

Coreference Semantics from Web Features



Mohit Bansal and Dan Klein
UC Berkeley

Motivation



Example

*When **Obama** met **Jobs**, **the president** discussed the economy, technology, and education. **His** election campaign is expected to ...*

- ▶ World knowledge needed:
 - ▶ Obama is the president of the US
 - ▶ Presidents, not CEOs, have election campaigns



Example

(Obama, president)

(Jobs, president)

When Obama met Jobs, the president discussed ...



Example

(Obama's election campaign)

(Jobs' election campaign)

When Obama met Jobs ... His election campaign ...



This Work

- ▶ No new model
- ▶ Simple, principled features that subsumes previous work
- ▶ Features computed from Google n-grams only

Baseline System

Reconcile (Stoyanov et al., 2009)



Pairwise Supervised Coreference

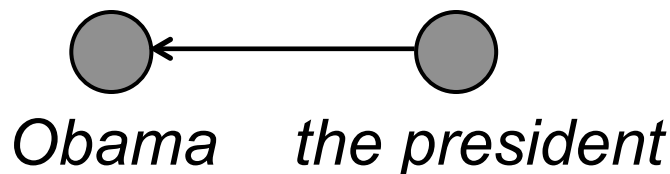
$$f(m_1, m_2) \Rightarrow \text{score}(f(m_1, m_2))$$

When Obama met Jobs , the president discussed ...

m_1 m_2



Pairwise Features

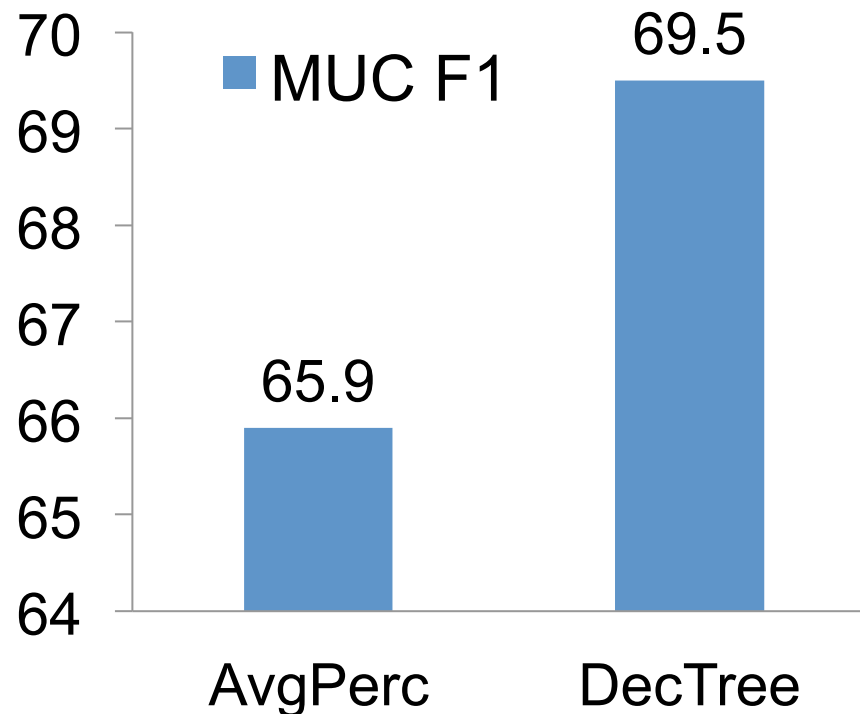


Category	Feature
POSITIONAL	Distance in # Sentences
LEXICAL	Approximate String Match
GRAMMATICAL	Number Agreement
	Gender Agreement
	Appositive Relationship
SEMANTIC	WordNet Synonyms
	Alias



One Important Change

► Averaged Perceptron ➡ Decision Tree



World Knowledge via Web Features



World Knowledge

Barack Obama, the 44th **president** of the United States, will be sworn in ...





World Knowledge

Barack Obama, the 44th **president** of the United States, will be sworn in ...



Korea and other **countries** will be participating in this important event ...





World Knowledge

Barack Obama, the 44th **president** of the United States, will be sworn in ...



Korea and other **countries** will be participating in this important event ...



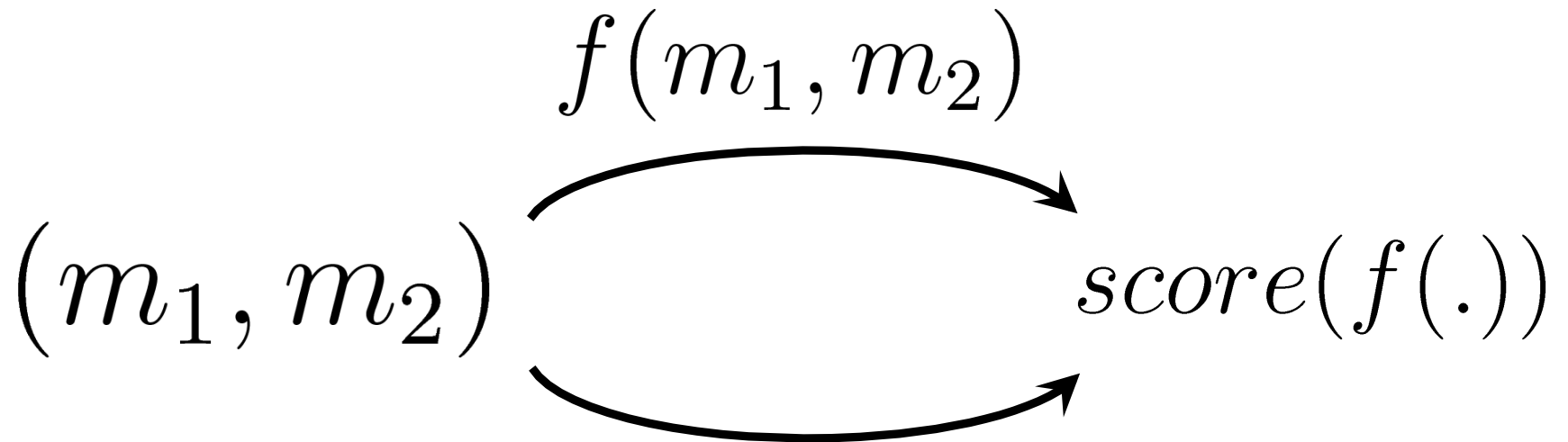
Asia is the largest **continent**, located primarily in the eastern and northern ...



WIKIPEDIA
The Free Encyclopedia



Web Features



$f(h_1, h_2)$ (Web)



Feature Categories

- ▶ General Co-occurrence
- ▶ Hearst Co-occurrence [Poesio et al., 2004; Markert & Nissim, 2005; Kobdani et al., 2011]*
- ▶ Entity-Based Context
- ▶ Distributional Soft Clustering [Daume' III & Marcu, 2005]*
- ▶ Pronoun Context [Yang et al., 2005; Bergsma & Lin, 2006]*



1. General Co-occurrence

$\text{count}(\textit{president} * \textit{leader}) = 11383$ $\text{count}(\textit{voter} * \textit{leader}) = 95$

When the president met a voter, the leader said ...



1. General Co-occurrence

$$C_{12}$$

where

$$c_{12} = \text{count}("h_1 \star h_2") + \text{count}("h_1 \star \star h_2") \\ + \text{count}("h_1 \star \star \star h_2")$$



1. General Co-occurrence

$$\frac{c_{12}}{c_1 \cdot c_2}$$

where

$$c_{12} = \text{count}("h_1 \star h_2") + \text{count}("h_1 \star \star h_2") \\ + \text{count}("h_1 \star \star \star h_2")$$

$$c_1 = \text{count}("h_1")$$

$$c_2 = \text{count}("h_2")$$



1. General Co-occurrence

$$\textit{bin} \left(\log \left(\frac{c_{12}}{c_1 \cdot c_2} \right) \right)$$

where

$$c_{12} = \textit{count}("h_1 \star h_2") + \textit{count}("h_1 \star \star h_2") \\ + \textit{count}("h_1 \star \star \star h_2")$$

$$c_1 = \textit{count}("h_1")$$

$$c_2 = \textit{count}("h_2")$$



2. Hearst Co-occurrence

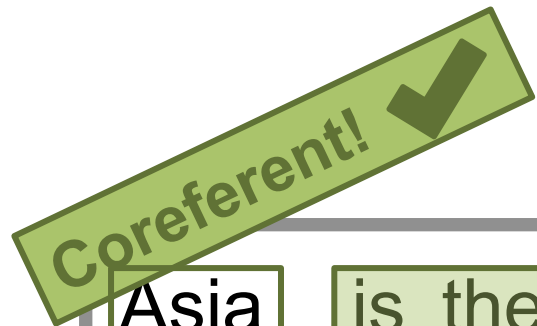
Not Coreferent! ✖

China and Japan are geographically separated only by a relatively narrow stretch of ocean. China has strongly influenced Japan with its writing system, architecture, culture, religion, philosophy, and law.





2. Hearst Co-occurrence



Asia is the largest continent of the world and has the highest population. It is located primarily in the eastern and northern hemispheres. It covers 8.7% of the Earth's total surface area and 30% of its land area.





2. Hearst Co-occurrence



ACL and other scientific societies are for people working on problems involving natural language and computation. An annual meeting is held each summer in locations where significant computational linguistics ...





2. Hearst Co-occurrence

$\text{count}(\textit{president} * \textit{leader}) = 752$ $\text{count}(\textit{voter} * \textit{leader}) = 0$

When the president met a voter, the leader said ...



2. Hearst Co-occurrence

► Hypernymy patterns:

- $h_1 \{is \mid are \mid was \mid were\} \{a \mid an \mid the\} ? h_2$



2. Hearst Co-occurrence

► Hypernymy patterns:

- $h_1 \{is \mid are \mid was \mid were\} \{a \mid an \mid the\} ? h_2$
- $h_1 \{and \mid or\} \{other \mid the \text{ other} \mid another\} h_2$



2. Hearst Co-occurrence

► Hypernymy patterns:

- $h_1 \{is \mid are \mid was \mid were\} \{a \mid an \mid the\} ? h_2$
- $h_1 \{and \mid or\} \{other \mid the \text{ other} \mid another\} h_2$
- $h_1 other \text{ than } \{a \mid an \mid the\} ? h_2$
- $h_1 such \text{ as } \{a \mid an \mid the\} ? h_2$



2. Hearst Co-occurrence

► Hypernymy patterns:

- $h_1 \{is \mid are \mid was \mid were\} \{a \mid an \mid the\} ? h_2$
- $h_1 \{and \mid or\} \{other \mid the other \mid another\} h_2$
- $h_1 other than \{a \mid an \mid the\} ? h_2$
- $h_1 such as \{a \mid an \mid the\} ? h_2$
- $h_1 , including \{a \mid an \mid the\} ? h_2$
- $h_1 , especially \{a \mid an \mid the\} ? h_2$



2. Hearst Co-occurrence

► Hypernymy patterns:

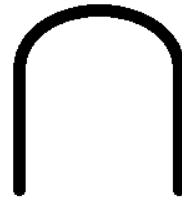
- $h_1 \{is \mid are \mid was \mid were\} \{a \mid an \mid the\} ? h_2$
- $h_1 \{and \mid or\} \{other \mid the \text{ other} \mid another\} h_2$
- $h_1 other \text{ than } \{a \mid an \mid the\} ? h_2$
- $h_1 such \text{ as } \{a \mid an \mid the\} ? h_2$
- $h_1 , including \{a \mid an \mid the\} ? h_2$
- $h_1 , especially \{a \mid an \mid the\} ? h_2$
- $h_1 of \{the \mid all\} ? h_2$



3. Entity-Based Context

$h_1 = \textit{president}$

president is elected
president is authorized
president is responsible
president is the chief
president is above
president is the head
...



$h_2 = \textit{leader}$

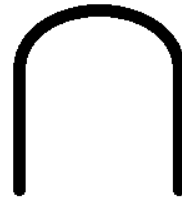
leader is responsible
leader is expected
leader is able
leader is elected
leader is chosen
leader is best
...



3. Entity-Based Context

$h_1 = \textit{president}$

president is elected
president is authorized
president is responsible
president is the chief
president is above
president is the head
...



$h_2 = \textit{leader}$

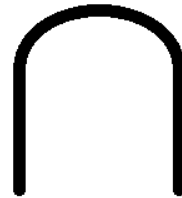
leader is responsible
leader is expected
leader is able
leader is elected
leader is chosen
leader is best
...



3. Entity-Based Context

$h_1 = \textit{president}$

president is elected
president is authorized
president is responsible
president is the chief
president is above
president is the head
...



$h_2 = \textit{leader}$

leader is responsible
leader is expected
leader is able
leader is elected
leader is chosen
leader is best
...



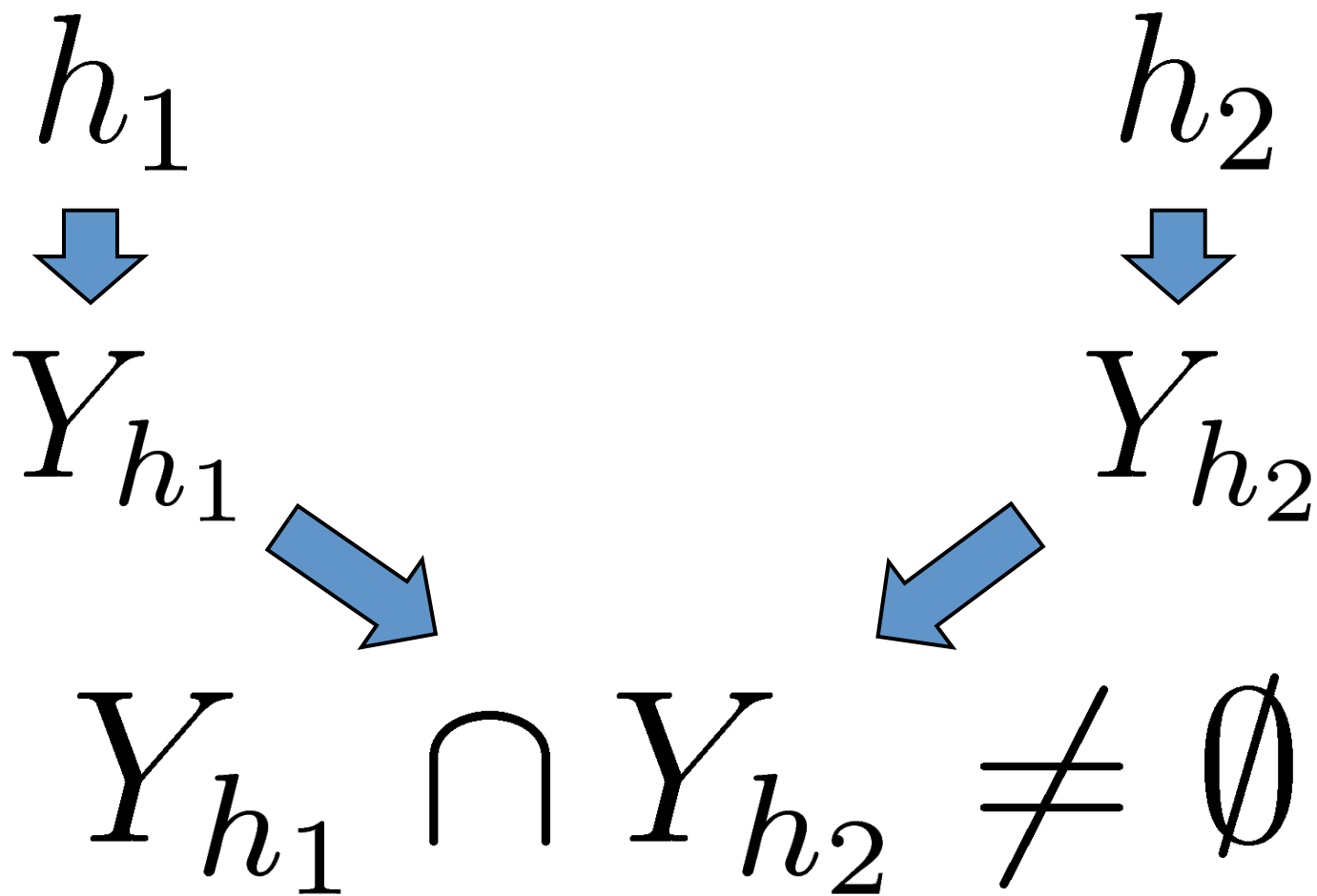
3. Entity-Based Context

$h \{is \mid are \mid was \mid were\} \{a \mid an \mid the\} ? y$

$$Y_h = \{y\}$$



3. Entity-Based Context





4. Distributional Soft Clustering

- ▶ Distributional hypothesis of Harris (1954)

Words that occur in similar contexts tend to have similar linguistic behavior.

- ▶ Applied to Web-scale clusters (Lin et al., 2010)
- ▶ Soft clustering assigns up to 20 clusters / word



4. Distributional Soft Clustering

president

C734
C284
C450
C976
C447
...

leader

C926
C985
C734
C974
C450
...

} Ranked



4. Distributional Soft Clustering

president

C734
C284
C450
C976
C447
...

i

leader

C926
C985
C734
C974
C450
...

j

Ranked

$bin(i + j)$



5. Pronoun Context

$h_2 + \text{context} = \textit{his election}$

$h_1 = \textit{Obama}$

$$\frac{\textit{count}(\textit{"Obama 's \star election"})}{\textit{count}(\textit{"\star 's \star election"})\textit{count}(\textit{"Obama"})}$$



5. Pronoun Context

$$\frac{\textit{count}(\textit{"}h_1's \star r\textit{"})}{\textit{count}(\textit{"} \star 's \star r\textit{"})\textit{count}(\textit{"}h_1\textit{"})}$$

Web Data



Datasets

► Google *n*-grams corpus (Brants & Franz, 2006)

president is the law	→	60
president is the leader	→	245
president is the least	→	58
president is the legal	→	50
president is the main	→	79

► Word-clusters from Lin et al. (2010)

presidency ”	→	C229	C127	C114	C129	C611	...
president	→	C734	C284	C450	C976	C447	...
president !	→	C548	C368	C645	C842	C583	...
president & CEO	→	C560	C293	C358	C944	C284	...
president's ability	→	C876	C754	C770	C212	C215	...

Coreference Experiments



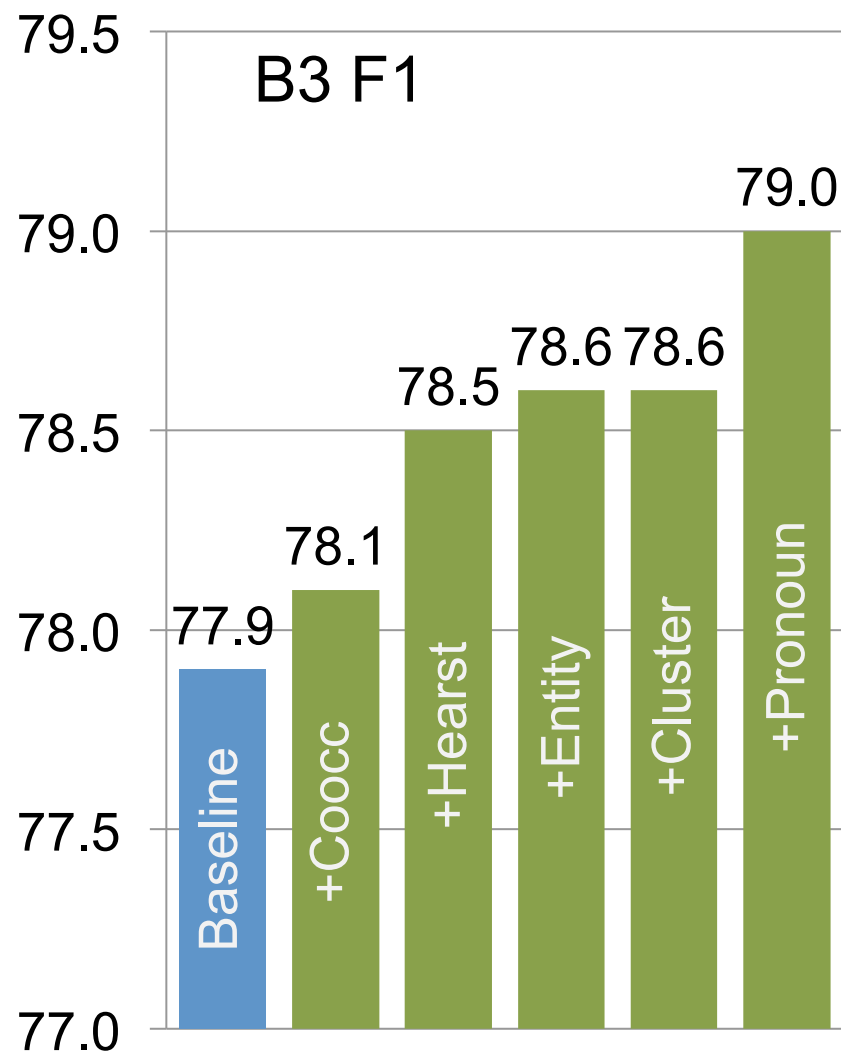
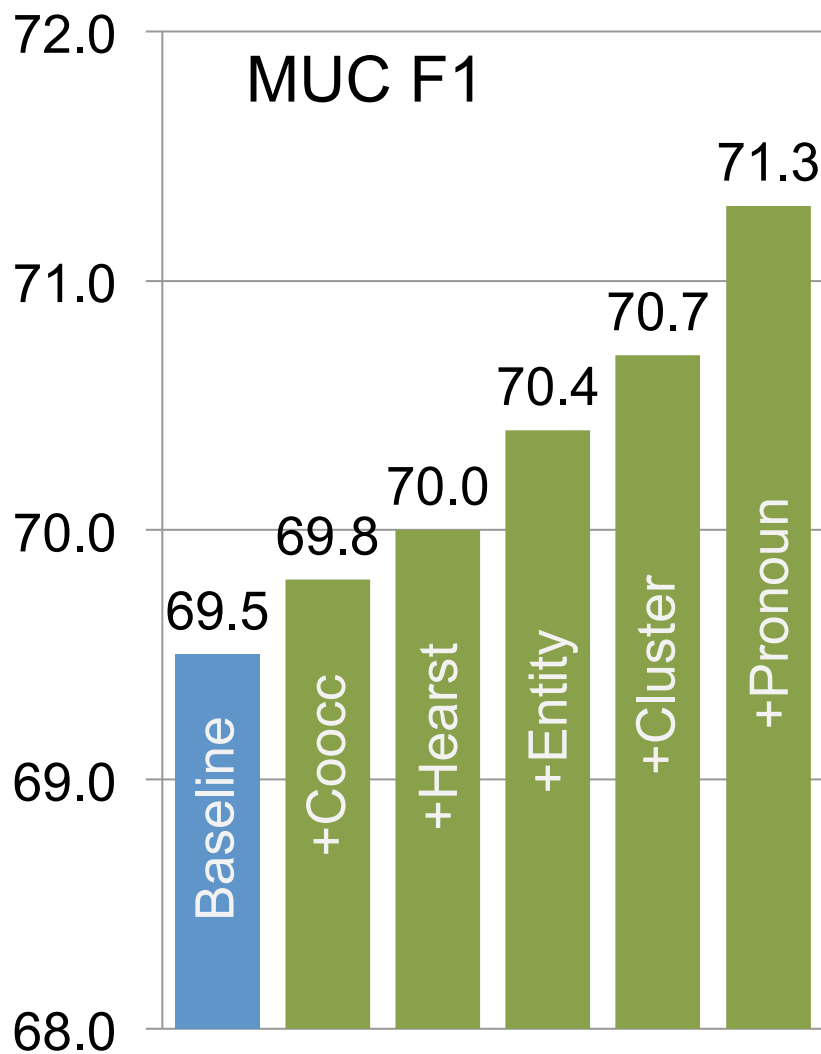
Datasets and Metrics

Dataset	#docs	#mentions	#chains
ACE04	128	3037	1332
ACE05	81	1991	775
ACE05-ALL	599	9217	3050

- ▶ 2 popular and complementary metrics
 - ▶ MUC
 - ▶ B³

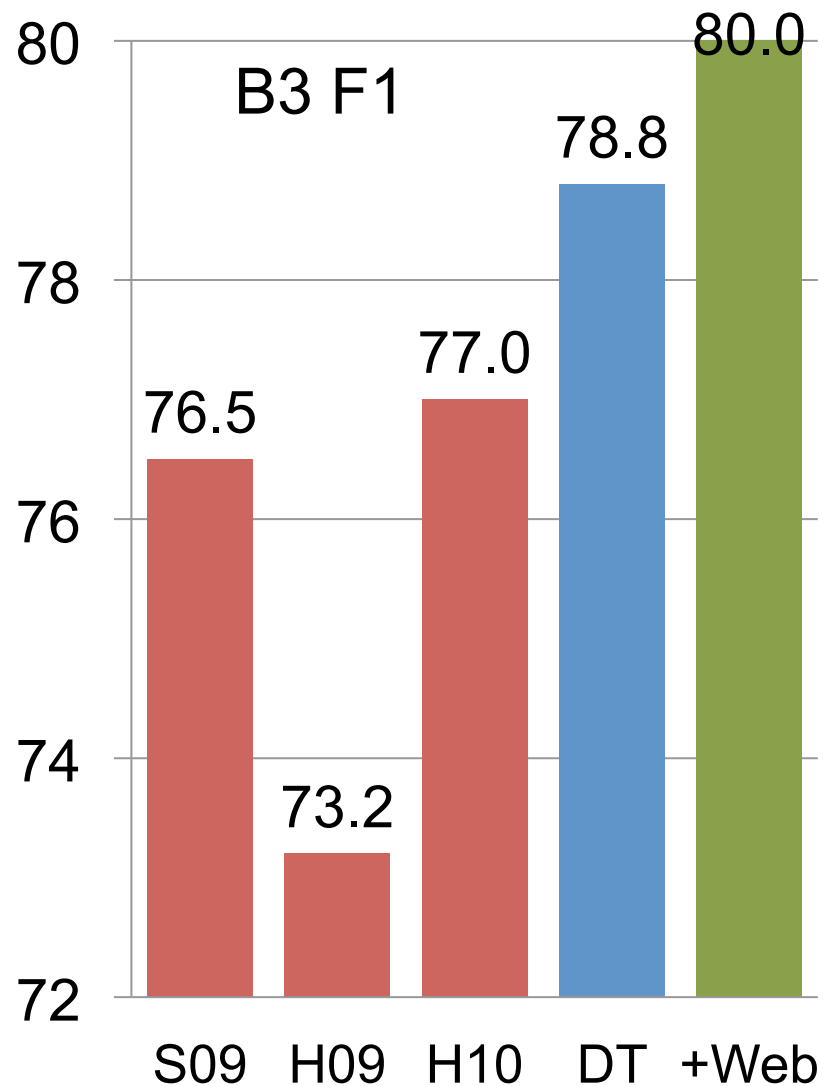
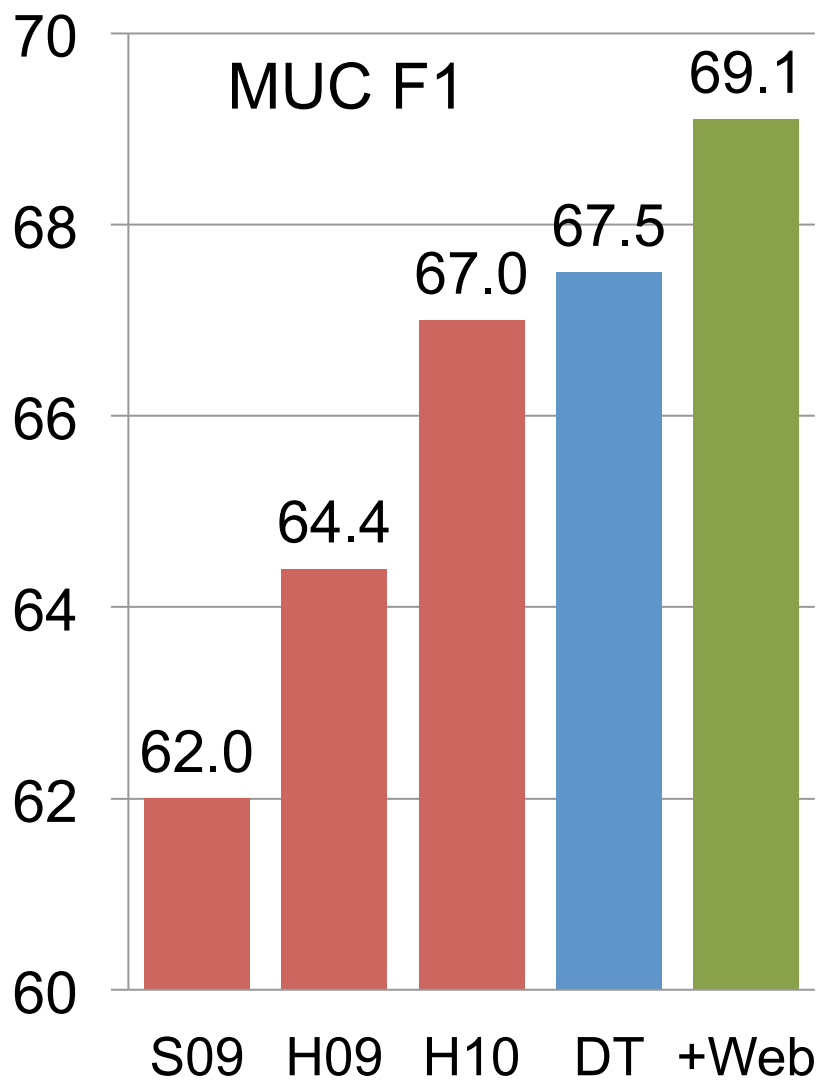


ACE04-DEV Incremental Results



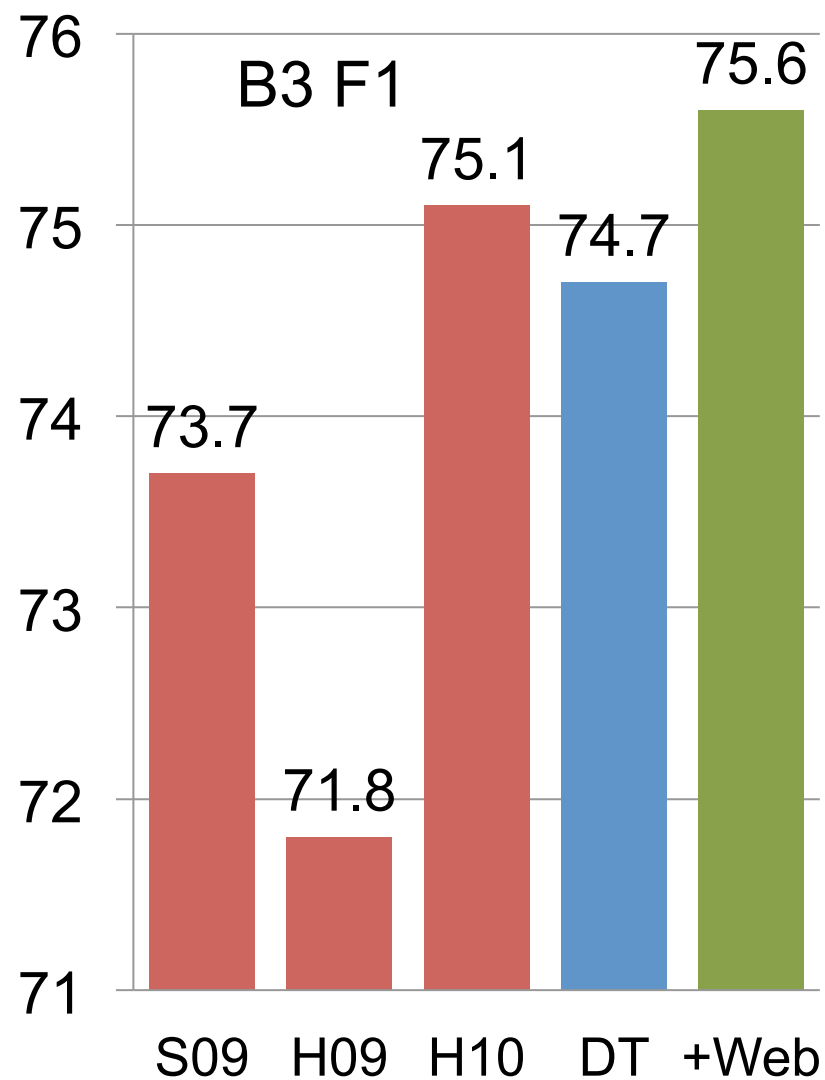
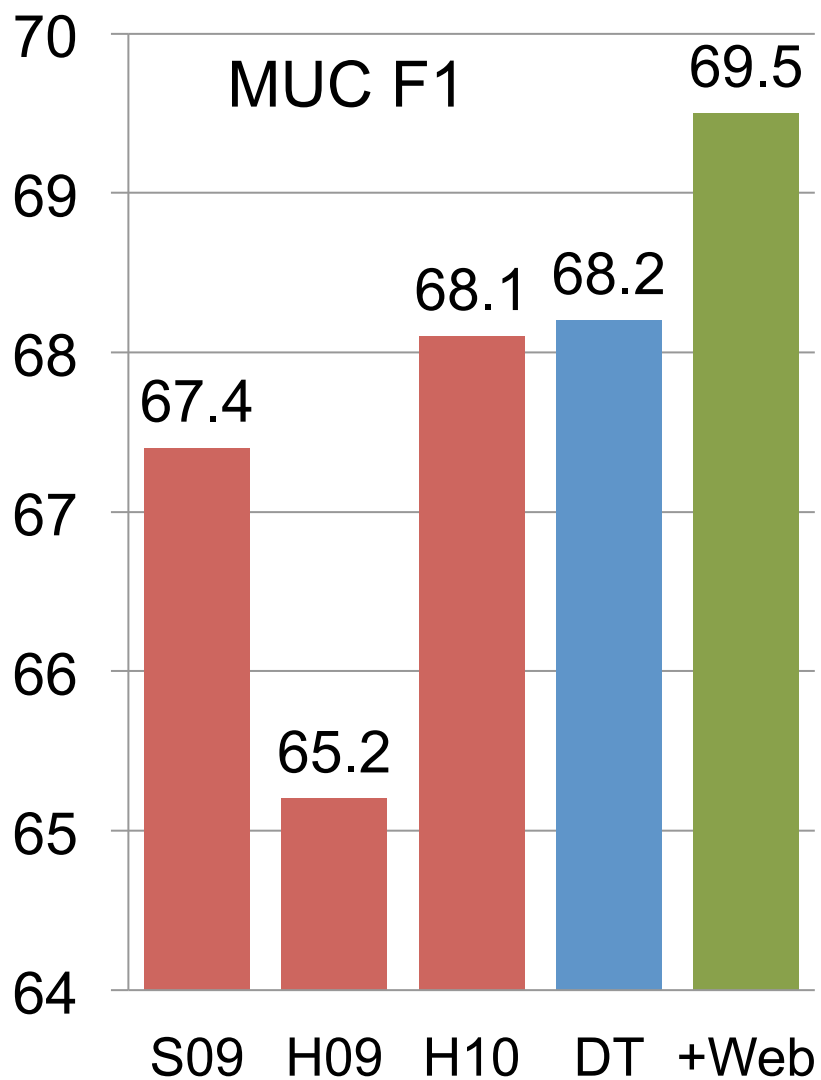


ACE04-TEST



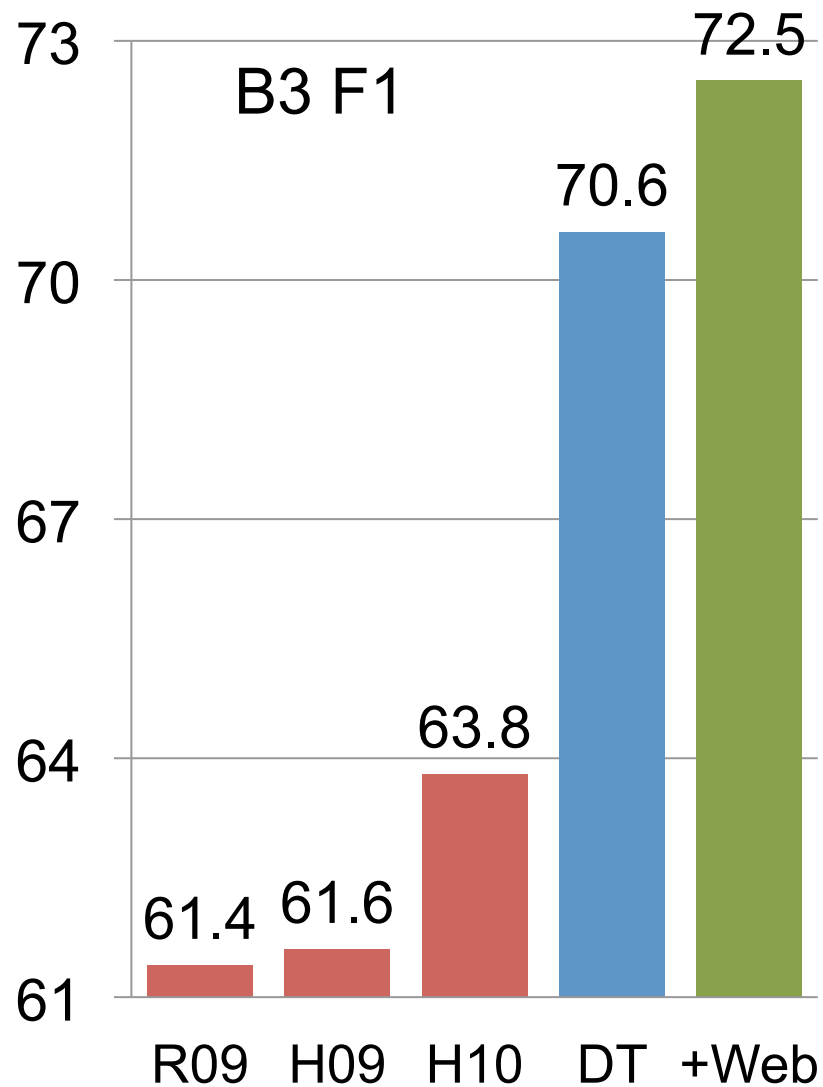
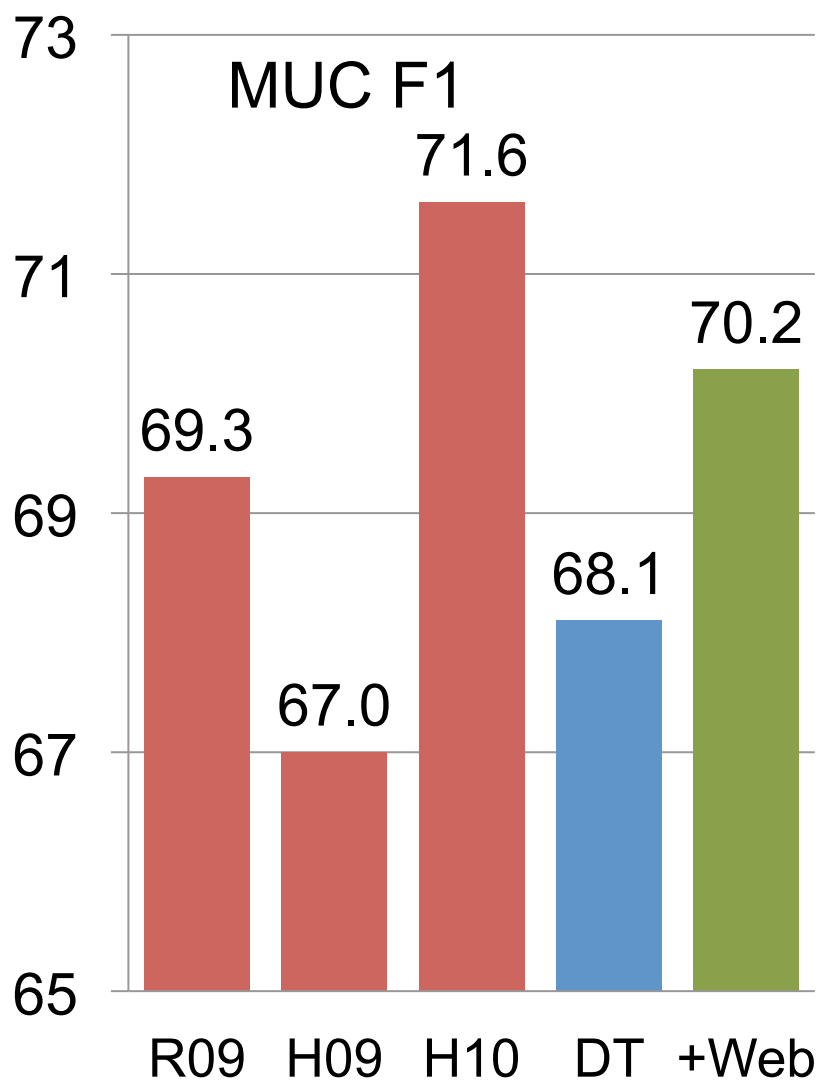


ACE05-TEST





ACE05-ALL-TEST



Analysis



Error Correction Analysis

- ▶ Coreferent pairs corrected by Web features:

Barry Bonds

the best baseball player

athletic director

Mulcahy

Democrat Al Gore

the vice president

Iran

the country

the EPA

the agency

Vojislav Kostunica

the pro-democracy leader



Decision Tree Analysis

- ▶ ~30% of the decision nodes are Web features
- ▶ Avg. classification error at Web leaves $< 3\%$
- ▶ Strongly discriminative nodes:
 - ▶ Hearst feature for its zero-count value
 - ▶ Cluster feature for its no-match value



Conclusion

- ▶ Simple Web features help significantly
- ▶ World knowledge via co-occurrence, context, hypernymy, and compatibility
- ▶ State-of-the-art results

Thank you!



Questions?