

**CSCI 5832**  
**Natural Language Processing**

---

Jim Martin  
Lecture 23

4/15/08 1

---

---

---

---

---

---

---

---

**Today: 4/15**

---

- Discourse structure
  - ♦ Referring expressions
  - ♦ Co-reference resolution

4/15/08 2

---

---

---

---

---

---

---

---

**What Makes a Discourse Coherent?**

---

The reason is that these utterances, when juxtaposed, will not exhibit coherence. Almost certainly not. Do you have a discourse? Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book.

4/15/08 3

---

---

---

---

---

---

---

---

## Better?

Assume that you have collected an arbitrary set of well-formed and independently interpretable utterances, for instance, by randomly selecting one sentence from each of the previous chapters of this book. Do you have a discourse? Almost certainly not. The reason is that these utterances, when juxtaposed, will not exhibit coherence.

4/15/08

4

---

---

---

---

---

---

---

---

## What makes a text coherent?

- Appropriate use of coherence relations between subparts of the discourse -- rhetorical structure
- Appropriate sequencing of subparts of the discourse -- discourse/topic structure
- Appropriate use of referring expressions

4/15/08

5

---

---

---

---

---

---

---

---

## Referring Expressions

- Referring expressions provide a kind of glue that makes texts cohere.

4/15/08

6

---

---

---

---

---

---

---

---

## Referring Expressions: Definition

- Referring expressions are words or phrases, the *semantic interpretation* of which is a *discourse entity* (also called referent)
  - ♦ Discourse entities are *semantic objects* and they can have multiple *syntactic realizations* within a text

4/15/08

7

---

---

---

---

---

---

---

---

## NY Times Example

A college student accused of faking her own kidnapping last month was charged Wednesday with lying to police in what they suggested was a desperate attempt to get her boyfriend's attention. Audrey Seiler, a 20-year-old sophomore at the University of Wisconsin, was charged with two misdemeanor counts of obstructing officers. Each charge carries up to nine months in jail and a \$10,000 fine.

Seiler disappeared from her off-campus apartment March 27 without her coat or purse. She was discovered curled in a fetal position in a marsh four days later, and told police that a man had abducted her at knifepoint.

But police concluded Seiler made up the story after obtaining a store videotape that showed her buying the knife, duct tape, rope and cold medicine she claimed her abductor used to restrain her. Seiler confessed after she was confronted with the tape, according to authorities.

4/15/08

8

---

---

---

---

---

---

---

---

## Referring Expressions: Example

A pretty woman entered the restaurant. She sat at the table next to mine and only then I recognized her. This was Amy Garcia, my next door neighbor from 10 years ago. The woman has totally changed! Amy was at the time shy...

4/15/08

9

---

---

---

---

---

---

---

---

### Pronouns vs. Full NPs

A pretty woman entered the restaurant. She sat at the table next to mine and only then I recognized her. This was Amy Garcia, my next door neighbor from 10 years ago. The woman has totally changed! Amy was at the time shy...

4/15/08

10

---

---

---

---

---

---

---

---

### Definite vs. Indefinite NPs

A pretty woman entered the restaurant. She sat at the table next to mine and only then I recognized her. This was Amy Garcia, my next door neighbor from 10 years ago. The woman has totally changed! Amy was at the time shy...

4/15/08

11

---

---

---

---

---

---

---

---

### Common Noun vs. Proper Noun

A pretty woman entered the restaurant. She sat at the table next to mine and only then I recognized her. This was Amy Garcia, my next door neighbor from 10 years ago. The woman has totally changed! Amy was at the time shy...

4/15/08

12

---

---

---

---

---

---

---

---

## Modified vs. Bare head NP

A pretty woman entered the restaurant. She sat at the table next to mine and only then I recognized her. This was Amy Garcia, my next door neighbor from 10 years ago. The woman has totally changed! Amy was at the time shy...

4/15/08

13

---

---

---

---

---

---

---

---

## Premodified vs. postmodified

A pretty woman entered the restaurant. She sat at the table next to mine and only then I recognized her. This was Amy Garcia, my next door neighbor from 10 years ago. The woman has totally changed! Amy was at the time shy...

4/15/08

14

---

---

---

---

---

---

---

---

## More NP types

- Inferrables
  - ♦ *Sally bought a used car. The tires need to be replaced.*
- Discontinuous sets
  - ♦ *John has known Bill for many years now. They often go hiking together.*

4/15/08

15

---

---

---

---

---

---

---

---

## Break

- Quiz Thursday
- Then
  - ♦ Q/A and Summarization
    - Read Chapter 23 for Tuesday
  - ♦ MT
    - Read Chapter 25 for the following Thursday

4/15/08

16

---

---

---

---

---

---

---

---

## Quiz Areas

- From 17
  - ♦ Basics of FOL
    - Syntax and semantics
- From 18
  - ♦ Rule to rule approach
  - ♦ Semantic attachments
    - Lambdas
  - ♦ Problems with quantifiers
- From 20
  - ♦ Different approaches to WSD (20.1 to 20.5)

4/15/08

17

---

---

---

---

---

---

---

---

## Quiz Areas

- From 21
  - ♦ 21.3 to 21.8
    - Focus on Hobbs and ML approaches in 21.6.1 and 21.6.3 for pronominal resolution
- From 22
  - ♦ All

4/15/08

18

---

---

---

---

---

---

---

---

## Quiz

- Don't forget all the earlier stuff.
  - ♦ In particular, syntax/parsing plays a big role in
    - Semantic analysis
    - Co-reference
  - ♦ And it plays a supporting role in IE and WSD
    - Provides features

4/15/08

19

---

---

---

---

---

---

---

---

## Anaphora resolution

- Finding in a text all the referring expressions that have one and the same denotation
  - ♦ Pronominal anaphora resolution
  - ♦ Anaphora resolution between named entities
  - ♦ Full noun phrase anaphora resolution
  - ♦ Zero anaphora detection/resolution

4/15/08

20

---

---

---

---

---

---

---

---

## Automatic Anaphora Resolution

- Resolving the referents of pronouns
  - ♦ Hand built rule systems vs. trained systems
  - ♦ Statistical systems vs rules
  - ♦ Systems that require a full syntactic parse vs. systems that assume a weaker chunking segmentation
  - ♦ Systems that deal with text vs. systems that deal with spoken discourse

4/15/08

21

---

---

---

---

---

---

---

---

## Basic ML Approach for Pronominal Reference

- Acquire an annotated corpus
  - ♦ Pronoun/antecedent pairs constitute + examples
  - ♦ Pronoun/non-referent pairs are - examples
    - Lots more - data than + data
- Extract features from the pairs
- Train a binary classifier
  - ♦ SVMs or MaxEnt

4/15/08

22

---

---

---

---

---

---

---

---

## Basic Features

- Strict grammatical number match (1/0)
  - ♦ Not the number of the elements, just whether they match
- Compatible match
- Same 2 for gender
- Sentence distance
  - ♦ Distance from pronoun sentence to the referent
  - ♦ Hobbs distance
- Grammatical role information
- Form
  - ♦ Form of the antecedent (full NP, etc)

4/15/08

23

---

---

---

---

---

---

---

---

## Classification

- Proceed through a text linearly. When a pronoun is encountered look backward and apply the classifier to each potential antecedent
  - ♦ In some order
  - ♦ Need a way to break ties
  - ♦ Result is a set of referent chains

4/15/08

24

---

---

---

---

---

---

---

---



## Full Co-reference Analysis

- Locating all the entity mentions in a text and then clustering them into groups that represent the actual entities.

4/15/08

25

---

---

---

---

---

---

---

---

## NY Times Example

A college student accused of faking her own kidnapping last month was charged Wednesday with lying to police in what they suggested was a desperate attempt to get her boyfriend's attention. Audrey Seiler, a 20-year-old sophomore at the University of Wisconsin, was charged with two misdemeanor counts of obstructing officers. Each charge carries up to nine months in jail and a \$10,000 fine.

Seiler disappeared from her off-campus apartment March 27 without her coat or purse. She was discovered curled in a fetal position in a marsh four days later, and told police that a man had abducted her at knifepoint.

But police concluded Seiler made up the story after obtaining a store videotape that showed her buying the knife, duct tape, rope and cold medicine she claimed her abductor used to restrain her. Seiler confessed after she was confronted with the tape, according to authorities.

4/15/08

26

---

---

---

---

---

---

---

---

## NY Times Example

A college student accused of faking her own kidnapping last month was charged Wednesday with lying to police in what they suggested was a desperate attempt to get her boyfriend's attention. Audrey Seiler, a 20-year-old sophomore at the University of Wisconsin, was charged with two misdemeanor counts of obstructing officers. Each charge carries up to nine months in jail and a \$10,000 fine.

Seiler disappeared from her off-campus apartment March 27 without her coat or purse. She was discovered curled in a fetal position in a marsh four days later, and told police that a man had abducted her at knifepoint.

But police concluded Seiler made up the story after obtaining a store videotape that showed her buying the knife, duct tape, rope and cold medicine she claimed her abductor used to restrain her. Seiler confessed after she was confronted with the tape, according to authorities.

4/15/08

27

---

---

---

---

---

---

---

---

## Example

- There are
  - ♦ 33 entity mentions
  - ♦ 16 actual entities
  - ♦ Seiler accounts for 12 of the 33 mentions, with the police in second place at 6
  - ♦ There are 11 entities with 1 mention each

4/15/08

28

---

---

---

---

---

---

---

---

## 2 Basic ML Approaches

- Scan the text from the top. When a new entity mention is encountered apply a binary classifier to each previous mention.
  - ♦ Result is a set of chains of references
- Extract all entity mentions and represent them as feature vectors. Apply an unsupervised partition-based clustering technique to all the mentions.
  - ♦ Result is a set of (soft) clusters that partition the mentions.

4/15/08

29

---

---

---

---

---

---

---

---

## Cross-Document Co-Reference

- Take a set of documents (possibly from the Web), extract all the named referents and resolve them into unique entities.
  - ♦ I.e. Google "Michael Jordan" and resolve the results into classes based on the real entities.

4/15/08

30

---

---

---

---

---

---

---

---

# Easy Money

**THE SPOCK CHALLENGE**

One Challenging Problem. One Compelling Prize (\$50,000)

At Spock, we love finding and meeting new people. It's our business.

**What is Spock?**  
Spock, an industry leading people search application, helps users find and discover people on the web. With over one hundred million individuals indexed and millions added every day, Spock is the largest and most comprehensive people specific search application.

**Our Technology**  
At the core, we organize relevant information around people and have developed unique technologies to do so. Not only is this a very fun product for us and our users, we are also fortunate enough to be working on some of the most interesting problems in computer science!

**The Challenge**  
To improve our technology and to create a better user experience, we decided to share the fun! We have selected one of our most interesting problems, namely Entity Resolution, to share with the community, allowing other leading computer scientists and engineers to compete in an open contest. The winners of this global competition will reap a handsome reward, and perhaps even employment at Spock.

You can work individually and in teams. The competition will last 4 months and the winning team will win a Grand Prize of \$50,000. Most important, you'll be working on a very important and widely applicable problem. We will also be paying prizes for 2nd and 3rd place.

[Click here to learn more about the Entity Resolution and Extraction Problem](#)

[About Spock](#)   [Rules](#)   [Discussion](#)   [FAQ](#)   [Jobs at Spock](#)

4/15/08 31

---

---

---

---

---

---

---

---

---

---

# Easy Money

- 100,000 files provided for development
- Set of target names of interest
- Ground truth provided with a F-measure based scoring metric
  - ◆ Precision/Recall computed on a pairwise document basis.
  - ◆ Assumes that target names are unique within each document (ie. All the *Michael Jordan* mentions in a single document are the same person).

4/15/08 32

---

---

---

---

---

---

---

---

---

---