CSE 517: Winter 2013
Discourse and Co-reference

Luke Zettlemoyer

Based on slides by Julia Hockenmaier, Frank Keller, Bonnie Webber, and others

Topics

- Mini-overview of Discourse
  - Co-reference
  - Rhetorical Structure
  - Entity Structure
- Approaches to Co-reference Resolution
  - Clustering
  - Supervised Learning
  - Seives

Discourse

The structure of text that goes beyond the sentence level

We can say that a text is *coherent* if it has well formed structure:

- coreference: the linguistic expressions refer correctly to real-world entities;
- rhetorical structure: the utterances in the discourse have to be connected a meaningful ways;
- entity structure: the entities referred to in the discourse have to be ordered in a certain way.
- And many other things too…

Coreference Resolution

Goal: predict what the (primarily) noun phrases in the text refer to

- John_i hid Bill_j's car keys. $\text{He}_{i/j}$ was drunk.

Many different cues can be used to disambiguate:

- Mary_i hid Bill_j's car keys. $\text{He}_{*i/j}$ was drunk.

Many other factors play a role

- syntactic structure, discourse relations, world knowledge.

Rhetorical Structure

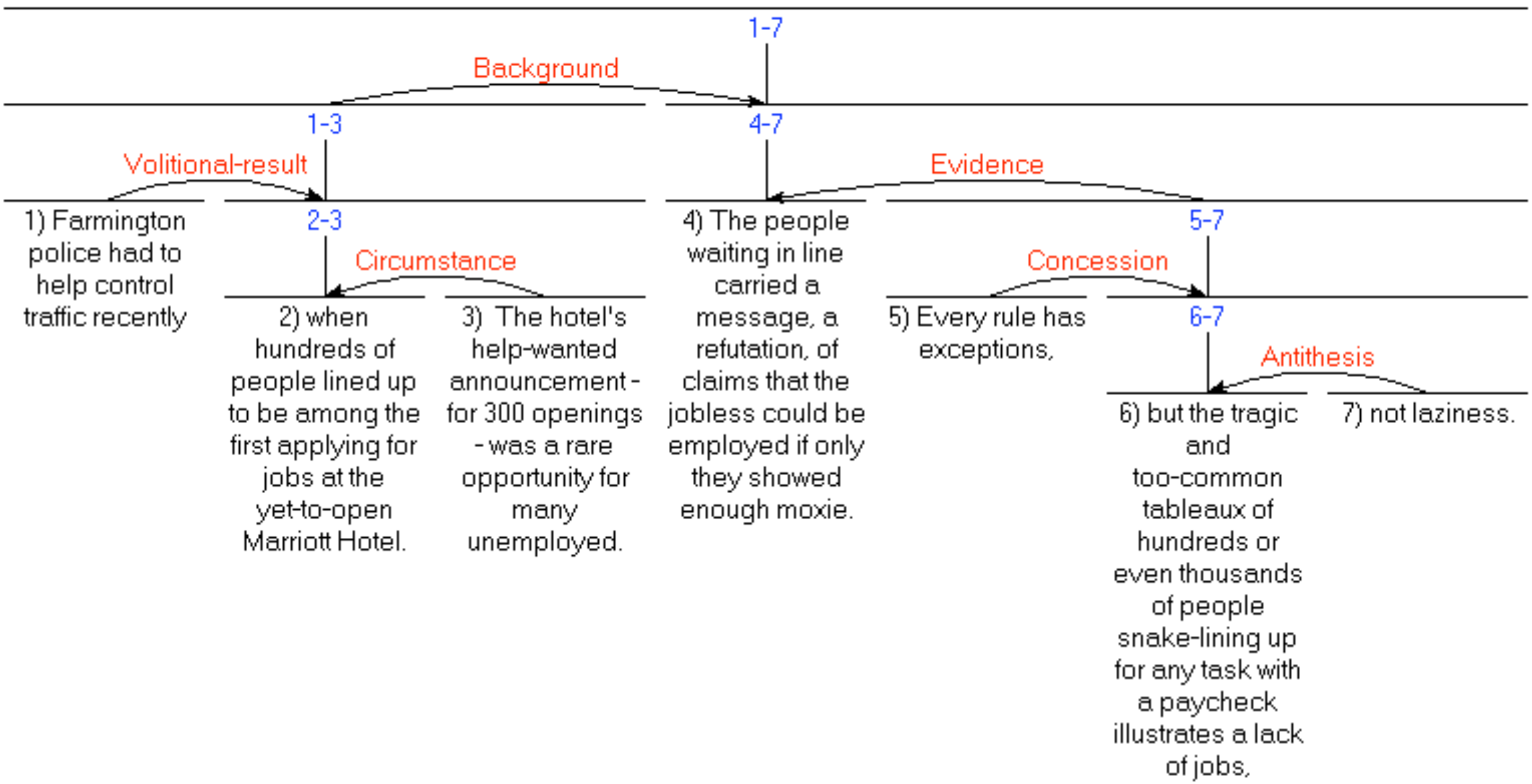
For a discourse to be *coherent*, utterances need to be juxtaposed in a meaningful way. Compare:

1. John_i hid Bill_j's car keys. $\text{He}_{i/j}$ was drunk.
2. John_i hid Bill_j's car keys. $\text{He}_{i/j}$ like spinach.

There is an likely explanation for (1), while (2) need a more elaborate back story…

- Relations such as EXPLANATION or CAUSE are called coherence relations (or discourse relations, or rhetorical relations).

Hierarchical Rhetorical Structure



RST website: http://www.sfu.ca/rst/

Entity Structure

(1):

a. John went to his favorite music store to buy a piano.
b. He had frequented the store for many years.
c. He was excited that he could finally buy a piano.
d. He arrived just as the store was closing for the day.

(2):

a. John went to his favorite music store to buy a piano.
b. It was a store that John had frequented for many years.
c. He was excited that he could finally buy a piano.
d. It was closing just as John arrived.

Question: Which text is more coherent? Why?

– Seems unnatural to alternate the *focus* between different entities?

Today: Focus on Co-reference

Problem definition

- Task, data, metrics, etc.

Many Different Approaches

- Clustering
- Classification
- Sieves
  - Error Analysis

The Problem: Find and Cluster Mentions

Victoria Chen, Chief Financial Officer of Megabucks banking corp since 2004, saw her pay jump 20%, to $1.3 million, as the 37 year old also became the Denver-based financial services company's president. It has been ten years since she came to Megbucks from rival Lotsabucks.

Mention Detection

[Victoria Chen]$_1$, [Chief Financial Officer of [Megabucks banking corp]$_2$ since 2004]$_3$, saw [[her]$_4$ pay]$_5$ jump 20%, to $1.3 million, as [the 37 year old]$_6$ also became the [[Denver-based financial services company]$_7$'s president]$_8$. It has been ten years since she came to [Megbucks]$_9$ from rival [Lotsabucks]$_{10}$.

The Problem: Find and Cluster Mentions

[Victoria Chen]$_{1}$, [Chief Financial Officer of [Megabucks banking corp]$_{2}$ since 2004]$_{3}$, saw [[her]$_{4}$ pay]$_{5}$ jump 20%, to $1.3 million, as [the 37 year old]$_{6}$ also became the [[Denver-based financial services company]$_{7}$'s president]$_{8}$. It has been ten years since she came to [Megbucks]$_{9}$ from rival [Lotsabucks]$_{10}$.

Mention Clustering

Co-reference chains:

1 {Victoria Chen, Chief Financial Officer...since 2004, her, the 37-year-old, the Denver-based financial services company's president}

2 {Megabucks Banking Corp, Denver-based financial services company, Megabucks}

3 {her pay}

4 {rival Lotsabucks}

Types of Noun Phrases

- Indefinite
  - **no determiner:** *walnuts*
  - **the indefinite determiner:** *a beautiful goose*
  - **numerals:** *three geese*
  - **indefinite quantifiers:** *some walnuts.*
  - **(indefinite) this:** *this beautiful Ford Falcon*
- Definite
  - **definite** article: ***the*** *book*
  - **demonstrative articles:** ***this/that*** *book,* ***these/those*** *books*
  - **possessives:** ***my/John's*** *book*
  - **personal pronouns:** *I, he*
  - **demonstrative pronouns:** *this, that, these, those*
  - universal **quantifiers**: *all, every*
  - (unmodified) **proper nouns:** *John Smith, Mary, Urbana*

Prince's Entity Information Status

- **Hearer-new vs. hearer-old**

  Is the speaker referring to something the hearer knows (even for the first time)?
  - Hearer-old: *I will call Sandra Thompson*.
  - Hearer-new: *I will call a colleague in California* (=Sandra Thompson)
  - Special case: **hearer-inferrable --** *My husband …*

- **Discourse-new vs. discourse-old:**
  Is the speaker introducing a new entity into the discourse?
  - *I will call her/Sandra now.*

An Unsupervised Clustering Approach

The coreference problem can be solved by assigning all NPs in the text to equivalence classes, i.e., by clustering. [Cardie and Wagstaff, 1999]

We need:

- a *representation* of NPs (as a set of features)
- a *distance metric*
- a clustering *algorithm*.

Data Sets

Corpora	# Documents	# Sentences	# Words	# Entities	# Mentions
OntoNotes-Dev	303	6,894	136K	3,752	14,291
OntoNotes-Test	322	8,262	142K	3,926	16,291
ACE2004-Culotta-Test	107	1,993	33K	2,576	5,455
ACE2004-nwire	128	3,594	74K	4,762	11,398
MUC6-Test	30	576	13K	496	2,136

Table 3
Corpora statistics.

- Traditionally, systems have used different sets
  - Has made direct comparison surprisingly difficult…
- Differing assumptions about mentions
  - We will assume gold standard in this lecture

Evaluation Metrics

- Difficult to agree on the single best metric
  - 5-6 are used in practice, often with an average score
- For gold mentions, can use: G – gold, S -- system
  - MUC (Vilain et al. 1995) – cluster level -- p(X) is partitions of X
    - Roughly, number of clusters to be merge to make S match G

$$R = \frac{\sum (|G_i| - |p(G_i)|)}{\sum (|G_i| - 1)} \quad P = \frac{\sum (|S_i| - |p(S_i)|)}{\sum (|S_i| - 1)}$$

  - B^3 (Bagga and Baldwin 1998) – mention level
    - Roughly, cluster overlap between S and G, averaged over mention m_i

$$R = \sum_i \frac{|G_{m_i} \cap S_{m_i}|}{|G_{m_i}|}, P = \sum_i \frac{|G_{m_i} \cap S_{m_i}|}{|S_{m_i}|},$$

An Unsupervised Clustering Approach

The coreference problem can be solved by assigning all NPs in the text to equivalence classes, i.e., by clustering. [Cardie and Wagstaff, 1999]

We need:

- a *representation* of NPs (as a set of features)
- a *distance metric*
- a clustering *algorithm*.

Representing Mentions

Each NP is represented as a set of features:

- head noun: last word of the NP;
- position in the document;
- pronoun type: nominative, accusative, possessive, ambiguous;
- article: indefinite, definite, none;
- appositive: based on heuristics (commas, etc.)
- number: plural, singular;
- proper name: based on heuristics (capitalization, etc.);
- semantic class: based on Wordnet;
- gender: masculine, feminine, either, neuter;
- animacy: based on semantic class.

Example Mentions

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

Clustering

Distance Metric

$$dist(NP_1, NP_2) = \sum_{f \in F} w_f \cdot incompatibility_f(NP_1, NP_2)$$

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 Algorithm

- start from end of document, repeatedly merge compatible classes, compute transitive closure

Two Recent Unsupervised Learners

- Hierarchical Bayesian Model
  - [Haghighi & Klein, 2007, 2010]
  - Aims to learn head-word semantics at scale, more fine grained NP types, includes a discourse model, etc.
  - ~70 MUC F1 (approx.; used different test, but beat strong supervised system)

- Markov Logic Networks
  - [Poon & Domingos, 2008]
  - Joint inference across mentions
  - Many decisions are "easy" others more difficult
  - 70.9 MUC F1

Supervised Learning Approaches

- Treat co-reference as a classification problem
- Binary:
  - for all mention pairs m_i and m_j, are they coreferent?
  - Challenge: how to make coherent clusters
- Ranking:
  - for each mention m_i, select from $\{null, m_1, \ldots, m_{i-1}\}$
  - Questions: what are the advantages / disadvantages

Pairwise Model: Features matter! [Bengston & Roth, 2008]

Category	Feature	Source
Mention Types	Mention Type Pair	Annotation and tokens
String Relations	Head Match	Tokens
	Extent Match	Tokens
	Substring	Tokens
	Modifiers Match	Tokens
	Alias	Tokens and lists
Semantic	Gender Match	WordNet and lists
	Number Match	WordNet and lists
	Synonyms	WordNet
	Antonyms	WordNet
	Hypernyms	WordNet
	Both Speak	Context
Relative Location	Apposition	Positions and context
	Relative Pronoun	Positions and tokens
	Distances	Positions
Learned	Anaphoricity	Learned
	Name Modifiers Predicted Match	Learned
Aligned Modifiers	Aligned Modifiers Relation	WordNet and lists
Memorization	Last Words	Tokens
Predicted Entity Types	Entity Types Match	Annotation and tokens
	Entity Type Pair	WordNet and tokens

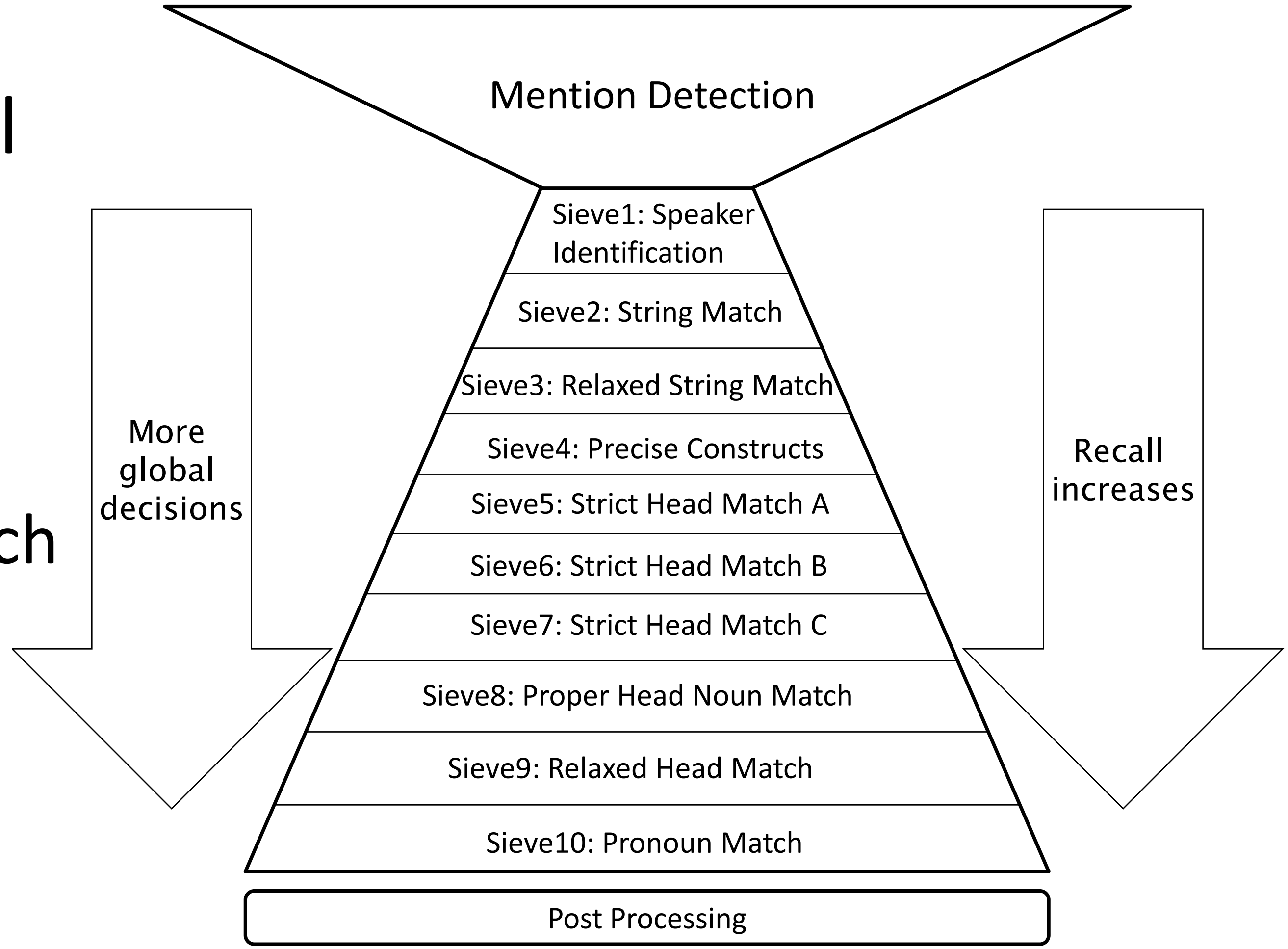
Two Recent Supervised Learners

- Linear Model
  - [Bengston & Roth 2008]
  - Pairwise classification
  - Careful experimental setup with tons of features!
  - 80.8 B^3 F1

- FOL-based approach
  - [Culotta et al. 2007]
  - Includes global constraints on clusters
  - 79.3 B^3 F1

Multi-pass Sieve

- Basically, a ranking model with no machine learning!
  - 10 sieves, each very simple
  - Winner of CONLL 2011 competition!



A Carefully Constructed Example

Input:	John is a musician. He played a new song. A girl was listening to the song. "It is my favorite," John said to her.
Mention Detection:	[John]$^{1}_{1}$ is [a musician]$^{2}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{6}_{6}$. "[It]$^{7}_{7}$ is [[my]$^{9}_{9}$ favorite]$^{8}_{8}$," [John]$^{10}_{10}$ said to [her]$^{11}_{11}$.
Speaker Sieve:	[John]$^{1}_{1}$ is [a musician]$^{2}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{6}_{6}$. "[It]$^{7}_{7}$ is [**[my]**$^{9}_{9}$ favorite]$^{8}_{8}$," **[John]**$^{9}_{10}$ said to [her]$^{11}_{11}$.
String Match:	**[John]**$^{1}_{1}$ is [a musician]$^{2}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{6}_{6}$. "[It]$^{7}_{7}$ is [[my]$^{1}_{9}$ favorite]$^{8}_{8}$," **[John]**$^{1}_{10}$ said to [her]$^{11}_{11}$.
Relaxed String Match:	[John]$^{1}_{1}$ is [a musician]$^{2}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{6}_{6}$. "[It]$^{7}_{7}$ is [[my]$^{1}_{9}$ favorite]$^{8}_{8}$," [John]$^{1}_{10}$ said to [her]$^{11}_{11}$.
Precise Constructs:	**[John]**$^{1}_{1}$ is **[a musician]**$^{1}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{6}_{6}$. "**[It]**$^{7}_{7}$ is **[[my]**$^{1}_{9}$ **favorite]**$^{7}_{8}$," [John]$^{1}_{10}$ said to [her]$^{11}_{11}$.
Strict Head Match A:	[John]$^{1}_{1}$ is [a musician]$^{1}_{2}$. [He]$^{3}_{3}$ played **[a new song]**$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to **[the song]**$^{4}_{6}$. "[It]$^{7}_{7}$ is [[my]$^{1}_{9}$ favorite]$^{7}_{8}$," [John]$^{1}_{10}$ said to [her]$^{11}_{11}$.
Strict Head Match B,C:	[John]$^{1}_{1}$ is [a musician]$^{1}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{4}_{6}$. "[It]$^{7}_{7}$ is [[my]$^{1}_{9}$ favorite]$^{7}_{8}$," [John]$^{1}_{10}$ said to [her]$^{11}_{11}$.
Proper Head Noun Match:	[John]$^{1}_{1}$ is [a musician]$^{1}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{4}_{6}$. "[It]$^{7}_{7}$ is [[my]$^{1}_{9}$ favorite]$^{7}_{8}$," [John]$^{1}_{10}$ said to [her]$^{11}_{11}$.
Relaxed Head Match:	[John]$^{1}_{1}$ is [a musician]$^{1}_{2}$. [He]$^{3}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{4}_{6}$. "[It]$^{7}_{7}$ is [[my]$^{1}_{9}$ favorite]$^{7}_{8}$," [John]$^{1}_{10}$ said to [her]$^{11}_{11}$.
Pronoun Match:	**[John]**$^{1}_{1}$ is [a musician]$^{1}_{2}$. **[He]**$^{1}_{3}$ played [a new song]$^{4}_{4}$. **[A girl]**$^{5}_{5}$ was listening to **[the song]**$^{4}_{6}$. "**[It]**$^{4}_{7}$ is [[my]$^{1}_{9}$ favorite]$^{4}_{8}$," [John]$^{1}_{10}$ said to **[her]**$^{5}_{11}$.
Post Processing:	[John]$^{1}_{1}$ is **a musician**. [He]$^{1}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{4}_{6}$. "[It]$^{4}_{7}$ is **[my]**$^{1}_{9}$ **favorite**," [John]$^{1}_{10}$ said to [her]$^{5}_{11}$.
Final Output:	[John]$^{1}_{1}$ is a musician. [He]$^{1}_{3}$ played [a new song]$^{4}_{4}$. [A girl]$^{5}_{5}$ was listening to [the song]$^{4}_{6}$. "[It]$^{4}_{7}$ is [my]$^{1}_{9}$ favorite," [John]$^{1}_{10}$ said to [her]$^{5}_{11}$.

Table 1

The Most Useful Sieves

- **2: Exact string match --** e.g., *[the Shahab 3 ground- ground missile]* and *[the Shahab 3 ground-ground missile]*. Precision is over 90% B3 **[+16 F1]**
- **5: Entity head match** – The mention head word matches *any* head word of mentions in the antecedent entity. Also, looks ar modifiers, e.g. to separate *Harvard University* and *Yale Universit*y. **[+3 F1]**
- **10: Pronominal Coreference Resolution** – observe constraints on number, gender, person, animacy, and NER types. Link to closest, with a maximum distance. **[+10 F1]**
- Most others get between 0-2 points improvement, but are cumulative

Some Results

System	MUC			B^3		
	R	P	F1	R	P	F1
ACE2004-Culotta-Test						
This paper	70.2	82.7	75.9	74.5	88.7	81.0
Haghighi and Klein (2009)	77.7	74.8	79.6	78.5	79.6	79.0
Culotta et al. (2007)	–	–	–	73.2	86.7	79.3
Bengston and Roth (2008)	69.9	82.7	75.8	74.5	88.3	80.8
ACE2004-nwire						
This paper	75.1	84.6	79.6	74.1	87.3	80.2
Haghighi and Klein (2009)	75.9	77.0	76.5	74.5	79.4	76.9
Poon and Domingos (2008)	70.5	71.3	70.9	–	–	–
Finkel and Manning (2008)	58.5	78.7	67.1	65.2	86.8	74.5
MUC6-Test						
This paper	69.1	90.6	78.4	63.1	90.6	74.4
Haghighi and Klein (2009)	77.3	87.2	81.9	67.3	84.7	75.0
Poon and Domingos (2008)	75.8	83.0	79.2	–	–	–
Finkel and Manning (2008)	55.1	89.7	68.3	49.7	90.9	64.3

Table 5
Comparison of our system with the other reported results on the ACE and MUC corpora. All these systems use gold mention boundaries.

[Lee et al, 2013]

Error Analysis

Error type	Percentage
Semantics, discourse	41.7
Pronominal resolution errors	28.7
Non-referential mentions	14.8
Event mentions	6.1
Miscellaneous	8.7

Error type	Example
Semantics, discourse	• Lincoln's parent company, American Continental Corp., entered bankruptcy - law proceedings this April 13, and regulators seized *the thrift* the next day. . . . Mr. Keating has filed his own suit, alleging that **his property** was taken illegally. • *New pictures* reveal the sheer power of that terrorist bomb . . . In **these photos obtained by NBC News**, the damage much larger than first imagined . . . • Of all the one-time expenses incurred by a corporation or professional firm, few are larger or longer term than *the purchase of real estate or the signing of a commercial lease* . . . To take full advantage of the financial opportunities in **this commitment**, . . .
Pronominal resolution errors	Under the laws of the land, *the ANC* remains an illegal organization , and **its** headquarters are still in Lusaka, Zambia.
Non-referential mentions	When you become a federal judge, all of a sudden **you** are relegated to a paltry sum.
Event mentions	"*Support* the troops, not the regime" **That** 's a noble idea until you're supporting the weight of an armoured vehicle on your chest.
Miscellaneous (inconsistent annotations, parser or NER errors, enumerations)	• Inconsistent annotation - Inclusion of *'s*: . . . that's without adding in [*Business Week 's*] charge . . . Small wonder that [*Britain*] 's Labor Party wants credit controls. • Parser or NER error: Um alright uh *Mister Zalisko* do **you** know anything from your personal experience of having been on the cruise as to what happened? – *Mister Zalisko* is not recognized as a PERSON • Enumerations: This year, the economies of the five large special economic zones, namely, Shenzhen, Zhuhai, Shantou, Xiamen and Hainan, have maintained strong growth momentum. . . . A three dimensional traffic frame in **Zhuhai** has preliminarily taken shape and the investment environment improves daily.

[Lee et al, 2013]

Table 12

Joint Models of Entities and Events

- E.g., "Joint Entity and Event Coreference Resolution across Documents" [Lee et al, 2012]

1. (a) **One of the key suspected Mafia bosses arrested yesterday** <u>has hanged</u> himself.
   (b) Police said **Lo Presti** <u>had hanged</u> himself.
   (c) <u>His suicide</u> appeared to be related to clan feuds.
2. (a) **The New Orleans Saints** <u>placed</u> **Reggie Bush** on the injured list on Wednesday.
   (b) **Saints** <u>put</u> **Bush** on I.R.