

## Introduction to Human Language Technologies

### 10. Coreference resolution



UNIVERSITAT POLITÈCNICA DE CATALUNYA  
BARCELONATECH

---

Facultat d'Informàtica de Barcelona



# Outline

# Outline

# The goal of coreference resolution

**Determining which mentions in a discourse refer to the same real world entity, property or situation.**

Example:

FC Barcelona president Joan Laporta has warned Chelsea off star strike Lionel Messi.

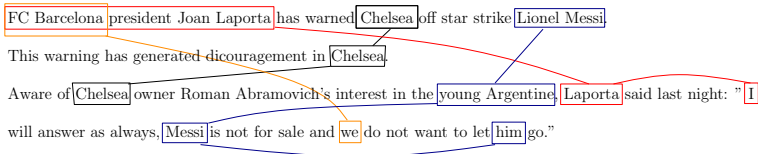
This warning has generated discouragement in Chelsea.

Aware of Chelsea owner Roman Abramovich's interest in the young Argentine, Laporta said last night: "I will answer as always, Messi is not for sale and we do not want to let him go."

# The goal of coreference resolution

**Determining which mentions in a discourse refer to the same real-world entity, property or situation.**

Example:



# The goal of coreference resolution

**Determining which mentions in a discourse refer to the same real world entity, property or situation.**

Example:

FC Barcelona president Joan Laporta has warned Chelsea off star strike Lionel Messi

This warning has generated discouragement in Chelsea.

Aware of Chelsea owner Roman Abramovich's interest in the young Argentine, Laporta said last night: "I will answer as always, Messi is not for sale and we do not want to let him go."

# Identity noun-phrase coreference resolution

- Most commonly investigated type of coreference.
- Determining which **noun-phrases** in a discourse refer to the **same real-word entity**

Ex: [Messi]<sub>1</sub> is not for sale. We do not want to let [him]<sub>1</sub> go.

Ex: [The car]<sub>1</sub> hit a tree. [The vehicle]<sub>1</sub> was found one day later.

Ex: [Bruce Springsteen]<sub>1</sub> will play in Barcelona. [The Boss]<sub>1</sub> is well liked in that place.

We will focus on identity noun-phrase coreference resolution.  
(For simplicity, we will refer to it simply as *coreference resolution*).

# Motivation

Coreference resolution is involved in many NLP applications

- Reading machine: required to fully understand text
- Information extraction
  - Ex: Extract organizations in which a person has worked.
- Automatic summarization
  - Ex: Find relevant sentences related to a particular person.
- Question Answering:
  - Ex: Answer factual questions such as *Where has Mary Doe been working?*

*She worked at UPC. ... Mrs. Doe moved to IBM in 1976*



# Outline

# Types of referent

- Referring expressions:

world knowledge can be necessary for the resolution

Ex: Laporta, the president, FC Barcelona president Joan Laporta

Ex: Lionel Messi, the young Argentine

- Pronouns:

linguistic information can be useful for the resolution  
(number, genre, grammatical constraints)

Ex: Laporta said: "I will answer"

Ex: the president said: "I will answer"

# Coreferent, Anaphor, Cataphor

## Orthogonal concepts

- Coreferent: two mentions refer to the same world entity
- Anaphora: a mention (*anaphor*) refers to a preceding mention (*antecedent*) and the interpretation of the anaphor depends on the antecedent.  
*John Smith* had been writting for months. *He* ended up sleeping on the bed.
- Cataphora: the antecedent occurs after the anaphor (*cataphor*)  
*He* had been writting for months. *John Smith* ended up sleeping on the bed.
- Anaphor is more frequently used than cataphor

# Coreference vs Anaphora

Anaphora resolution is different to coreference resolution

	Example	Anaphora	Coreference
	[ <i>John</i> ] <sub>1</sub> wrote [ <i>his</i> ] <sub>1</sub> book.	X	X
	[ <i>Steve Jobs</i> ] <sub>2</sub> set up Apple in 1976. [ <i>The genius</i> ] <sub>2</sub> died in 2011.	X	X
	[ <i>The man</i> ] <sub>3</sub> wrote [ <i>his</i> ] <sub>3</sub> book.	X	X
(1)	[ <i>Every dog</i> ] <sub>3</sub> has [ <i>its</i> ] <sub>3</sub> day.	X	
(2)	The boy entered [ <i>the room</i> ] <sub>4</sub> . The [ <i>door</i> ] <sub>4</sub> closed automatically.	X	
(3)	[ <i>Apple</i> ] <sub>5</sub> launched the iPhoneX today. [ <i>Apple</i> ] <sub>5</sub> has already won 100 millions.		X

- (1) Not a real world entity - a quantified concept
- (2) Not the same entity
- (3) They don't need each other to be interpreted

# Outline

# General methodology of a coreference solver



## Mention detection:

- Detects the boundaries of the mentions in the input text.
- $m = (m_1, m_2, \dots, m_n)$  ordered as found in the document.

## Coreference prediction:

- find the coreference chains.
- Heuristic-driven approaches: based on the centering theory of the discourse [Grosz et al., 83, 95]. See details in [Walker et al., 98].
- ML-based approaches
- Hand-crafted rules

# Outline

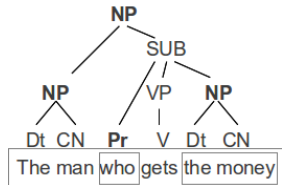
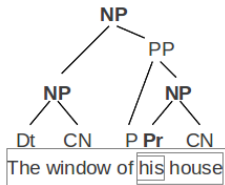
# Mention detection

- Preprocess: POS-tagging, NERC and parsing (constituent parsing or dependence parsing).
- Recursively visiting the parse tree, accept the following as mention
  - Pronouns (filter out pleonastic pronouns, e.g., **It** *is raining*)
  - Proper names
  - Maximal noun-phrase projections
  - Coordinated noun phrases



## Examples

Examples of maximal NP projections with constituent parses:



Drop out NPs sharing the same head.

(head: essentially, right-most noun in first sub-constituent)

# Outline

# Outline

# Mention-Pair model

- Examples:  $(m_i, m_j)$  classified as CO/NC.
- Two steps:
  - 1 Learn a classifier of mention pairs. Ex:
    - Decision Trees [McCarthy & Lehnert, 95], [Soon et al., 01]
    - Rule induction (RIPPER) [Ng & Cardie, 02]
    - Maximum Entropy [Denis & Baldrige, 07], [Ji et al., 05]
    - SVMs [Yang et al., 06]
  - 2 Generate chains. Ex:
    - Closest-first strategy [Soon et al., 01]
    - Best-first strategy [Ng & Cardie, 02][Bengtson & Roth, 08]
    - Clustering [Klenner & Ailloud 2008]...
    - Global optimization (ILP) [Klenner, 07][Finkel & Manning, 08]
    - Graph partitioning [McCallum & Wellner, 05][Nicolae & Nicolae, 06][Sapena et al, 10]...

## 1) Learn a mention-pair classifier

Creating training examples from an annotated corpus:

- **[Soon et al, 01]**

- the classifier is biased to select the closest antecedent
- $e^+ \implies (m_i, m_j)$  Anaphor  $m_j$  and closest antecedent  $m_i$
- $e^- \implies \forall k : i < k < j : (m_k, m_j)$

- **[Ng & Cardie, 02]**

- the classifier is biased to select the best antecedent
- $e^+ \implies (m_i, m_j)$  Anaphor  $m_j$  and closest antecedent  $m_i$   
but for non-pronominal anaphor  $m_j$  select the closest non-pronominal antecedent  $m_i$
- $e^- \implies \forall k : i < k < j : (m_k, m_j)$

## Exercise

- 1 Identify all the mentions for identity noun-phrase coreference resolution in the following text:

Mr. Smith was traveling when Lara came back home. He had never been far from his wife. Mrs. Smith closed the door and went to bed thinking of John.

- 2 Extract all positive and negative examples of coreferent mention pairs for closest-first and for best-first strategies.

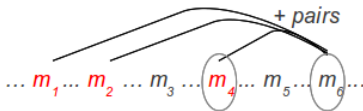
# 1) Learn a mention-pair classifier

Examples of feature functions for mention pair  $(m_i, m_j)$ :

Type	Feature	Description
Structural	DIST_SEN_k	distance in sentences is k: y,n
	DIST_SEN_>2	distance in sentences greater than 2: y,n
	DIST_MEN_k	distance in mentions is k: y,n
	DIST_MEN_>2	distance in mentions greater than 2: y,n
	APPPOSITIVE	One mention in apposition with the other: y,n
Lexical	STR_MATCH	String matching: y,n
	ALIAS	One mention is an alias of the other: y,n,u
Morphological	NUMBER	The number of both mentions match: y,n,u
	GENDER	The gender of both mentions match: y,n,u
Syntactic	DEF_NP	$m_i$ is a definitive NP: y,n
	DEM_NP	$m_i$ is a demonstrative NP: y,n
Semantic	SEMCLASS	Semantic class match: y,n,u
	ANIMACY	Animacy of both mentions match: y,n

## 2) Generate chains

### Closest-first strategy [Soon et al., 01]

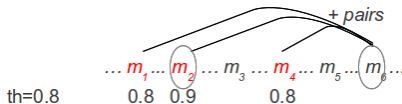


- for a given  $m_j$ , select as antecedent the closest preceding  $m_k$  from the  $+ \text{ pairs}$  provided by a binary mention-pair classifier
- if a probabilistic classifier is used then define a threshold above which a pair is considered coreferent



## 2) Generate chains

**Best-first strategy** [Ng & Cardie, 02][Bengtson & Roth, 08]



- for a given  $m_j$ , select as antecedent the most probable precedent  $m_k$  from the +pairs achieved by a probabilistic mention-pair classifier
- aims to improve the Precision of closest-first strategy by taking profit of the probabilities of the +pairs

## Exercise

Assume we have already learned a mention-pair classifier. Consider  $m_1, \dots, m_9$  as the ordered sequence of mentions in a text. Given the following probabilities for the mention pairs:

$$\begin{aligned}P(CO | \langle m_1, m_7 \rangle) &= 0.8; & P(CO | \langle m_2, m_7 \rangle) &= 0.6; \\P(CO | \langle m_3, m_7 \rangle) &= 0.4; & P(CO | \langle m_4, m_7 \rangle) &= 0.5; \\P(CO | \langle m_5, m_7 \rangle) &= 0.7; & P(CO | \langle m_6, m_7 \rangle) &= 0.6; \\P(CO | \langle m_7, m_8 \rangle) &= 0.9; & P(CO | \langle m_7, m_9 \rangle) &= 0.5;\end{aligned}$$

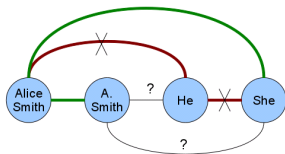
provide the  $m_7$  antecedent that results from applying:

- 1 Closest-first strategy
- 2 Best-first strategy

assuming a coreference threshold of 0.6.

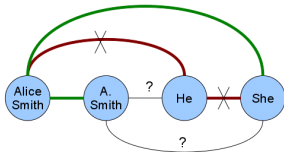
## Drawbacks of the Mention-Pair model

Lack of information.

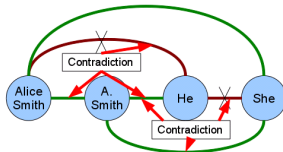


# Drawbacks of the Mention-Pair model

Lack of information.

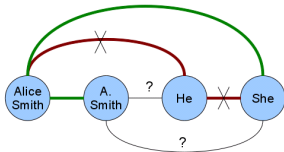


Contradictions in classification.

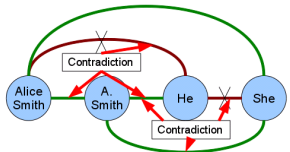


# Drawbacks of the Mention-Pair model

Lack of information.



Contradictions in classification.

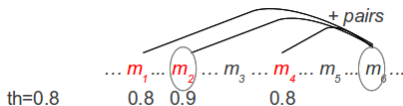


Mention-ranking models and Entity-mention model are different perspectives to deal with the problem.

# Outline

# Mention-ranking models

**Mention-pair models** take profit of independent mention pair decisions between  $m_j$  and each possible antecedent.



**Mention-ranking models** take profit of decisions between  $m_j$  and all its possible antecedents.

The mention-pair classifier is replaced by a ranker.

For a given  $m_j$ , provide a ranking of the set of possible antecedents

# Mention-ranking models

Ex: rankers [Denis and Baldrige, 08]

- Learn a ranker from examples
- Example =  $(m_i, \alpha_i, A_i)$ , where  $\alpha_i$  is the first antecedent of  $m_i$  (the non-pronominal one if  $m_i$  is non-pronominal) and  $A_i$  is the set of non-antecedents in a context of 2 sentences around  $\alpha_i$
- Exponential model:

$$P(\alpha_i | m_i) = \frac{\exp \sum_k \lambda_k f_k(m_i, \alpha_i)}{\sum_{m_s \in A_i \cup \{\alpha_i\}} \exp \sum_k \lambda_k f_k(m_i, m_s)}$$

- Resolution:  $\alpha_i$  is any potential antecedent of  $m_i$ .



## Mention-ranking models

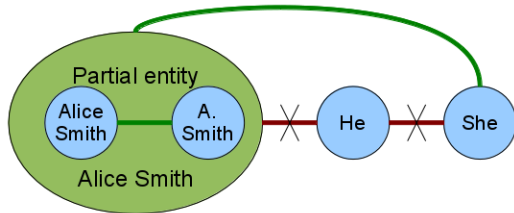
- Pros: take profit of decisions involving all the candidate antecedents.
- Cons: always pick an antecedent from the candidates, although the mention in course is not anaphoric.

(a classifier of anaphoricity improves the results)

# Outline

# Entity-Mention model

- The classifier classifies a partial entity and a mention (or two partial entities in some approaches)
- Entities characterization:
  - **Partial entity**: a set of mentions considered coreferent during the resolution
  - Each partial entity is represented as the set of features of its mentions.
  - Each partial entity has its representative mention



# Outline

# Stanford Easy-first approach

- 1 Exact string match
- 2 Relaxed string match  
Ex: "Clinton", "Clinton, whose term ends in January"
- 3 Precise constructs (appositives, predicate nominatives, ...)  
Ex: "[Donald Trump] is [the president of the USA]"
- 4 Strict head matching: (entity-mention model)  
Ex: {"the Florida Supreme Court"}, {"the Florida court"}
- 5 Strict head matching – Variant 1:  
Ex: {"American President Clinton", "Clinton"}, {"President Clinton"}
- 6 Strict head matching – Variant 2:  
Ex: {"an organization of journalists", "The Gridiron Club"}, {"The Gridiron Club at Greenbrier Hotel"}
- 7 Proper head noun matching  
Ex: "southern Lebanon", "Lebanon"
- 8 Relaxed Head Matching (entity-mention model)  
Ex: {"the judge", "Circuit Judge N. Sanders Sauls"}, {"Sanders"}
- 9 Pronouns

# Outline

# References

- B.J. Grosz, A.K. Joshi and S. Weinstein, **Providing a unified account of definite noun phrases in discourse**. Proceedings of ACL, 1983.
- B.J. Grosz, A.K. Joshi and S. Weinstein, **Centering: A framework for modeling the local coherence of discourse**. Computational Linguistics, 21(2), 1995.
- M. Walker, A. Joshi, and E. Prince, editors. 1998. **Centering theory in discourse**. Oxford University Press
- W.M. Soon, H.T. Ng and D.C.Y. Lim, **A machine learning approach to coreference resolution of noun phrases**. Computational Linguistics, 27(4), 2001
- V. Ng and C. Cardie, **Improving machine learning approaches to coreference resolution**. Proceedings of ACL, 2002

# References

- C. Nicolae and A. Nicolae, **Best Cut: A graph algorithm for coreference resolution**. Proceedings of EMNLP, 2006.
- V. Ng, **Supervised noun phrase coreference research: The First Fifteen Years**. Proceedings of ACL, 2010.
- E. Sapena, L. Padró and J. Turmo. 2010. **A global relaxation labeling**. Proceedings of COLING 2010