



Co-funded by the
Erasmus+ Programme
of the European Union

REFERENCE PHENOMENA

Prof. Jesús Vilares

jesus.vilares@udc.es



UNIVERSIDADE DA CORUÑA

Introduction

- ▶ **Reference resolution:** determining what entities are referred to by which linguistic expression

REFERRING EXPRESSIONS

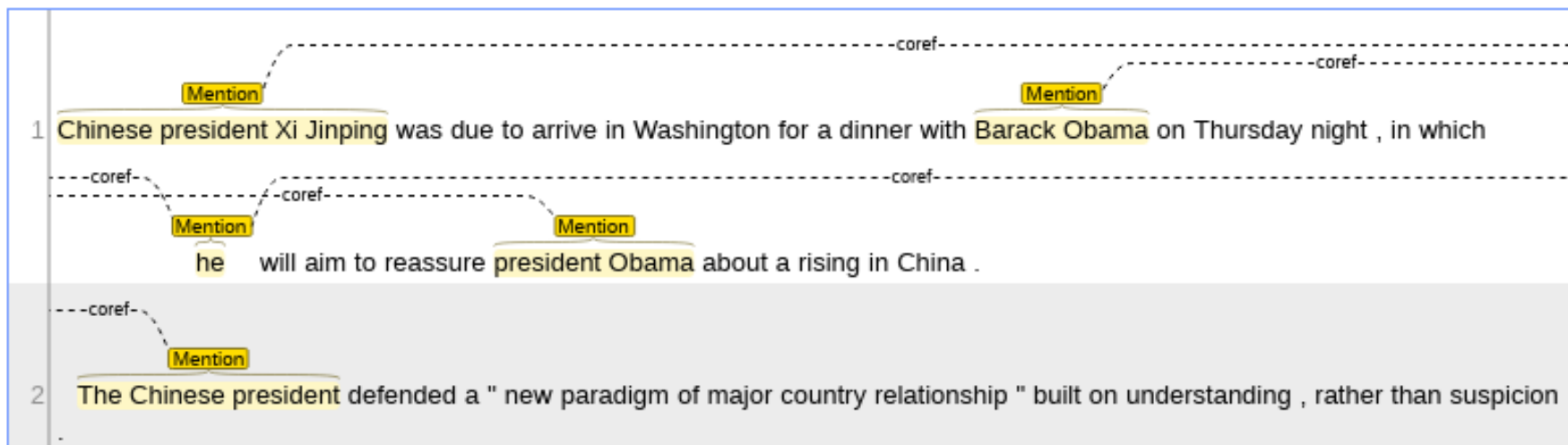
"Barack Obama"
"president Obama"

REFERENTS



REFERRING EXPRESSIONS

"Chinese president Xi Jinping"
"he"
"The Chinese president"





Tasks

- ▶ We describe two reference resolution tasks:
 1. **[Pronominal] anaphora resolution:** given a pronoun in a text, finding the NP of the text referred by the pronoun (i.e. finding its *antecedent*)

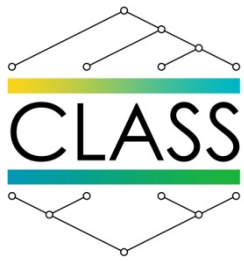
"Chinese president Xi Jinping was due to (...) in which he will aim to reassure (...) "

2. **[General] coreference resolution:** finding all referring expressions (NPs) in a text that refer to the same real-world entity (i.e. finding expressions that *corefer*)

"Barack Obama"
"president Obama"



"Chinese president Xi Jinping"
"he"
"The Chinese president"



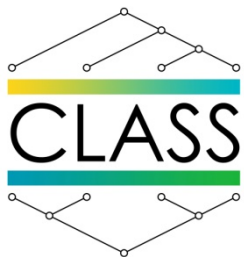
[Pronominal] Anaphora resolution

- ▶ Requires a **hand-labeled training corpus** where each pronoun has been linked by hand with its correct *antecedent* (NP).
- ▶ A **classifier** is trained using positive and negative examples of anaphoric relations:
 - **Positive examples:** those already labeled in the corpus

“Chinese president Xi Jinping was due to (...) in which he will aim to reassure (...)”

- **Negative examples:** obtained by pairing the pronouns in the corpus with other NPs of their previous contexts different from their respective antecedents

“ (...) with Barack Obama on Thursday night, in which he will aim to reassure (...)”



[Pronominal] anaphora resolution:

Features & restrictions

1. **Number agreement:** pronoun and antecedent NP must agree in *number*.

e.g. Mary has adopted two puppies. They are lovely!

Mary has adopted ~~two puppies~~. She is lovely!

2. **Gender agreement:** pronoun and antecedent NP must agree in *gender*.

e.g. ~~John~~ married Mary last year. He is very lucky.

John married Mary last year. She is very lucky.

3. **Person agreement:** pronoun and antecedent NP must agree in *grammatical person*.

e.g. The boys lost contact with John and me. They were worried.

~~The boys~~ lost contact with John and me. We were worried.



[Pronominal] anaphora resolution:

Features & restrictions

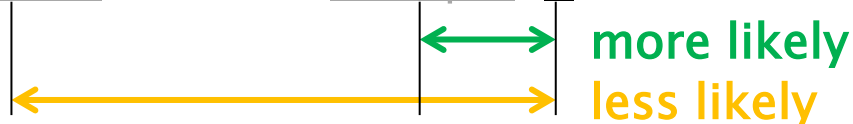
4. **Binding Theory constraints:** when pronoun and antecedent NP occur in the same sentence, they may be constrained by their *syntactic relationship*.

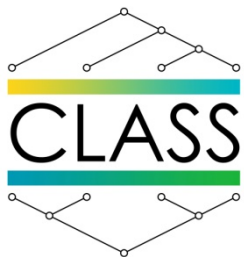
e.g. John said that ~~Bill~~ bought him a new car. [him \neq Bill]
~~John~~ said that Bill bought himself a new car. [himself = Bill]

5. **Distance:** the further pronoun and its candidate antecedent are, the less probable they are connected through a reference.

- Different measure units: no. of words in-between both, no. of NPs in-between, no. of sentences, etc.

e.g. Lex bought a Ford and Mike an Opel. It has a diesel engine.





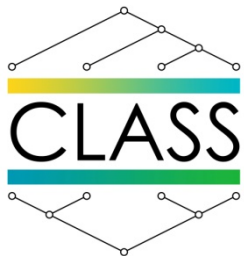
[Pronominal] anaphora resolution:

Features & restrictions

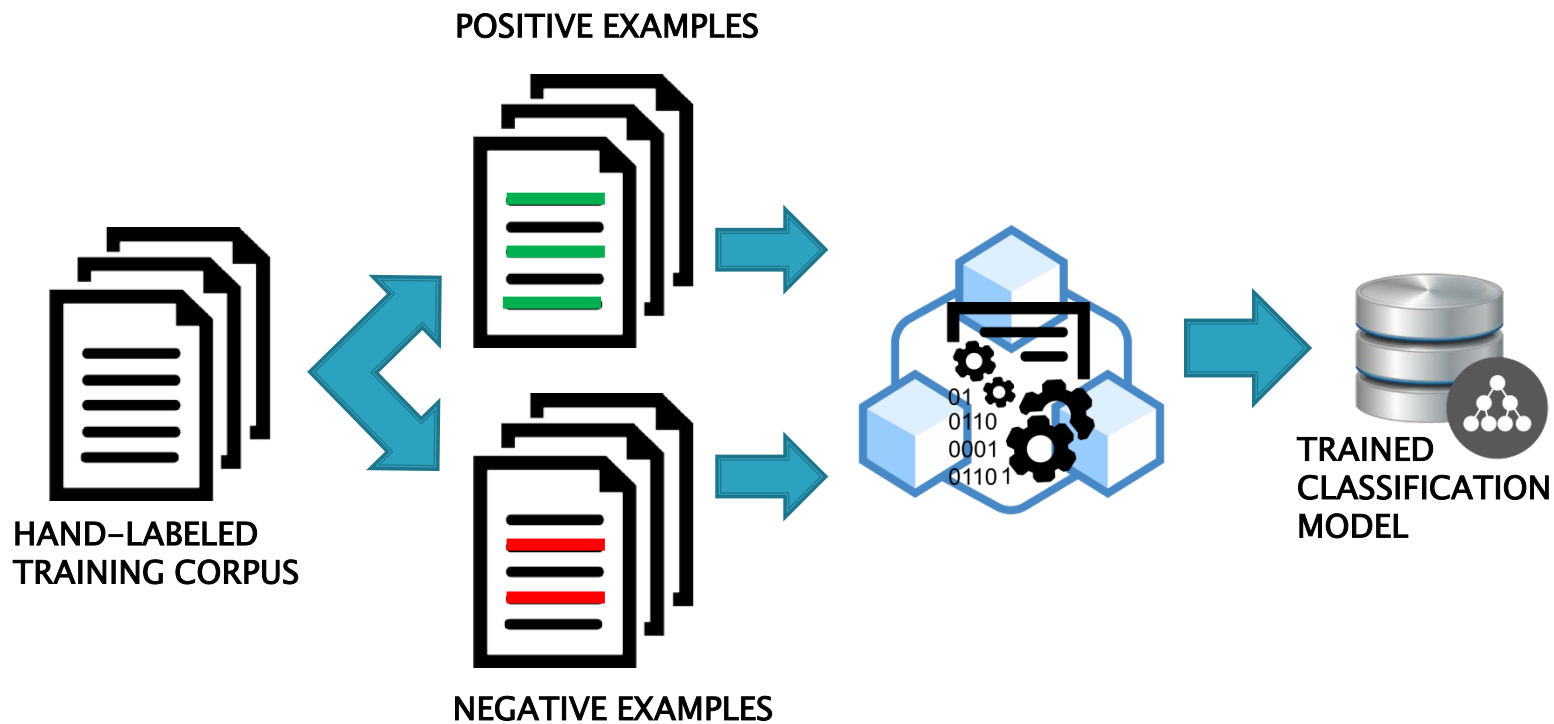
6. **Selectional restrictions:** semantic-type constraints that a verb imposes on the kind of concepts that are allowed to be its arguments

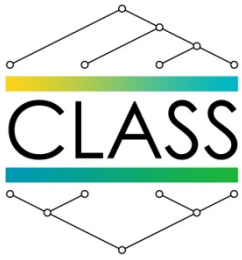
e.g. Olga sat on ~~the car~~, took her sandwich and began to **ate** it.

- “it” is being **eaten** (“to ate it”)
- To **eat** something, it must be **eatable**
- Two candidates: “the car”, “her sandwich”
- A “car” is **not eatable**; thus, it is not a valid candidate
- but a “sandwich” is **eatable**



[Pronominal] anaphora resolution: Training





[Pronominal] anaphora resolution: Reference resolution

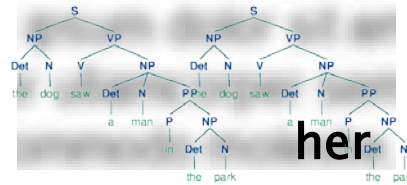
PRONOUN
+ PRECEDING TEXT

...
...
her

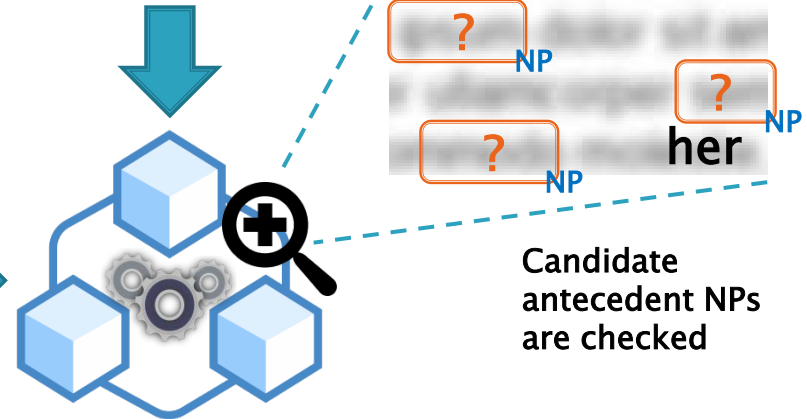
SYNTACTIC PARSER



PARSED TEXT



TRAINED
CLASSIFICATION
MODEL



Candidate
antecedent NPs
are checked

Mary NP ← her

SOLVED REFERENCE



[General] Coreference resolution

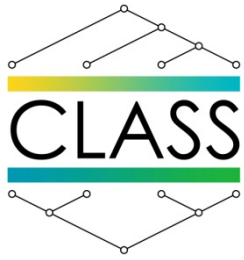
- ▶ Now **any pair of NPs** may corefer.
- ▶ Requires a **hand-labeled training corpus** where each referring expression (NP) has been linked by hand with its correct *antecedent* (other NP).
- ▶ A **classifier** is trained using positive and negative examples of anaphoric relations:

- **Positive examples:** those already labeled in the corpus

" (...) dinner with **Barack Obama** (...) to reassure **president Obama** about (...) "

- **Negative examples:** obtained by pairing the anaphor NPs of the positive examples with those preceding NPs between themselves and their respective correct antecedents.

" (...) in which **he** will aim to reassure **president Obama** about (...) "



[General] Coreference resolution:

Features & restrictions

- ▶ The same as for pronominal anaphora and some others
- 1. **String similarity** between the potential antecedent and the anaphor NP. For example, **minimum edit distances** from the potential antecedent to the anaphor NP and viceversa.
 - Note: The *minimum edit distance* from string A to string B is the minimum number of character editing operations (removals, insertions and substitutions) needed to transform A into B.

e.g. (...) than Cristiano Ronaldo or Lionel Messi. Leo Messi, however (...)

["Lionel Messi" ~ "Leo Messi"
"Cristiano Ronaldo" ≠ "Leo Messi"



[General] Coreference resolution:

Features & restrictions

2. **Alias** (NER required): given two named entities (A,B) of the same type, A is an *alias* of B if they can be matched by applying a given set of possible operations. For example:
 - **PERSON**: by removing titles (e.g. “Mr.”), posts (e.g. “president”), etc.

e.g. Trump met Kim Jong Un. President Trump has travelled to (...)
“~~President~~ Trump” → “Trump” = “Trump”
 - **ORGANIZATION**: by checking for acronyms, etc.

e.g. “European Union” ↔ “EU”
3. **Apposition**: two NPs linked through syntactic apposition.

e.g. The ex-President of the USA, Barack Obama, has visited (...)



Bibliography

- ▶ **[Jurafsky & Martin, 2009]** Jurafsky, D. & Martin, J.H. (2009). Chapter 21: Computational Discourse. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition (2nd ed.)*. Pearson-Prentice Hall.
- ▶ **[Mitkov, 2002]** Mitkov, R. (2002). *Anaphora Resolution*. Pearson Education.
- ▶ **[Mitkov, 2010]** Mitkov, R. (2010). Chapter 21: Discourse Processing. In Clark, A., Fox, C. & Lappin, S. (Eds.), *The Handbook of Computational Linguistics and Natural Language Processing*. Wiley-Blackwell.
- ▶ **[Nugues, 2006]** Nugues, P.M. (2006). Chapter 14: Discourse. *An Introduction to Language Processing with Perl and Prolog*. Springer.