Discourse, Pragmatics, Coreference Resolution

Many slides are adapted from Roger Levy, Chris Manning, Vicent Ng, Heeyoung Lee, Altaf Rahman

A pragmatic issue

- Just how are pronouns interpreted (resolved) in a discourse?

  (1) Jane likes Mary.
  (2) She often brings her flowers.
  (3) She chats with the young woman for ages

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"

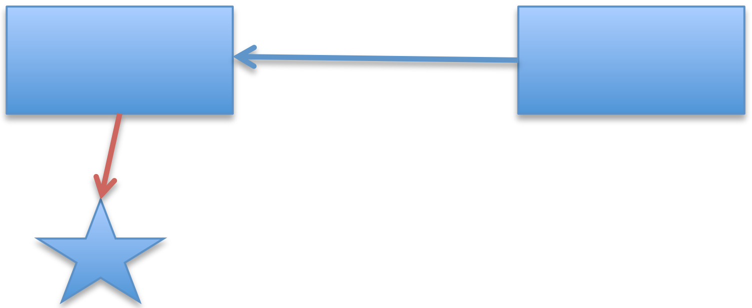
Not all NPs are referring!

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

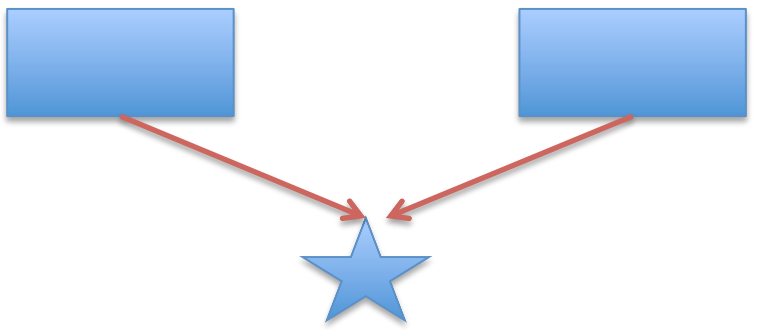
- There are three NPs in each of these sentences; because the first one is non-referential, the other two aren't either.

Two different things…

- Anaphora
  - Text
  - World



- (Co)Reference
  - Text
  - World



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

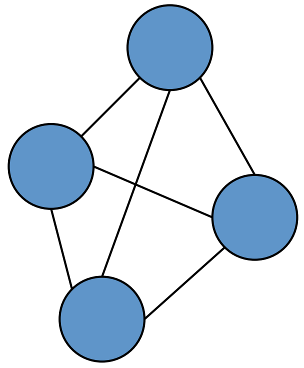
- Preferences:
  - Verb Semantics: Certain verbs seem to bias whether the subsequent pronouns should be referring to their subjects or objects
    - John telephoned Bill. He lost the laptop.
    - John criticized Bill. He lost the laptop.
  - Selectional Restrictions: Restrictions because of semantics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

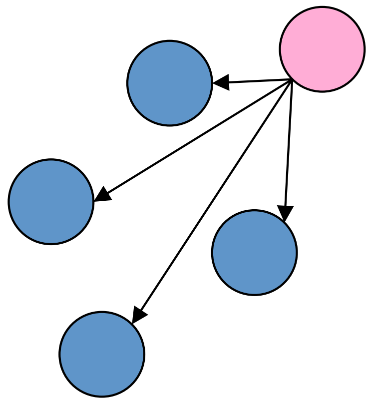
Machine learning models of coref

- 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

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]





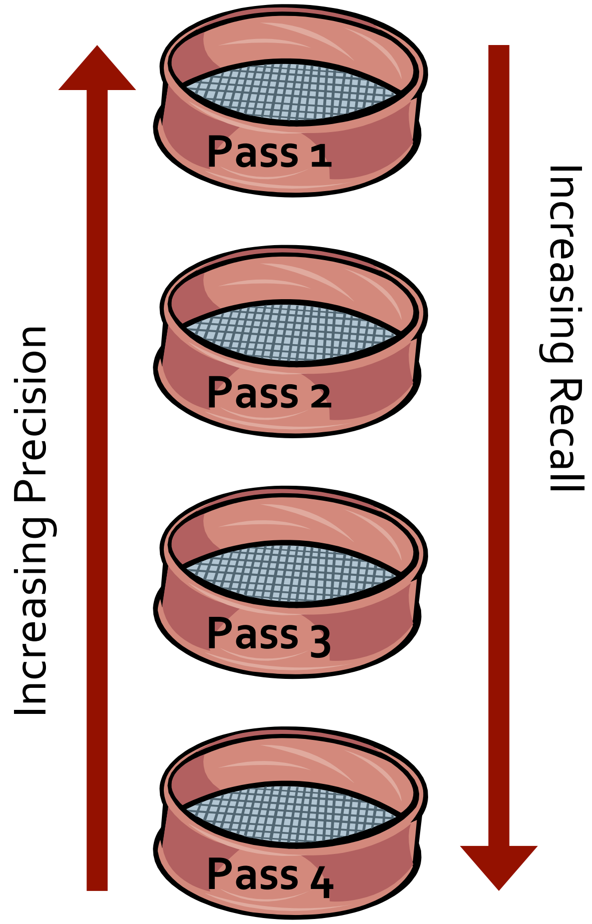
Pairwise Features

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]

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")



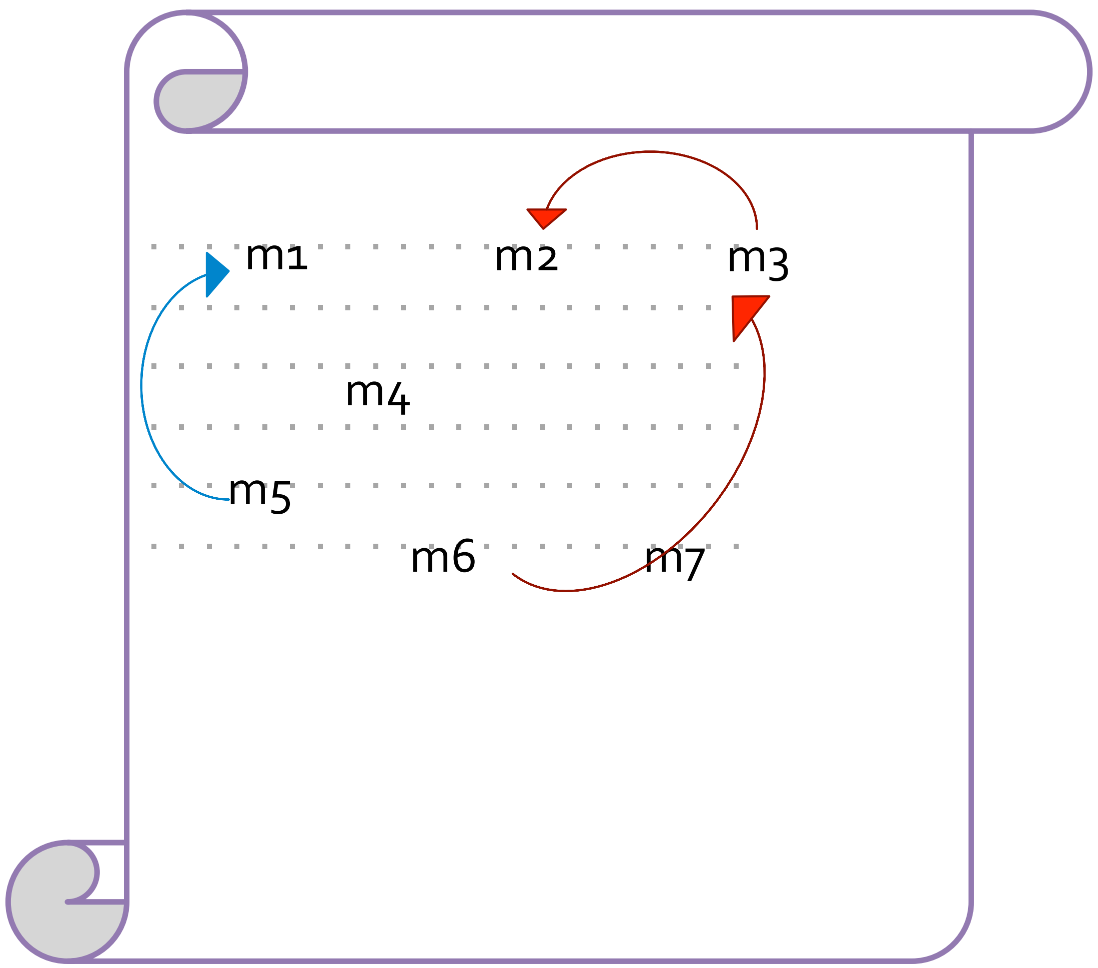
Approach: start with high precision clumpings

E.g.

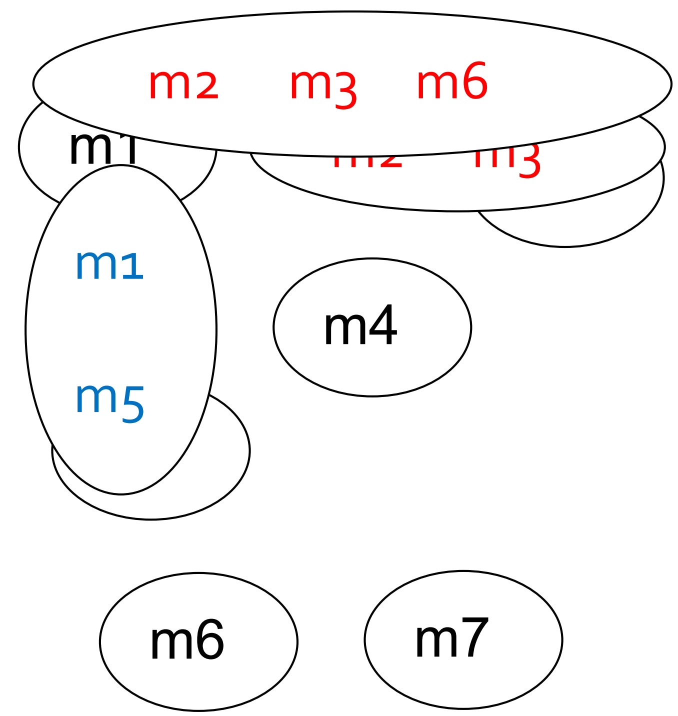
Pepsi hopes to take Quaker oats to a whole new level. ... Pepsi says it expects to double Quaker's snack food growth rate. ... the deal gives Pepsi access to Quaker oats' Gatorade sport drink as well as ....

Exact String Match: A high precision feature

Entity-mention model: Clusters instead of mentions



Clusters:

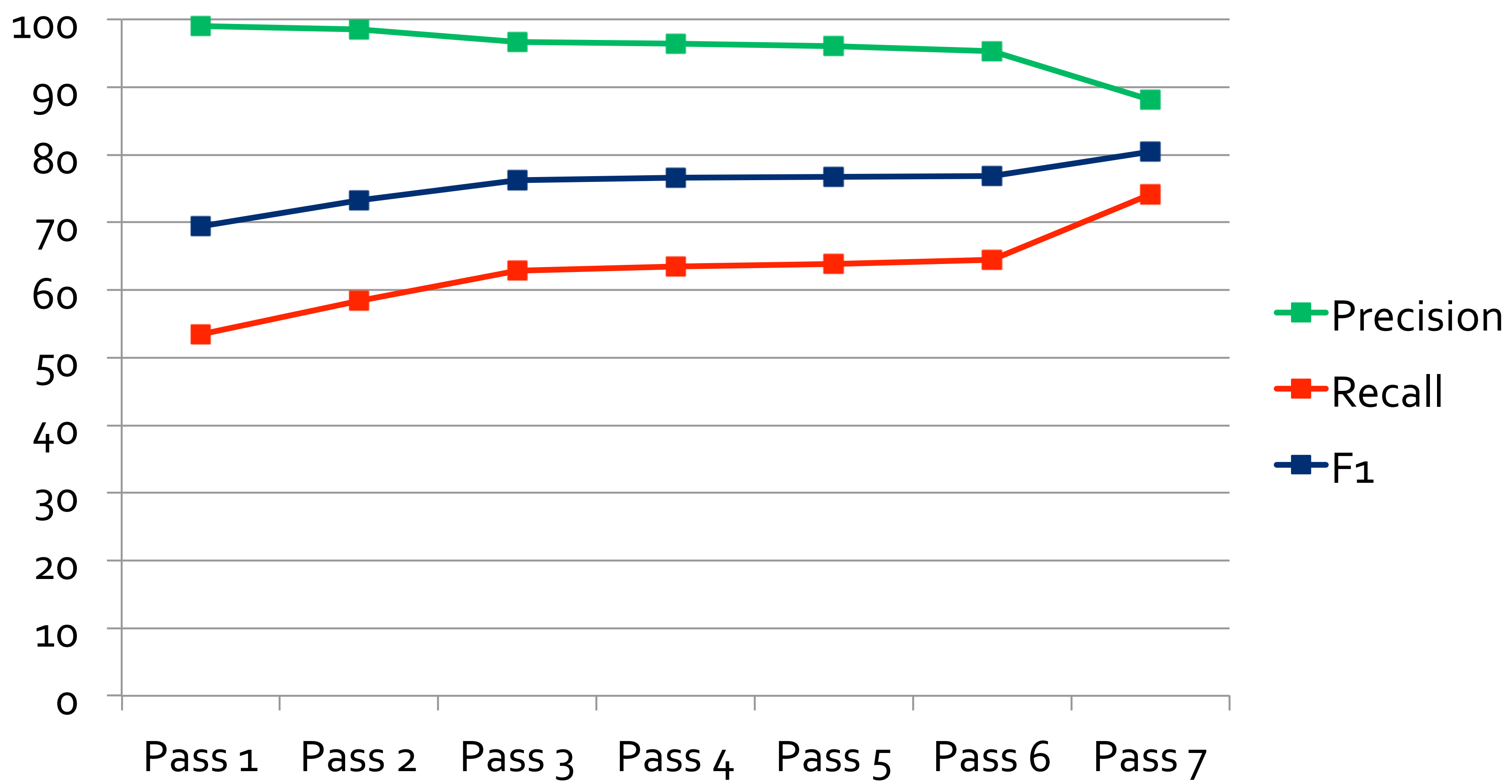


Detailed Architecture

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

Cumulative performance of passes♪



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

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 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

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??