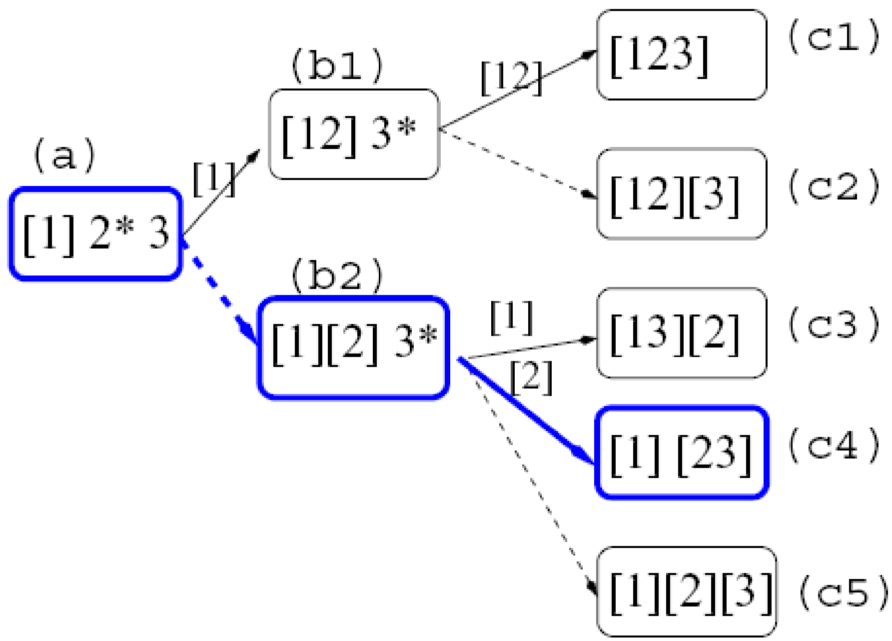
Mention-Pair (MP) Model

- Soon et al. 2001 ; Ng and Cardie 2002
- Classifies whether **two mentions** are coreferent or not.
- Weaknesses
  - Insufficient information to make an informed coreferenced decision.
  - Each candidate antecedent is considered independently of the others.

Barack Obama ………Hillary Rodham Clinton …….his ………..
secretary of state ………..the President……..He …………**her**

An Entity Mention Model

- Example: [Luo et al. 04]
- Bell Tree (link vs. start decision list)
- Entity centroids, or not?
  - Not for [Luo et al. 04], see [Pasula et al. 03]
  - Some features work on nearest mention (e.g. recency and distance)
  - Others work on "canonical" mention (e.g. spelling match)
  - Lots of pruning, model highly approximate
  - (Actually ends up being like a greedy-link system in the end)



Entity-Mention (EM) Model

- Pasula et al. 2003 ; Luo et al. 2004 ; Yang et al. 2004
- Classifies whether **a mention** and **a preceding, possibly partially formed cluster** are coreferent or not.
- Strength
  - Improved expressiveness.
    - Allows the computation of cluster level features
- Weakness
  - Each candidate cluster is considered independently of the others.

Barack Obama ………………Hillary Rodham Clinton …….his

……….. secretary of state ……………………..He …………her

Mention-Ranking (MR) Model

- Denis & Baldridge 2007, 2008
- Imposes a **ranking** on a set of candidate antecedents

- Strength
  - Considers all the candidate antecedents simultaneously
- Weakness
  - Insufficient information to make an informed coreference decision.

Barack Obama ………………Hillary Rodham Clinton …….his
……….. secretary of state …………………….He ………….her