

CSCI 5832

Natural Language Processing

Lecture 24
Jim Martin

4/24/07

CSCI 5832 Spring 2007

1

Today: 4/19

- **Discourse structure**

4/24/07

CSCI 5832 Spring 2007

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.

4/24/07

CSCI 5832 Spring 2007

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.

4/24/07

CSCI 5832 Spring 2007

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**

4/24/07

CSCI 5832 Spring 2007

5

Rhetorical Structure Theory

- One theory of discourse structure, based on identifying relations between segments of the text
 - **Nucleus/satellite** notion encodes asymmetry
 - Some **rhetorical relations**:
 - **Elaboration** (set/member, class/instance, whole/part...)
 - **Contrast**: multinuclear
 - **Condition**: Sat presents precondition for N
 - **Purpose**: Sat presents goal of the activity in N

4/24/07

CSCI 5832 Spring 2007

6

Relations

- **A sample definition**
 - Relation: **evidence**
 - Constraints on N: H might not believe N as much as S think s/he should
 - Constraints on Sat: H *already believes or will believe* Sat
- **An example:**
 - The governor supports big business.**
 - He is sure to veto House Bill 1711.**

4/24/07

CSCI 5832 Spring 2007

7

Automatic Rhetorical Structure Labeling

- **Same old story by now...**
 - Get a group of annotators to assign a set of RST relations to a text
 - Extract a set of surface features from the text that might signal the presence of the rhetorical relations in that text
 - Train a supervised ML system based on the training set

4/24/07

CSCI 5832 Spring 2007

8

Features

- Explicit markers: *because, however, therefore, then, etc.*
- Tendency of certain syntactic structures to signal certain relations: **Infinitives are often used to signal purpose relations:** *Use rm to delete files.*
- Ordering
- Tense/aspect
- Intonation

4/24/07

CSCI 5832 Spring 2007

9

Some Problems with RST

- How many Rhetorical Relations are there?
- How can we use RST in dialogue as well as monologue?
- RST does not model overall structure of the discourse.
- Difficult to get annotators to agree on labeling the same texts

4/24/07

CSCI 5832 Spring 2007

10

Break

- **Final project format.**
 - Full deliverable varies based on your project
 - But all projects will need to produce an 8-page conference style paper.
 - I'll provide LaTeX and MS Word templates for you to use from the ACM.
 - This paper should be self-contained. As with a program committee, everything needed to assess the project should be in the paper.

Referring Expressions

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

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

4/24/07

CSCI 5832 Spring 2007

13

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.

4/24/07

CSCI 5832 Spring 2007

14

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...

4/24/07

CSCI 5832 Spring 2007

15

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...

4/24/07

CSCI 5832 Spring 2007

16

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...

4/24/07

CSCI 5832 Spring 2007

17

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...

4/24/07

CSCI 5832 Spring 2007

18

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...

4/24/07

CSCI 5832 Spring 2007

19

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...

4/24/07

CSCI 5832 Spring 2007

20

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.*

4/24/07

CSCI 5832 Spring 2007

21

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**

4/24/07

CSCI 5832 Spring 2007

22

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

4/24/07

CSCI 5832 Spring 2007

23

Full Co-reference Analysis

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

4/24/07

CSCI 5832 Spring 2007

24

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.

4/24/07

CSCI 5832 Spring 2007

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.

4/24/07

CSCI 5832 Spring 2007

26

Example

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

4/24/07

CSCI 5832 Spring 2007

27

Basic ML Approaches

- **Scan the text from the top. When a new entity mention is encountered apply a binary classifier to each previous mention (refers or not).**
 - **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.**

4/24/07

CSCI 5832 Spring 2007

28

Cross-Document Co-Reference

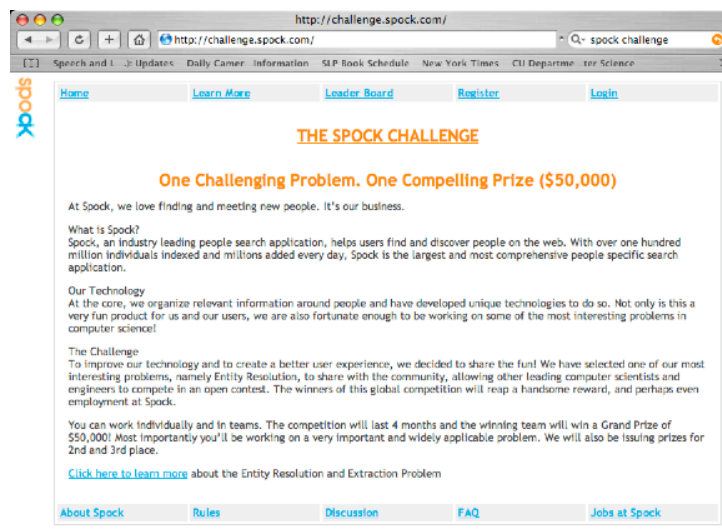
- Take a set of documents (possibly from the Web), extract all the named referents and resolve them into unique entities.
 - Ie. Google "Michael Jordan" and resolve the results into classes based on the real entities.

4/24/07

CSCI 5832 Spring 2007

29

Easy Money



The screenshot shows a web browser window with the URL <http://challenge.spock.com/>. The page features a navigation bar with links for Home, Learn More, Leader Board, Register, and Login. The main heading is "THE SPOCK CHALLENGE" in orange. Below this, the text reads "One Challenging Problem. One Compelling Prize (\$50,000)". The page describes Spock as a people search application and details the challenge: to improve Entity Resolution technology. It offers a Grand Prize of \$50,000 and mentions that the competition will last 4 months. A link is provided to learn more about the Entity Resolution and Extraction Problem. At the bottom, there are links for About Spock, Rules, Discussion, FAQ, and Jobs at Spock.

4/24/07

CSCI 5832 Spring 2007

30

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).

4/24/07

CSCI 5832 Spring 2007

31

Next Time

- Machine Translation
 - Read Chapter 24

4/24/07

CSCI 5832 Spring 2007

32