Lecture 22: Discourse and Referring Expressions

Julia Hockenmaier
juliahmr@illinois.edu
3324 Siebel Center
Part 1: 
Brief Introduction to Discourse
What we’ve covered so far

**Lexical Semantics** (meaning of words)
We’ve mostly focused on content words
(nouns, verbs, adjectives)

**Compositional Semantics** (meaning of sentences)
— Principle of compositionality:
The meaning of sentences depends recursively
(compositionally) on the meaning of their words and
constituents.
— Logically, declarative sentences correspond to
propositions that can either be true or false.
On Monday, John went to Einstein’s. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.

‘Discourse’: Any linguistic unit that consists of multiple sentences

Speakers describe “some situation or state of the real or some hypothetical world” (Webber, 1983)

Speakers attempt to get the listener to construct a similar model of the situation.
Why study discourse?

For natural language **understanding**: 
Most information is not contained in a single sentence. 
The system has to **aggregate** information across sentences, paragraphs or entire documents.

For natural language **generation**: 
When systems generate text, that text needs to be easy to understand — it has to be **coherent**.
**What makes text coherent?**
How can we understand discourse?

On Monday, John went to Einstein’s. He wanted to buy lunch. But the cafe was closed. That made him angry, so the next day he went to Green Street instead.

Understanding discourse requires (among other things):

1) doing **coreference** resolution:
   ‘*the cafe*’ and ‘*Einstein’s*’ refer to the same entity
   *He* and *John* refer to the same person.
   *That* refers to ‘*the cafe was closed*’.

2) identifying **discourse (‘coherence’) relations**:
   ‘*He wanted to buy lunch*’ is the **reason** for
   ‘*John went to Bevande*.’
Discourse models

An explicit representation of:

— the **entities, events and states** that a discourse talks about
— the **relations** between them (and to the real world).

This representation is often written in some form of logic.

What does this logic need to capture?
Discourse models should capture...

**Entities** (physical or abstract):  
John, Einstein’s, lunch, hope, computer science, …

**Eventualities** (events or states):  
— **Events**: On Monday, John went to Einstein’s  
  involve entities, take place at a point in time  
— **States**: It was closer. Water is a liquid.  
  involve entities and hold for a period of time (or are generally true)

**Temporal relations** between events/states  
afterwards, during,

**Rhetorical** (‘discourse’) **relations** between propositions  
so, instead, if, whereas
Part 2: Referring expressions
How do we refer to entities?


‘the book’

‘this book’

‘a book’

‘my book’

‘that one’

‘the book I’m reading’

‘it’
Some terminology

Referring expressions (‘this book’, ‘it’) refer to some entity (e.g. a book), which is called the referent.

Co-reference: two referring expressions that refer to the same entity co-refer (are co-referent).

*I saw a movie last night. I think you should see it too!*

The referent is evoked in its first mention, and accessed in any subsequent mention.
Indefinite NPs

No determiner:  I like walnuts.
Indefinite determiner: She sent her a beautiful goose
Numerals:  I saw three geese.
Indefinite quantifiers: I ate some walnuts.
(Indefinite) this: I saw this beautiful Ford Falcon today

Indefinite NPs usually introduce a new discourse entity.

They can refer to a specific entity or not:

I’m going to buy a computer today.

(unclear if the speaker has a particular computer in mind (e.g. his friends’ old computer), or just any computer)
Definite NPs

The definite article (*the book*),
Demonstrative articles (*this/that book, these/those books*),
Possessives (*my/John’s book*)

Definite NPs can also consist of

Personal pronouns (*I, he*)
Demonstrative pronouns (*this, that, these, those*)
Universal quantifiers (*all, every*)
(unmodified) proper nouns (*John Smith, Mary, Urbana*)

Definite NPs refer to an identifiable entity
(previously mentioned or not)
Information status

Every entity can be classified along two dimensions:

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

Speaker assumes entity is (un)known to the hearer

- **Hearer-old**: *I will call Sandra Thompson.*
- **Hearer-new**: *I will call a colleague in California* (=Sandra Thompson)

**Special case of hearer-old**: hearer-inferrable

*I went to the student union. *The food court was really crowded.*

**Discourse-new vs. discourse-old:**

Speaker introduces new entity into the discourse, or refers to an entity that has been previously introduced.

- **Discourse-old**: *I will call her/Sandra now.*
- **Discourse-new**: *I will call my friend Sandra now.*
Anaphoric pronouns refer back to some previously introduced entity/discourse referent:

John showed Bob his car. He was impressed.

John showed Bob his car. This took five minutes.

The antecedent of an anaphor is the previous expression that refers to the same entity.

There are number/gender/person agreement constraints: girls can’t be the antecedent of he.

Usually, we need some form of inference to identify the antecedents.
Salience/Focus

Only some recently mentioned entities can be referred to by pronouns:

John went to Bob’s party and parked next to a classic Ford Falcon.
He went inside and talked to Bob for more than an hour.
Bob told him that he recently got engaged.
He also said he bought it (???) / the Falcon yesterday.

Key insight (also captured in Centering Theory)
Capturing which entities are salient (in focus) reduces the amount of search (inference) necessary to interpret pronouns!
Part 3: Coreference resolution
The coreference resolution task

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 **Megabucks** from rival **Lotsabucks**.

Return **Coreference Chains**

(sets of mentions that refer to the same entities)

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}
Special case: Pronoun resolution

Task: Find the antecedent of an anaphoric pronoun in context

1. *John saw a beautiful Ford Falcon at the dealership.*
2. *He showed it to Bob.*
3. *He bought it.*

\[ \text{he}_2, \text{it}_2 = \text{John, Ford Falcon, or dealership?} \]
\[ \text{he}_3, \text{it}_2 = \text{John, Ford Falcon, dealership, or Bob?} \]
Coref as binary classification

Represent each NP-NP pair (+context) as a feature vector.

**Training:**
Learn a binary classifier to decide whether NP$_i$ is a possible antecedent of NP$_j$.

**Decoding** (running the system on new text):
— Pass through the text from beginning to end
— For each NP$_i$:
  Go through NP$_{i-1}$...NP$_1$ to find best antecedent NP$_j$.
  Corefer NP$_i$ with NP$_j$.
  If the classifier can’t identify an antecedent for NP$_i$, it’s a new entity.
Example features for Coref resolution

What can we say about each of the two NPs?
Head words, NER type, grammatical role, person, number, gender, mention type (proper, definite, indefinite, pronoun), #words, …

How similar are the two NPs?
— Do the two NPs have the same head noun/modifier/words?
— Do gender, number, animacy, person, NER type match?
— Does one NP contain an alias (acronym) of the other?
— Is one NP a hypernym/synonym of the other?
— How similar are their word embeddings (cosine)?

What is the likely relation between the two NPs?
— Is one NP an appositive of the other?
— What is the distance (#sentences, #words, #mentions) between the two NPs?
Lee et al.’s neural model for coref resolution

Joint model for mention identification and coref resolution:
Use word embeddings + LSTM to get a vector \( g_i \) for each span \( i \)
\( i = \text{START}(i) \ldots \text{END}(i) \) in the document (up to a max. span length \( L \))
Use \( g_i \) + neural net \( \text{NN}_m \) to get a mention score \( m(i) \) for each \( i \)
(used to identify most likely mention spans at inference time)
Use \( g_i, g_j \) + \( \text{NN}_c \) to get antecedent scores \( c(i,j) \) for all span pairs \( i, j<i \)
Compute overall score \( s(i,j) = m(i)+m(j)+c(i,j) \) for all span pairs \( i,j<i \)
and set overall score \( s(i,\epsilon) = 0 \) [score for \( i \) being discourse-new]
Identify the most likely antecedent for each span \( i \) according to
\[
y_i^* = \arg\max_{y_i \in \{1, \ldots i-1, \epsilon\}} P(y_i) \quad \text{with} \quad P(y_i) = \frac{\exp(s(i, y_i))}{\sum_{y' \in \{1, \ldots i-1, \epsilon\}} \exp(s(i, y'))}
\]
Perform a forward pass over all (most likely) spans
to identify their most likely antecedents
Lee et al.’s neural model for coref resolution

**Span representation** \( g_i \):
Computed by a biLSTM over word embeddings:
- LSTM’s hidden state of i’s first word,
- LSTM’s hidden state of i’s last,
- weighted avg of word embeddings in span i; length of span
\[
[h_{\text{START}}(i), h_{\text{END}}(i), h_{\text{ATT}}(i), \varphi(i)]
\]

**Scoring function** \( s(i,j) \):
- a) for \( j=\epsilon \) (i has no antecedent): \( s(i,\epsilon) = 0 \)
- b) for \( j\neq\epsilon \): \( s(i,j) = m(i) + m(j) + c(i,j) \)
  - \( m(i) \): is span i a mention?
    - binary classifier (feedforward net) with \( g_i \) as input
  - \( c(i,j) \): is j an antecedent of i?
    - input: \( g_i, g_j, g_i \circ g_i \) [element-wise multiplication]
Evaluation metrics for coref resolution

Compare hypothesis H against (gold) reference R by:

MUC score:
— Precision/Recall over #coref links
— Ignores singleton mentions
— Rewards long coref chains/clusters

B³ score:
— Precision/Recall over mentions in same cluster
— May count same mention multiple times

CEAF score:
— Precision/Recall, based on mention alignments

CoNLL F1: combines MUC, B³, CEAF

Challenge: How to handle predicted mentions (whose span may differ from gold mentions)?
The importance of world knowledge

Coreference resolution often needs world (“commonsense”) knowledge.

Compare:

The city councilmen refused the demonstrators a permit because they feared violence.

The city councilmen refused the demonstrators a permit because they advocated violence.

CF: The Winograd Schema Challenge
https://cs.nyu.edu/faculty/davise/papers/WinogradSchemas/WS.html
World knowledge may capture bias

Preferred attachments (both by humans and systems) often reflect stereotypes (e.g. about occupations and gender)

A man and his son get into a terrible car crash. The father dies, and the boy is badly injured. In the hospital, the surgeon looks at the patient and exclaims, “I can’t operate on this boy, he’s my son!”

https://www.aclweb.org/anthology/N18-2002/