Statistical NLP
Spring 2010

Lecture 21: Coreference Resolution

Dan Klein – UC Berkeley

Natural Language Processing

Los Angeles Times

President Barack Obama received the Serve America Act after congress' vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

Question Answering

Q: Who signed the Serve America Act?

A: Barack Obama

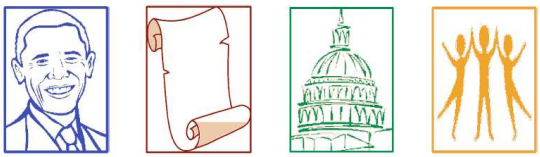
> **Los Angeles Times**
>
> President Barack Obama received the Serve America Act after congress' vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

Machine Translation

او	قانونرا	امظاءكرد
oo	ganoon-ra	emza-kard
person	*law-obj*	*signed-past*

He signed the bill

Discourse Structure



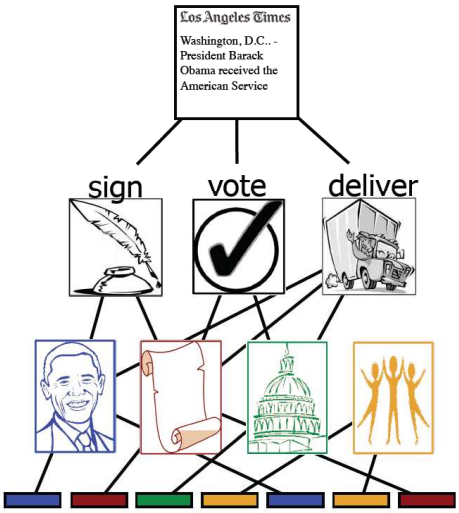
President Barack Obama received the Serve America Act after congress' vote. He signed the bill last Thursday. The president said it would greatly increase service opportunities for the American people.

Discourse Structure

Narrative

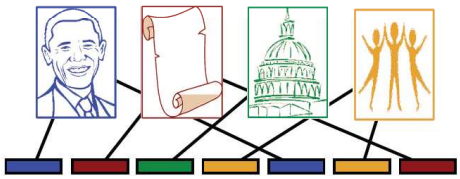
Events

Entities

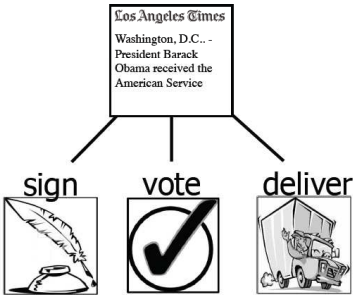


Outline

- Entity Reference Resolution



- Multi-Document Summarization



Entity Reference Resolution

Input:

President Barack Obama received the Serve America Act after congress' close vote.

He signed the bill last Thursday.

The president said it would greatly increase service opportunities for the American people.

Entity Reference Resolution

Output:

President Barack Obama received the Serve America Act after congress' close vote.

He signed the bill last Thursday.

The president said it would greatly increase service opportunities for the American people.

Entity Reference Resolution

President Barack Obama

He

The president

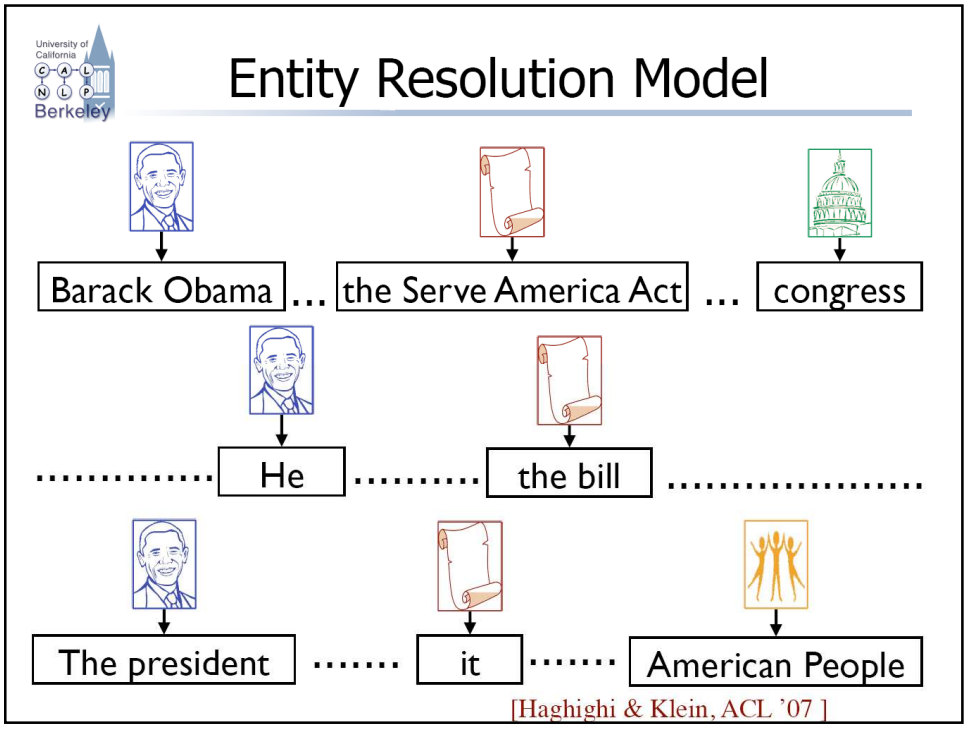
congress'

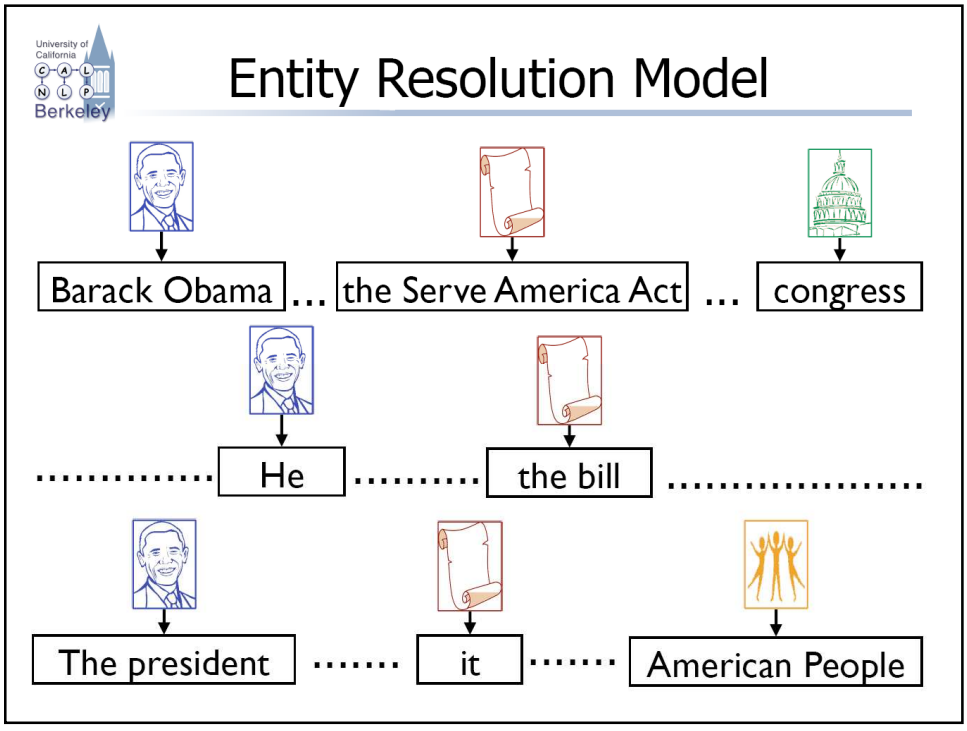
the Serve America Act

the bill

it

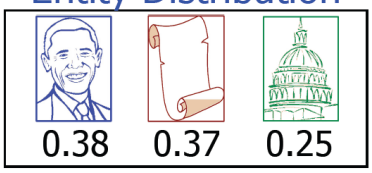
the American people


Entity Resolution Model
Barack Obama
the Serve America Act
congress
He
the bill
The president
it
American People
[Haghighi & Klein, ACL '07]


Entity Resolution Model
Barack Obama
the Serve America Act
congress
He
the bill
The president
it
American People

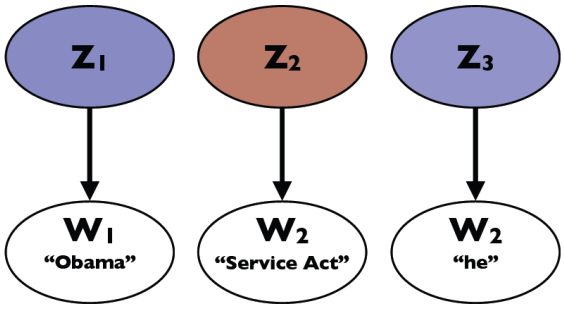
Finite Mixture Model

Entity Distribution



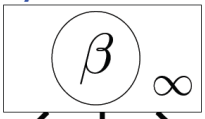
Barack Obama

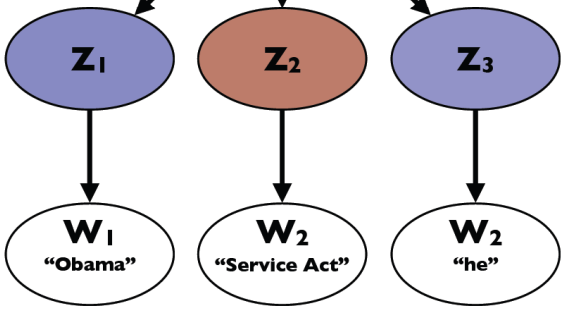
"Barack Obama": 0.30
"president": 0.20
"he": 0.15
"him": 0.20
....



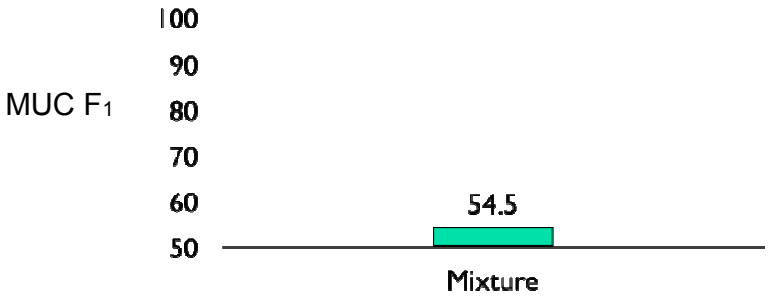
Infinite Mixture Model

Entity Distribution





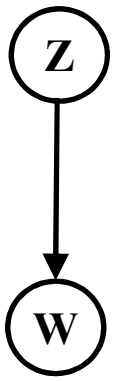
Infinite Mixture Model



The Weir Group , whose headquarters is in the U.S is a large specialized corporation. This power plant , which , will be situated in Jiangsu, has a large generation capacity.

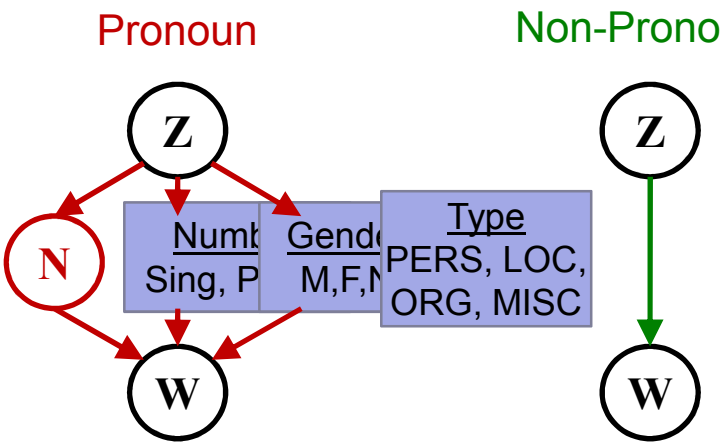
Enriching the Mention Model

Mention Model

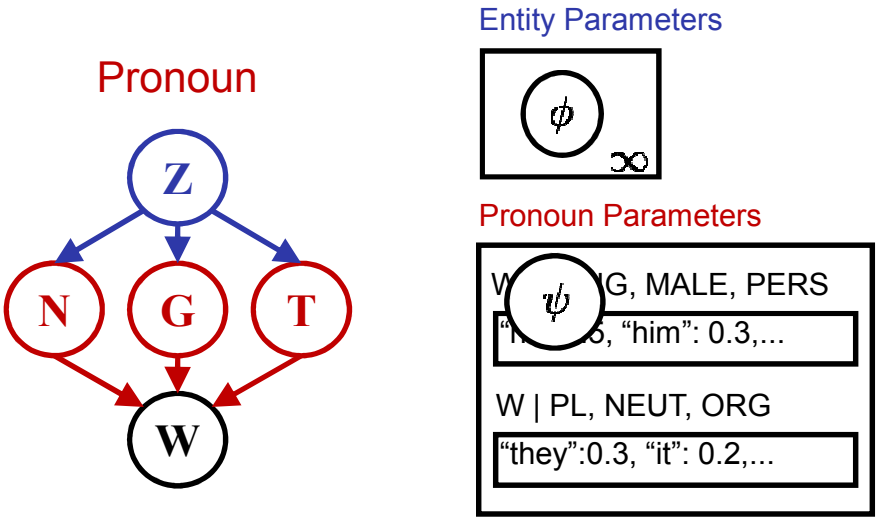


P(W | *Weir Group*):
"Weir Group"=0.4,
"whose"=0.2,
.....

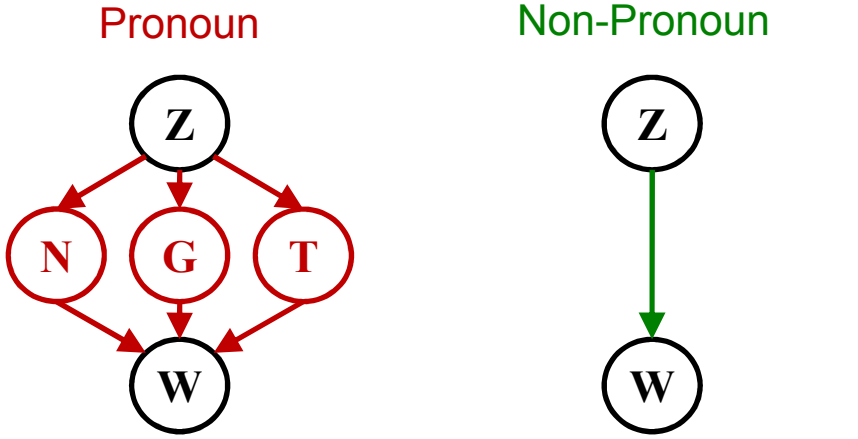
Enriching the Mention Model



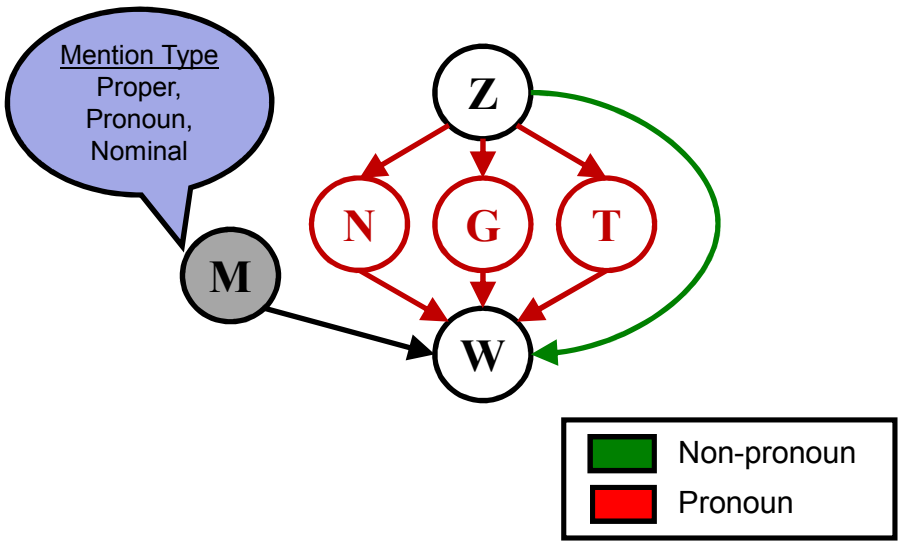
Enriching the Mention Model

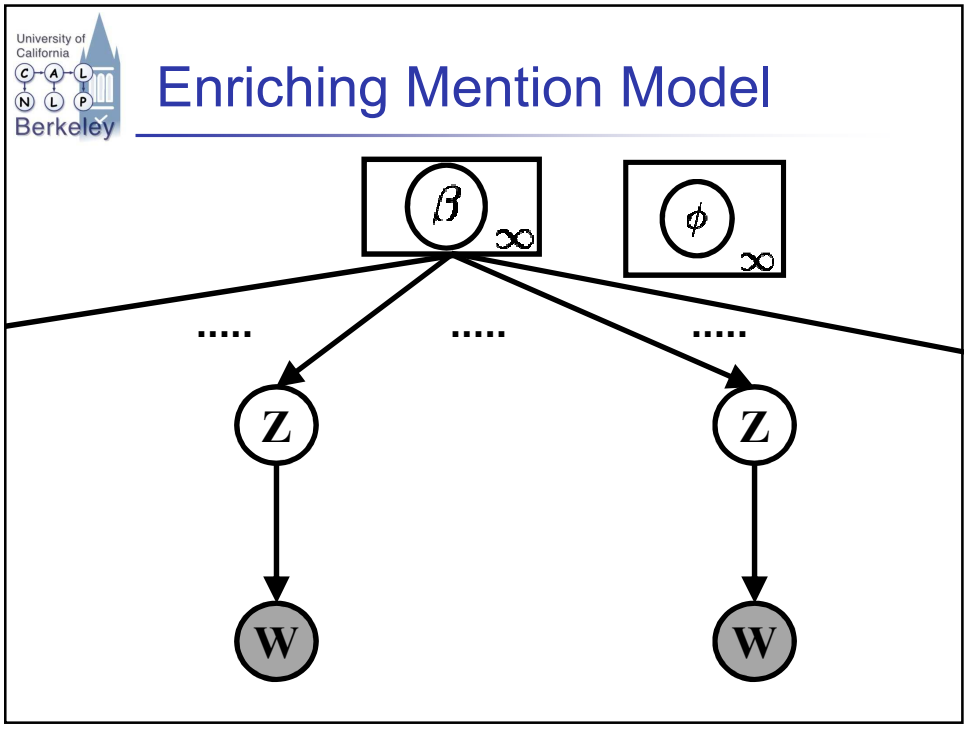


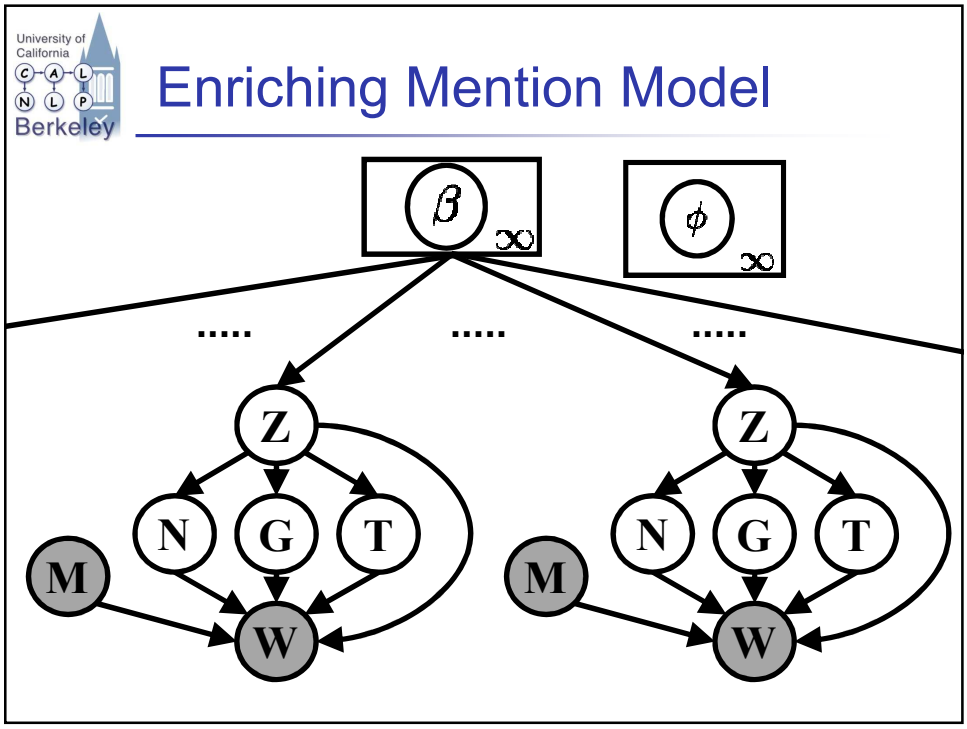
Enriching the Mention Model



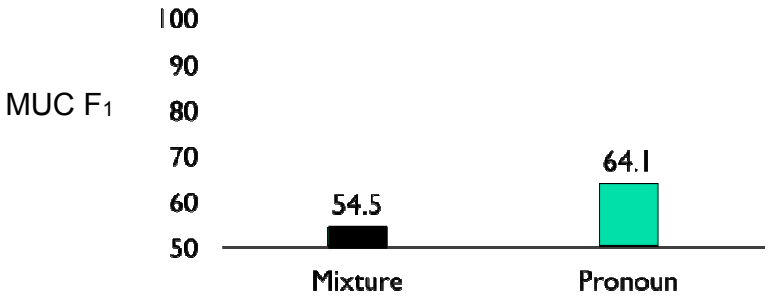
Enriching Mention Model




University of California
Berkeley
Enriching Mention Model
β
∞
ϕ
∞
.....
.....
.....
Z
Z
W
W


University of California
Berkeley
Enriching Mention Model
β
∞
ϕ
∞
.....
.....
.....
Z
Z
N
G
T
N
G
T
M
M
W
W

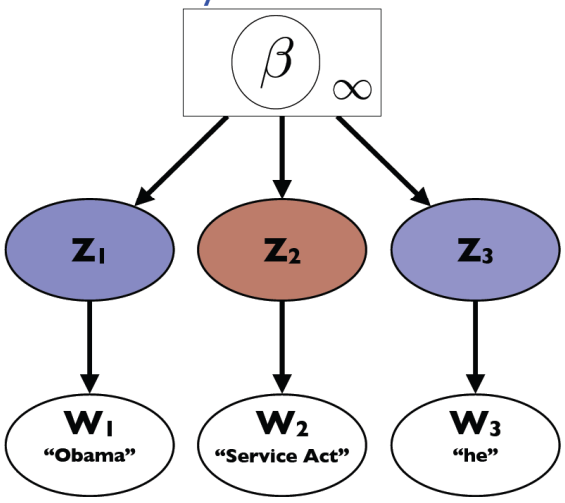
Pronoun Model



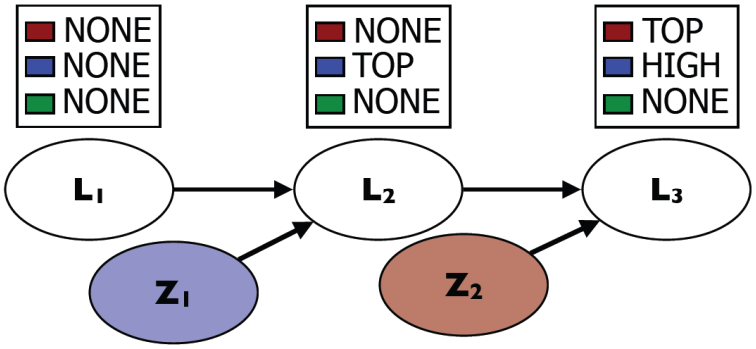
The Weir Group , whose headquarters is in the U.S is a large specialized corporation. This power plant , which , will be situated in Jiangsu, has a large generation capacity.

Discourse Salience

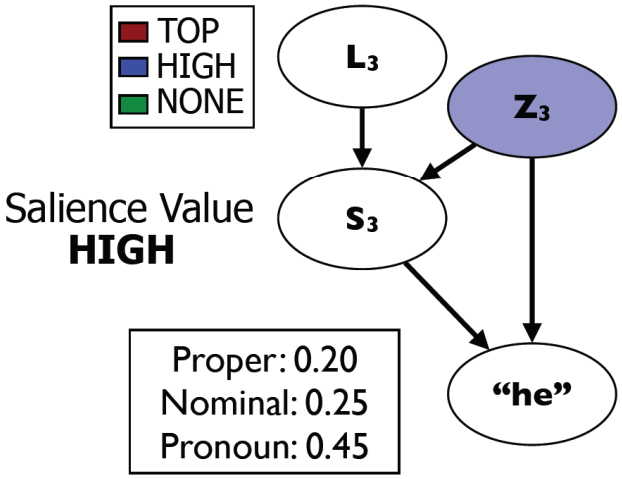
Entity Distribution



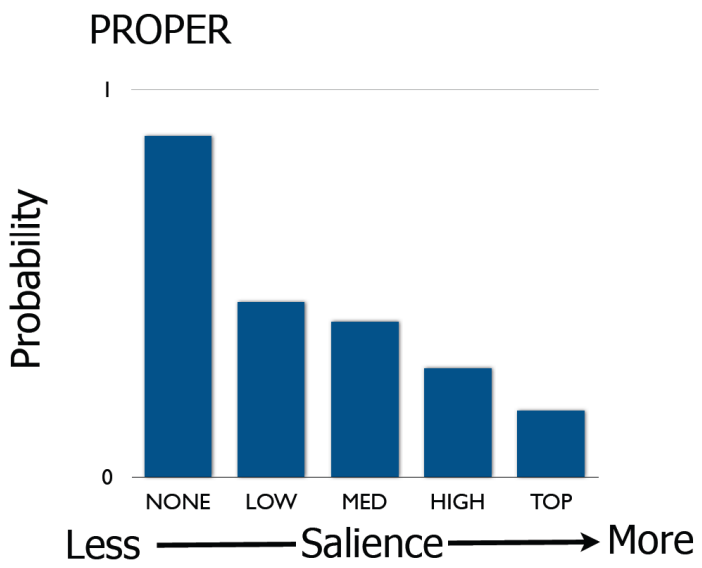
Salience List



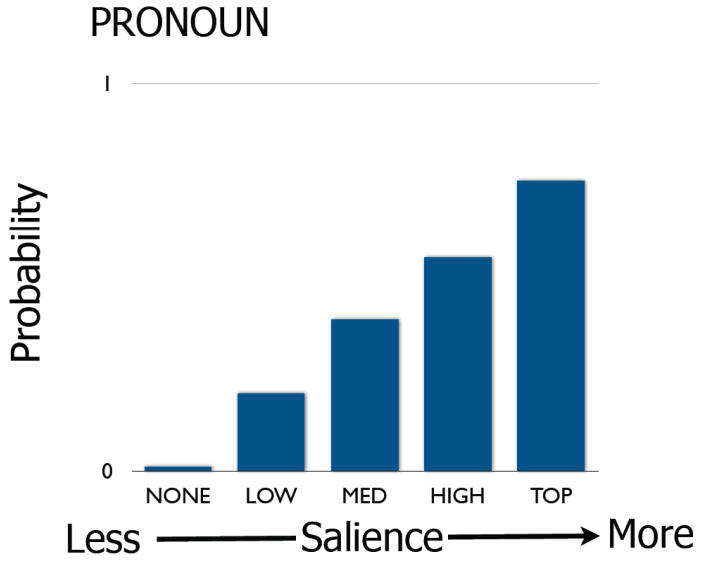
Discourse Salience



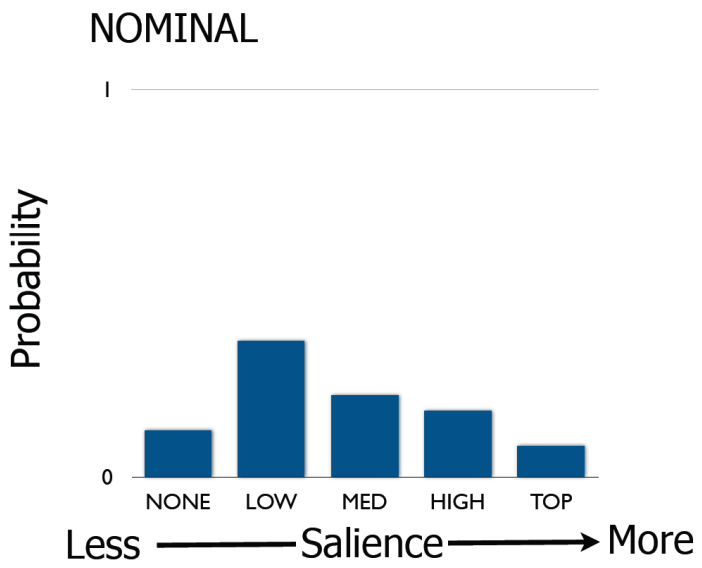
What is Learned?



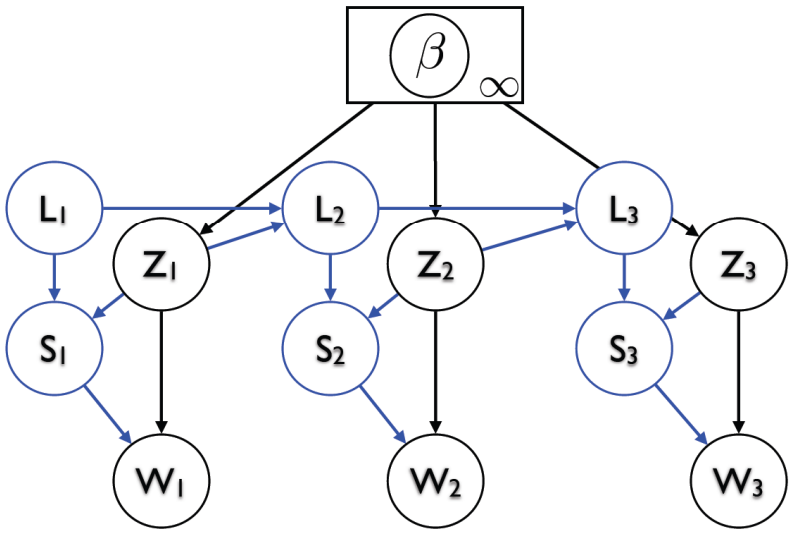
What is Learned?



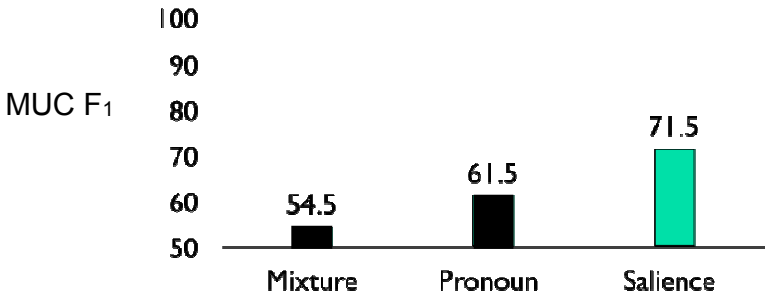
What is Learned?



Discourse Salience

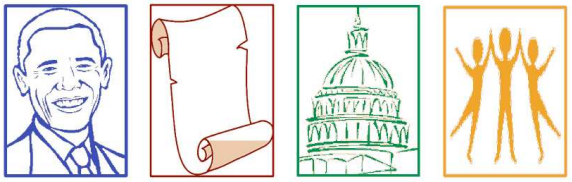


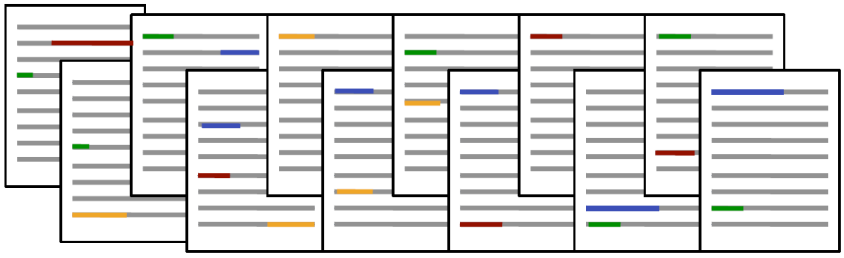
Salience Model



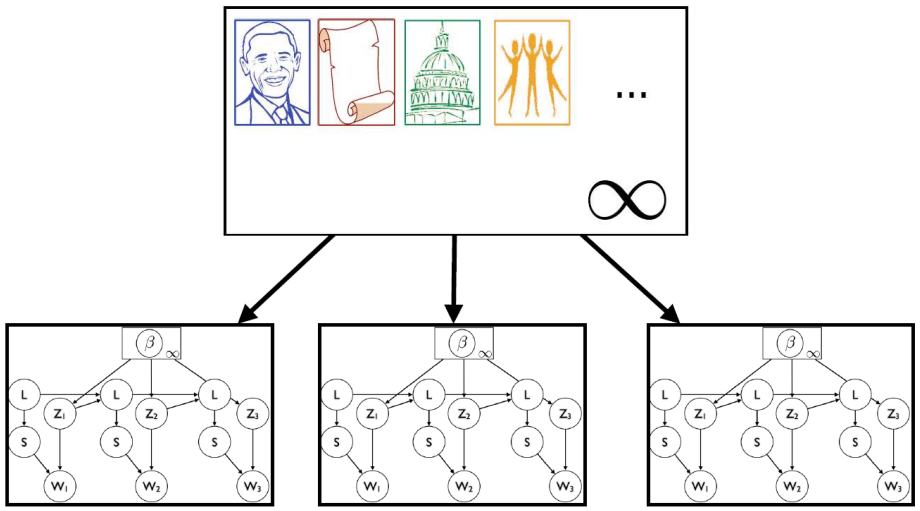
The Weir Group , whose headquarters is in the U.S is a large specialized corporation. This power plant , which , will be situated in Jiangsu, has a large generation capacity.

Cross-Document Model

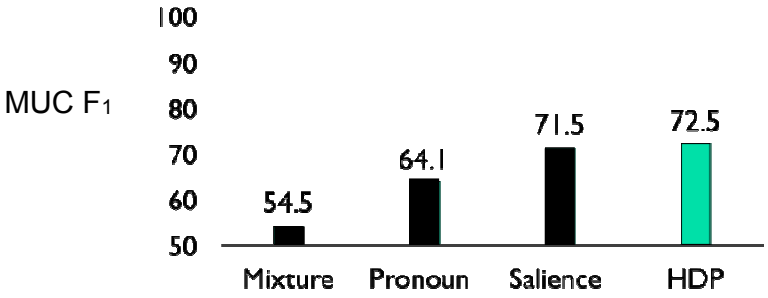




Cross-Document Model

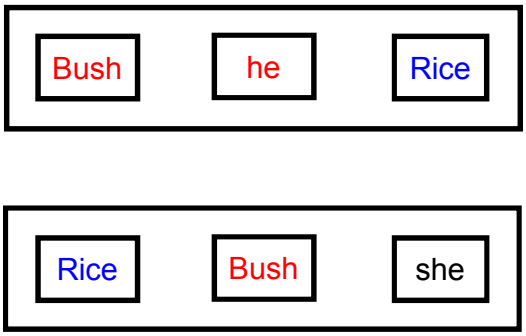


HDP Model



The Weir Group , whose headquarters is in the U.S is a large specialized corporation. This power plant , which , will be situated in Jiangsu, has a large generation capacity.

Global Entity Resolution



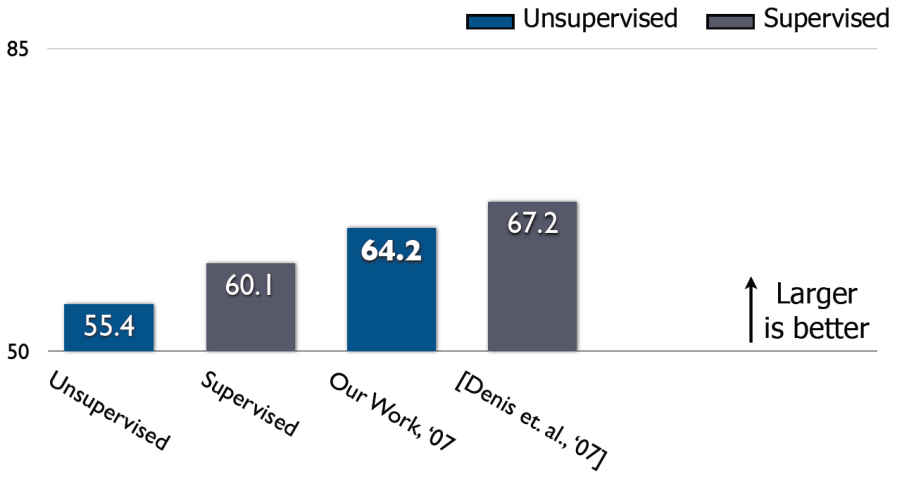
Experiments

- MUC6 English NWIRE (all mentions)
  - 53.6 F1* [Cardie and Wagstaff 99] Unsupervised
  - 70.3 F1 [Haghighi & Klein 07] Unsupervised
  - 73.4 F1 [McCallum & Wellner 04] Supervised
  - 81.3 F1 [Luo et al 04] Supervised+Oracle
- *MUC score, not all systems are completely comparable*
- *Why not: some assume oracle gives boundaries, NER types, etc.*

Evaluation

Cluster Similarity: MUC F_1

* Now system mentions!
These are "real numbers"



Errors

Input

America Online announced on Monday that the company plans to update its instant messaging service.

Correct

America Online | the company | its | instant messaging service

Guess

America Online | the company | its | instant messaging service

Adding Semantic Knowledge

America Online ⟷ **company**

America Online, LLC (commonly known as **AOL**) is an American global Internet services and media company operated by Time Warner. It is headquartered at 770 Broadway in Midtown Manhattan, New York City.[2][3] Founded in 1983 as **Quantum Computer Services**, it has franchised its services to companies in several nations around the world or set up international versions of its services.[4]

America Online

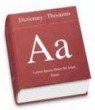
Type	Subsidiary of Time Warner
Founded	1983 as *Quantum Computer Services*

Adding Semantic Knowledge

America Online, LLC (commonly known as AOL) is an American Global Internet services and media **company**

George W. Bush, the 43rd **president**, is already the subject of more than 100 books

trade union: an **organization** whose members belong ...

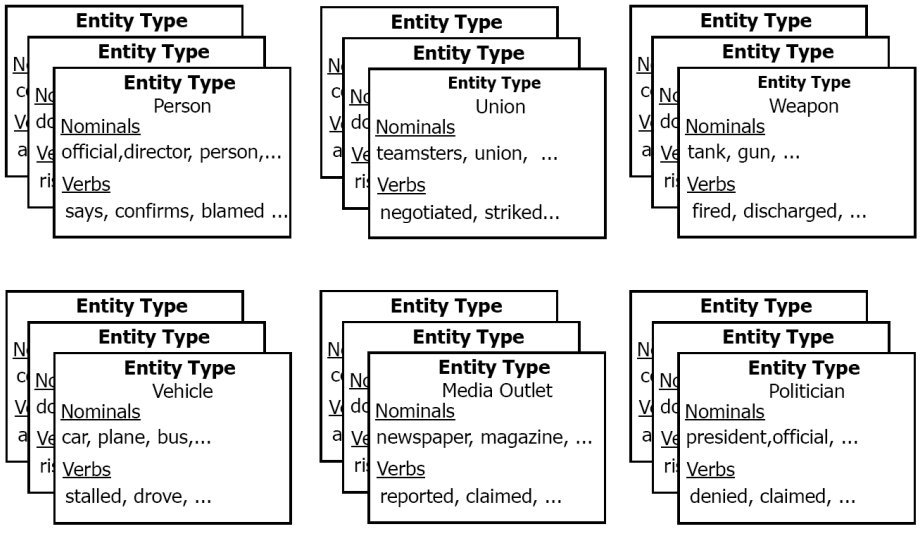


Richer Semantic Modeling

The Federal Reserve announced that the dollar dropped this week. It fell ...

The Federal Reserve announced that the dollar dropped this week. It predicted ...

Richer Semantic Modeling

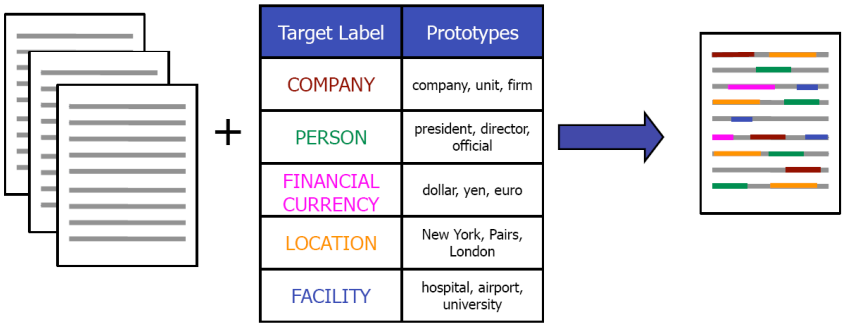


Prototype-Driven Learning

Unlabeled Data

Prototype List

Annotated Data



Target Label	Prototypes
COMPANY	company, unit, firm
PERSON	president, director, official
FINANCIAL CURRENCY	dollar, yen, euro
LOCATION	New York, Pairs, London
FACILITY	hospital, airport, university

[Haghighi & Klein, NAACL '06]

Evaluation

MUC F_1 - Cluster Similarity