CS 498 JH: **Introduction to NLP** (Fall '08)

Lecture 22: Coreference resolution

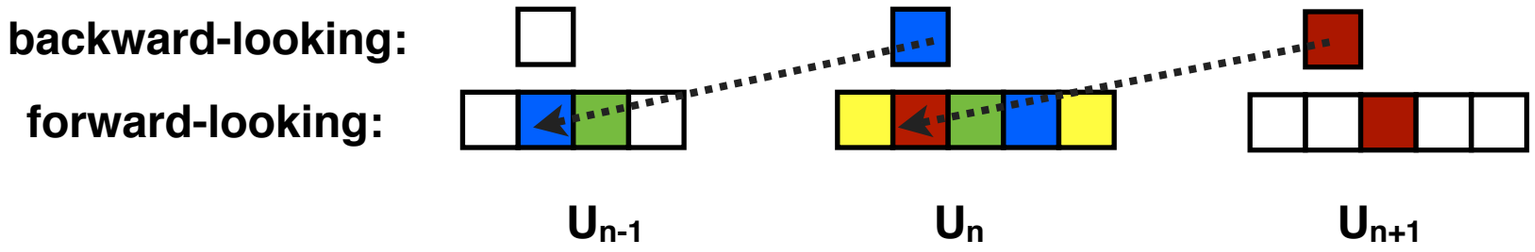
Julia Hockenmaier
juliahmr@illinois.edu
3324 Siebel Center
Office Hours: Tue, 2:00-3:30pm

http://www.cs.uiuc.edu/class/fa08/cs498jh

Quick revision:

Centering Theory

The two centers of an utterance



The **forward looking center** of an utterance is a **partially ordered set of entities** that are mentioned in the utterance.

The **ordering** reflects salience within the current utterance:

subject > direct object > object,….

The **backward looking center** of an utterance is the **highest ranked element** of the **forward looking center of the previous utterance** that is mentioned in the current utterance

Example: Coherent discourse

John went to **his** **favorite music store** to buy **a piano**.

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'* }

He *had frequented* ***the store*** *for many years.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *store'* }

He *was excited that* ***he*** *could finally buy* ***a piano****.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *piano'* }

He *arrived just as* ***the store*** *was closing for the day.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *store'* }

Example: incoherent discourse

John *went to* ***his favorite music store*** *to buy* ***a piano****.*

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'* }

It *was* ***a store John*** *had frequented for many years.*

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }

He *was excited that* ***he*** *could finally buy* ***a piano****.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *piano'* }

It *was closing just as* ***John*** *arrived.*

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }

Center realization and pronouns

Constraint/Rule 1:

If *any* element of FW(S_{n-1}) is realized as **a pronoun** in S_n,
then the **BW(S_n)** has to be realized **as a pronoun** in S_n as well

Example of a violation:

Susan *gave* ***Betsy*** ***a hamster****.*

She *told* ***her*** *to feed* ***the hamster*** *well.*

BW=Susan, **FW**={Susan, Betsy, hamster}

She *asked* ***Susan*** *what to feed* ***him****.*

BW=Susan, **FW**= {Betsy, Susan, hamster}

Violation: *hamster'* is realized as pronoun, but *Susan'* isn't.

Transitions between sentences

Center continuation:

$BW(S_n) = BW(S_{n-1})$ and $BW(S_n)$ = highest ranked element in $FW(S_n)$

Susan *gave* ***Betsy*** ***a hamster.***

She *told* ***her*** *to feed* ***it*** *well.* **BW=Susan**, **FW=**{Susan, Betsy, hamster}

She *asked* ***her*** *whether she liked the gift.* **BW=Susan**, **FW=**{**Susan**, Betsy, hamster}

Center retaining:

$BW(S_n) = BW(S_{n-1})$ but $BW(S_n) \neq$ highest ranked element in $FW(S_n)$

Susan *gave* ***Betsy*** ***a hamster.***

She *told* ***her*** *to feed* ***it*** *well.* **BW=Susan**, **FW=**{Susan, Betsy, hamster}

Betsy *asked* ***her*** *what to feed him.* **BW=Susan**, **FW=**{**Betsy, Susan** hamster}

Center shifting:

$BW(S_n) \neq BW(S_{n-1})$

Susan *gave* ***Betsy*** ***a hamster.***

She *told* ***her*** *to feed* ***it*** *well.* **BW=Susan**, **FW=**{Susan, Betsy, hamster}

The hamster *was very cute.* **BW=hamster, FW=**{hamster}

Rough vs. smooth shifts

Smooth shift:

$BW(S_n) \neq BW(S_{n-1})$ and $BW(S_n)$ = highest in $FW(S_n)$

Susan *gave* ***Betsy a hamster.***

Betsy *thanked* ***Susan****.* **BW**=**Susan**, **FW**={*Betsy*, Susan}

This *was* ***her*** *favorite gift.* **BW**=*Betsy* **FW**={hamster, *Betsy*}

Rough shift:

$BW(S_n) \neq BW(S_{n-1})$ and $BW(S_n) \neq$ highest in $FW(S_n)$

Susan *gave* ***Betsy a hamster.***

She *told* ***her*** *to feed* ***it*** *well.* **BW**=**Susan**, **FW**={Susan, Betsy, hamster}

John *said* ***the hamster*** *was very cute.* **BW**=**hamster,** **FW**={John, hamster}

Preferred transitions

Rule/Constraint 2:

- Sequences of center continuation are preferred over sequences of center retaining.
- Sequences of center retaining are preferred over sequences of center shifting.
- Sequences of smooth shifting are preferred over sequences of rough shifting

Local coherence is achieved by maximizing the number of center continuations.

Transitions: Summary

	$BW(S_{n+1}) = BW(S_n)$ or $BW(S_n)$ undef.	$BW(S_{n+1}) \neq BW(S_n)$
$BW(S_{n+1})$ = highest in $FW(S_n)$	Center continuation	Smooth shift
$BW(S_{n+1}) \neq$ highest in $FW(S_n)$	Center retaining	Rough shift

Preference over sequences of transitions:

Continuation > Retaining > Smooth Shift > Rough Shift

Example: Coherent discourse

John went to **his** **favorite music store** to buy **a piano**.

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'* }

He *had frequented* ***the store*** *for many years.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *store'* }

He *was excited that* ***he*** *could finally buy* ***a piano****.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *piano'* }

He *arrived just as* ***the store*** *was closing for the day.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *store'* }

Continuation

Continuation

Example: incoherent discourse

John *went to* ***his favorite music store*** *to buy* ***a piano****.*

backward-looking center: ? (no previous discourse)

forward-looking center: {*John'*, *store'*, *piano'* }

It *was* ***a store John*** *had frequented for many years.*

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }

He *was excited that* ***he*** *could finally buy* ***a piano****.*

backward-looking center: {*John'* }

forward-looking center: {*John'*, *piano'* }

It *was closing just as* ***John*** *arrived.*

backward-looking center: {*John'* }

forward-looking center: {*store'*, *John'* }

Continuation

Retention

More on Centering

There are different variants of Centering in the literature:

- Different versions of therules and constraints
- Different definitions of utterance and salience.

Poesio et al. (2004) tested different instantiations of centering against a corpus. Some of their conclusions:

- Choosing the right instantiation of Centering matters
- Centering's account of pronominalization is robust
- Entity-based accounts alone are not sufficient to capture local coherence
- Variety (e.g. not always using pronouns) matters too

Pronominal anaphora resolution

Pronouns: reminder

Anaphoric pronouns are a form of definite reference:

Emma smiled and chatted as cheerfully as she could.

Some pronouns are **bound by quantifiers:**
(they behave like bound variables)

Every dancer brought her left arm forward.

Reflexive pronouns are also **bound:**

Mary looked at herself in the mirror.

(compare: Mary looked at her in the mirror.)

Pronoun resolution

What is 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.*

He$_3$ = John or Bob?

Pronominal anaphora resolution constraints

- **Number/Gender/Person agreement:**
  girls can't be the antecedent of *he*

- **Binding theory constraints**: Reflexive pronouns (himself/myself) have to refer to the same entity as the subject of the sentence. Personal pronouns refer to a different entity as the subject:

John saw himself in the mirror
John saw him in the mirror (i.e. not John)

Anaphora resolution preferences

Recency:

The doctor found an old map. Jim found an even older map. It described an island.

Grammatical roles determine salience

Salience hierarchies: Subjects > direct objects > indirect objects

More salient entities are more likely to be referred to again.

Repeated mention:

If an entity has been mentioned several times, it is more likely to be referred to again

Parallelism:

John went with Jim to Paris. Bill went with him to Rome.

Verb semantics and plausible rhetorical relations:

John called Bill. He lost the laptop.

John criticized Bill. He lost the laptop.

Selectional preferences:

John parked his car in the garage after driving it around for hours

A centering-based pronoun resolution algorithm

Centering-based pronoun resolution

Rules:

- If any element in FW(S_n) is realized by a pronoun in Sn+1, then BW(S_{n+1}) must also be realized as a pronoun.
- Transition states are ordered. Prefer continuation over retaining over smooth shift over rough shift.

The algorithm:

For each utterance:

- Generate possible FW/BW centers
- Filter by constraints (agreement etc.) and above rules
- Rank the resulting assignments by transition orderings

An example

1. John saw a beautiful Ford Falcon at the dealership.

FW = {John, Ford, dealership} BW = undefined

2. He showed it to Bob.

Reading 1: He=John, it=Ford

*FW = {**John**, Ford, Bob} BW = John Transition: continue*

Reading 2: He=John, it=dealership

*FW = {**John**, dealership, Bob} BW = John Transition: continue*

There's a tie! Assume reading 1 (it=Ford)

3. He bought it.

Reading 1: *He = John*

FW = {John, Ford}, BW = John Transition: continue

Reading 2: *He = Bob*

FW = {Bob, Ford}, BW = Bob Transition: shift

Prefer reading 1 over reading 2 (he=John)

What does this capture?

This algorithm captures preferences for

- grammatical roles (salience hierarchy)

This preference can be overcome if it leads to a bad transition

- recency
- repeated mention

But:

Centering alone does not capture antecedents that occur in the same sentence as the pronoun:

Mary likes her coffee black.

President-elect Obama and his aides…

John claims that he did his homework alone.

Pronoun resolution with a classifier

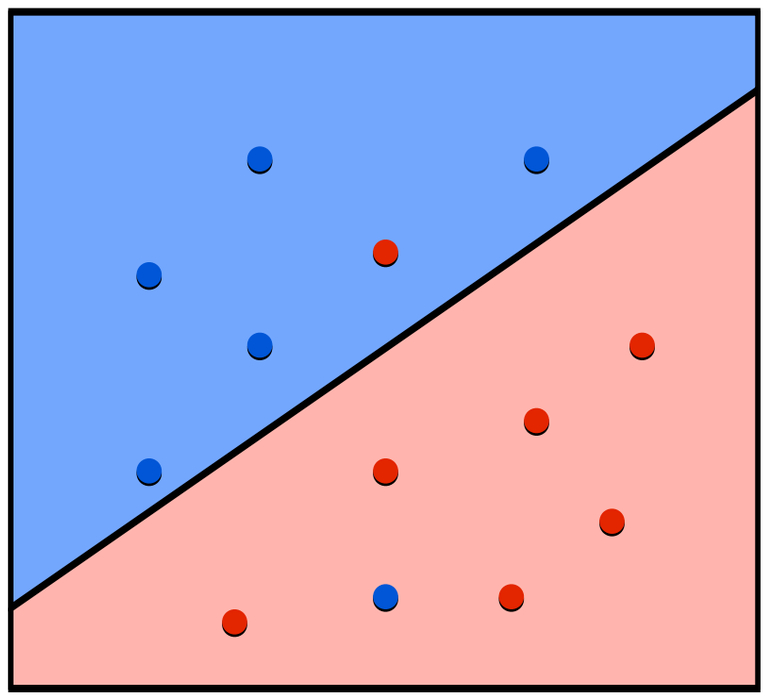
Pronoun resolution as classification

- Rephrase anaphora resolution as **binary classification:**

"Is NP …, which appears in context … an antecedent for the pronoun … ?"

- Represent NP(+context)-pronoun pairs as **feature vectors**
- Training data is labeled: NPs+context+pronoun + YES/NO
- Learn classifier to predict YES/NO for input feature vectors

When the classifier output is a number (e.g. between 0...1), define a threshold (e.g. 0.5) above which the answer is yes (and below which it is no).



Features

- Lexical features

The pronoun and the words (or the head noun) in the antecedent

- Number agreement of antecedent and pronoun
- Gender agreement of antecedent and pronoun
- Distance of antecedent to pronoun:

Number of sentences,
'Hobbs' distance (~ number of possible antecedents in parse tree)

- Grammatical role of antecedent (subject, object, PP)

More about pronouns

Non-referential uses of pronouns/demonstratives

Expletive *it* **and** *there:*

It is raining
It seems that I'm late.
There's a shirt over there.
It's you that I was waiting for (cleft)

Note: a good diagnostic to distinguish expletive from referential it:
(from Santorini&Kroch's syntax textbook)
Expletive 'it' can't be used as an answer. It can't even be asked for:
*"*What seems that I'm late? --*it.*

Non-physical entities as referents of pronouns

According to Doug, Sue just bought a 1961 Ford Falcon.

- *But <u>that</u> turned out to be a lie.*
- *But <u>that</u> was false.*
- *<u>That</u> struck me as a funny way to describe the situation.*
- *<u>That</u> caused a financial problem for Sue.*

(Webber, 1991)

Coreference resolution

The 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.

The 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.

Coreference chains:

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}

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:

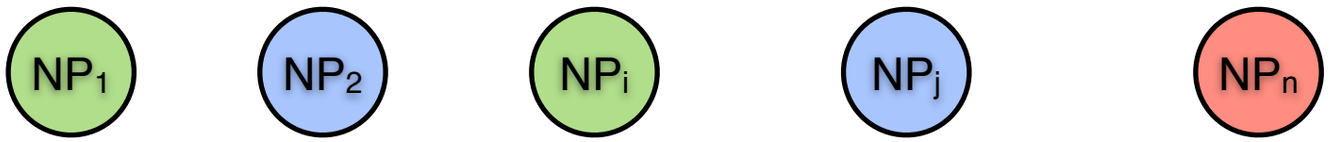
  Go through NP_{i-1}...NP_1 to find best antecedent NP_j.
  Corefer NP_i with NP_j.
  If the classifier can't identify an antecedent for NP_i, it's a new entity.

The entity-mention graph (Bengston/Roth '08)

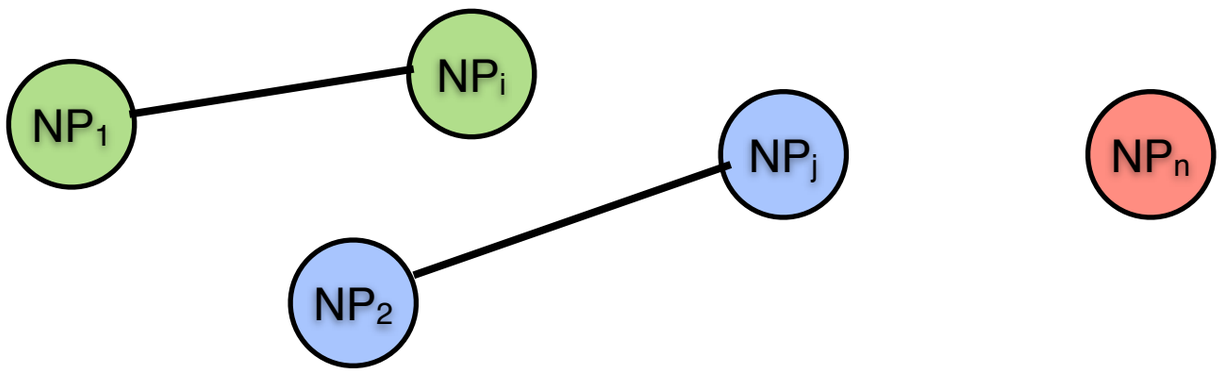
Assume a **pairwise coreference classifier** *pc(NP$_i$, NP$_j$)*

Construct an **entity-mention graph**:

- Every NP ('mention') + context in the document is a vertex



- Using *pc(NP$_i$, NP$_j$)*, generate a set of edges that partitions the graph such that each connected component corresponds to one entity: If *pc(NP$_i$, NP$_j$) > threshold*, add edge:



Features for Coref resolution

- Do the two NPs have the same **head noun**?
  (e.g. company)
- Do they contain the **same modifier**?
  (e.g. Denver-based)?
- Does the **gender** and **number** of the NPs match?
- Does one NP contain an alias (**acronym**) of the other?
  (United States = USA, Chief Executive Office = CEO)
- Is one NP a **hypernym/synonym** of the other?
- Is one NP an **appositive** of the other?
  [Victoria Chen], [CFO of Megabucks]
- Are both NPs **named entities** of the same type?

[CEO] = PERSON, Victoria Chen = PERSON

Evaluation: B-cubed F-score

The test data consists of *D* documents *d* with *N* total mentions *m* (mention boundaries are given as input)

- In the **gold standard**, each mention *m* belongs to a **'true' cluster** of mentions (=connected component) of size t_m
- In the **system output**, each mention *m* belongs to a **predicted cluster** of mentions (=connected component) of size p_m
- For each mention *m*, the **intersection** of the gold standard and system output clusters defines a **common cluster** of mentions of size c_m

$$\text{Precision } P = \frac{1}{N} \sum_{d \in D} \sum_{m \in d} \frac{c_m}{p_m}$$

$$\text{Recall } R = \frac{1}{N} \sum_{d \in D} \sum_{m \in d} \frac{c_m}{t_m}$$

$$\text{F-measure} = \frac{2PR}{P+R}$$