

# 12 – Anaphora resolution

## IA161 Advanced Techniques of Natural Language Processing

M. Medved'

NLP Centre, FI MU, Brno

December 8, 2021

# Anaphora resolution

## Example

- 1.1 H : all right system  
1.2 : first thing I would like you to do  
1.3 : is send **Engine1** off with a boxcar to **Brno** to pick up oranges  
1.4 : as soon as possible  
2.1 S : okay  
3.1 H : and while **it** is **there**, **it** should pick up **the tanker**  
4.1 S : okay  
4.2 : **it** can get that too  
4.3 : we can get **that** done by three  
5.1 H : good  
5.3 : can we please send **Engine** over to **Prague** to pick up boxcar  
5.4 : and then send **it** back to **Bratislava**  
6.1 S : okay  
6.2 : **it** will get to **Bratislava** at 6

## 1 Linguistic fundamentals

- Notation and terminology
- Coreference
- Anaphora
- Varieties of anaphora according to the form of the anaphora
- Types of anaphora according to the locations of the anaphora and the antecedent
- Indirect anaphora
- Location of anaphora
- Location of anaphora

## 2 The process of automatic anaphora resolution

- Anaphora resolution input
- Anaphora resolution in practice
- The resolution algorithm
  - The resolution algorithm - constraints
  - The resolution algorithm - preferences

### 3 Early anaphora resolution approaches

- Hobbs algorithm

### 4 Recent anaphora resolution approaches

- Main trends in recent anaphora resolution research
- Results

1.

## Linguistic fundamentals

## Used in areas

- question answering
- machine translation
- automatic summarization

# Linguistic fundamentals

- cohesion:

- ▶ collection of discourse sentences, phrases or words that are related

## Example (Small context length)

After an hour Elizabeth came home too. When he saw her ...

## Example (Appropriate context length)

Carl came home first. After an hour Elizabeth came home too. When he saw her ...

# Linguistic fundamentals

- anaphora:
  - ▶ **Informal**: cohesion which 'points back' to some previous discourse item
  - ▶ **Formal**: an expression the interpretation of which depends upon another expression in context before
- antecedent:
  - ▶ discourse entity to which the anaphora refers or for which it stands



# Notation and terminology

- anaphora resolution:
  - ▶ process of determining the antecedent of the anaphora
- coreference:
  - ▶ anaphora and its antecedent are coreferential when both have the same referent in real world

## Example

**The Queen** is not here yet but **she** is expected to arrive in the next half an hour.

## ATTENTION

The relation between the anaphor and antecedent is not to be confused with that between the anaphor and its referent.

## Referent vs. antecedent



Figure: Person from real word

The Queen is not here yet but she is expected to arrive in the next half an hour.

Figure: Linguistic form

# Coreference

- coreferential chain:
  - ▶ if more than one preceding noun phrases are coreferential

## Example

**This book** is about anaphora resolution. **The book** is designed to help beginners in the field and **its** author hopes that **it** will be useful.

- definite NPs in copular relation are considered as coreferential (including apposition):

## Example

David Beckham **is** the Manchester United midfielder .  
Dominique Voynet, the French Environment Minister ...

# Coreference

- anaphoric relation does not imply coreference relation between discourse entities:

## Example

**Every man** has **his** own destiny.

## Substitution test

Use the substitution test to establish the **coreference**.

**Correct:** **John** has **his** own destiny. -> **John** has **John's** destiny.

**Incorrect:** **Every man** has **his** own destiny. -> Every man has every man's destiny.

# Anaphora

We differentiate between:

- identity-of-sense anaphora
  - ▶ does not denote the same entity as its antecedent, but one of a similar description

## Example

The man who gave his **paycheck** to his wife was wiser than the man that gave **it** to his mistress.

- identity-of-reference anaphora
  - ▶ anaphora and the antecedent have the same referent in the real world and are therefore coreferential

## Example

**This book** is about anaphora resolution. **The book** is designed to help beginners in the field and **its** author hopes that **it** will be useful.

# Varieties of anaphora according to the form of the anaphora

- pronominal anaphora
- lexical noun phrase anaphora
- noun anaphora
- verb anaphora
- adverb anaphora
- zero anaphora (ellipsis)

# Varieties of anaphora according to the form of the anaphora

## (Pronominal anaphora)

- personal pronouns (he, him, she, her, it, they, them):

### Example

The most difficult for **David** was to tell her, that **he** loved her.

- possessive pronouns (his, her, hers, its, their, theirs):

### Example

**David** locked keys in **his** car.

- reflexive pronouns (himself, herself, itself, themselves):

### Example

**David** once again locked **himself** in his studio.

## Varieties of anaphora according to the form of the anaphora (Pronominal anaphora)

- demonstrative pronouns (this, that, these, those):

### Example

David, however, used **photographic precision** to transcribe the images of his dreams. **This** would become ...

- relative pronouns (who, whom, which, whose):

### Example

**David** , a Catalan **who** was addicted to fame and gold, painted a lot and talked a lot.



# Varieties of anaphora according to the form of the anaphora

## (Pronominal anaphora)

### Non-anaphoric usage of pronouns:

- pleonastic it:

#### Example

It is dangerous to be beautiful.

- generic usage of pronouns:

#### Example

He who dares wins.

- deictic usage of pronouns (pointing to specific person in given situation):

#### Example

He seems remarkably bright for a child of his age.

## Varieties of anaphora according to the form of the anaphora (Lexical noun phrase anaphora)

- definite description, proper names
- may have same head (these footprints -> footprints)
- synonyms (shop -> store)
- generalization (boutique -> the shop)

### Example

**Roy Keane** has warned Manchester United **he** may snub their pay deal.  
**United's skipper** is even hinting ... **Irishman Keane**, 27, still has 17  
months to run ... **Alex Ferguson's No. 1 player** confirmed ...

# Varieties of anaphora according to the form of the anaphora

## (Noun anaphora)

### ATTENTION

Noun phrase anaphora should not be confused with noun anaphora.

### Definition

**Noun anaphora** occurs when there is an anaphoric relation between a non-lexical proform (function word that expresses the same content) and the head noun or nominal group of a noun phrase.

### Example

I don't think I'll have a sweet **pretzel**, just a plain **one**.

# Varieties of anaphora according to the form of the anaphora

## (Verb anaphora)

### Example

... it was inevitable that his midfield prodigy would **follow**, and in 1981 he **did**.

# Varieties of anaphora according to the form of the anaphora (Adverb anaphora)

- locative: there
- temporal: then

## Example

Will you walk with me to **the garden**? I've got to go down **there** ...

# Varieties of anaphora according to the form of the anaphora

## (Zero anaphora)

- zero anaphora (ellipsis)

### Example

**Willie** paled and **\_he\_** pulled the sock up quickly.

# Types of anaphora according to the locations of the anaphora and the antecedent

- intrasentential:
  - ▶ anaphor and its antecedent are located in the same sentence
- intersentential:
  - ▶ antecedent is in a different sentence from the anaphor

## Example

**Rob** and **his** wife come to London last night. **He** ...

# Indirect anaphora

- indirect anaphora:
  - ▶ reference becomes part of the hearer's or reader's **knowledge** indirectly rather than by direct mention
  - ▶ **relationships** such as: part of, subset of, set membership

## Example

Although **the store** had just opened, **the food hall** was busy.



# Location of anaphora

- expected/possible distance between the anaphora and the closest antecedent
- narrow down the search scope of candidates for antecedents

# Location of anaphora

- expected/possible distance between the anaphora and the closest antecedent
- narrow down the search scope of candidates for antecedents

## Research

- about 90 % do not exceed 2-5 sentences (Hobbs(1978), McEnery et al.(1997) ...)
- distance between a pronominal anaphora and its antecedent reported in Hobbs (1978) is 13 sentences and in Fraurud (1988) is 15 sentences
- present writer found it quite common for proper names to refer to antecedents which are 30 or more sentences away

## 2.

The process of automatic anaphora resolution

# Anaphora resolution input

- sentence splitter
- tokenizer
- morphological and lexical knowledge
- syntactic knowledge
- semantic knowledge

## Example

The petrified kitten refused to come down from the tree. It gazed ...

- discourse knowledge
  - ▶ center or focus

## Example

Tilly's mother had agreed to make her a new dress for the party. She worked hard on the dress for weeks and finally it was ready for Tilly to try on. Impatient to see what it would look like, Tilly tried on the dress over her skirt and ripped it.

- real-world (common-sense) knowledge

# Anaphora resolution in practice

Process consists of:

- identification of anaphors
- location of the candidates for antecedents
- selection of antecedent from set of candidates

# The resolution algorithm - restrictions

once the anaphors have been detected, the program tries to select antecedents from set of candidates

- constraints:

- ▶ gender and number agreement
- ▶ c-command constraints
- ▶ selectional restrictions

- preferences:

- ▶ the most recent NP
- ▶ candidates in the main clause
- ▶ NPs which are positioned higher in the parse tree
- ▶ candidates in non-adjunct phrases
- ▶ syntactic parallelism
- ▶ center preference
- ▶ subject preference

# The resolution algorithm - constraints

- gender and number agreement

## Example

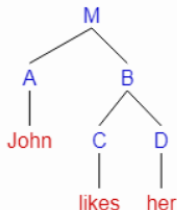
Because **Klein** tried 'dirty tricks', they refused to support **him**.

# The resolution algorithm - constraints

- c-command constraints (intrasentential)

## Definition

A node **A** c-commands a node **B** if and only if (i) **A** does not dominate **B**, (ii) **B** does not dominate **A**, (iii) the first branching node dominating **A** also dominates **B**.



**Figure:** Where [node A] John c-commands [node B]. This means that [node A] also c-commands [node C] and [node D], which means that John c-commands both [likes] and [her].



# The resolution algorithm - constraints

- selectional restrictions
  - ▶ semantic restrictions that apply to the anaphor should apply to the antecedent as well

## Example

George removed the disk from **the computer** and then disconnected **it**.

# The resolution algorithm - preferences

- syntactic parallelism:
  - ▶ noun phrases that have the same syntactic function as the anaphor

## Example

The programmer successfully combined **Prolog** with C, but he had combined **it** with Pascal last time.

# The resolution algorithm - preferences

- center preference:

## Definition

**Center** is most prominent entity in utterance.

- ▶ sentence that is likely to be pronominalised in a subsequent clause or sentence

## Example

Tilly's mother had agreed to make her **a new dress** for the party. She worked hard on **the dress** for weeks and finally **it** was ready for Tilly to try on. Impatient to see what **it** would look like, Tilly tried on **the dress** over her skirt and ripped **it**.

# The resolution algorithm - preferences

- subject preference:

## Example

**The customer** lost patience and called the waiter. **He** ordered two 12-inch pizzas.

# Simple anaphora resolution

- gender and number agreement
- c-command
- center

## Example

How poignant that **one of the television tributes** paid to **Jill Dando** shows **her** interviewing people just before the funeral of Diana Princess of Wales. Some of the words she used to describe the late princess could equally applied to her.

- number agreement removes "the television tributes"
- gender agreement removes "one of the television tributes"
- antecedent: Jill Dando

# Simple anaphora resolution

## Example

How poignant that **one of the television tributes** paid to **Jill Dando** shows her interviewing **people** just before **the funeral of Diana Princess of Wales**. **Some of the words** **she** used to describe the late princess could equally applied to her.

- number and gender agreement leaves only: "Jill Dando", "Diana Princess of Wales"
- antecedent (according center): Jill Dando

# Simple anaphora resolution

## Example

How poignant that **one of the television tributes** paid to **Jill Dando** shows her interviewing **people** just before **the funeral of Diana Princess of Wales**. **Some of the words** she used to describe the **late princess** could equally applied to her.

- number and gender agreement leaves only: "Jill Dando", "Diana Princess of Wales"
- antecedent (according c-command): "Diana Princess of Wales"

# 3.

## Early anaphora resolution approaches



# Hobbs algorithm

- one of the most influential works in the field
- for pronoun resolution
- algorithm traverses the surface parse tree in a particular order looking for a noun phrase of the correct gender and number

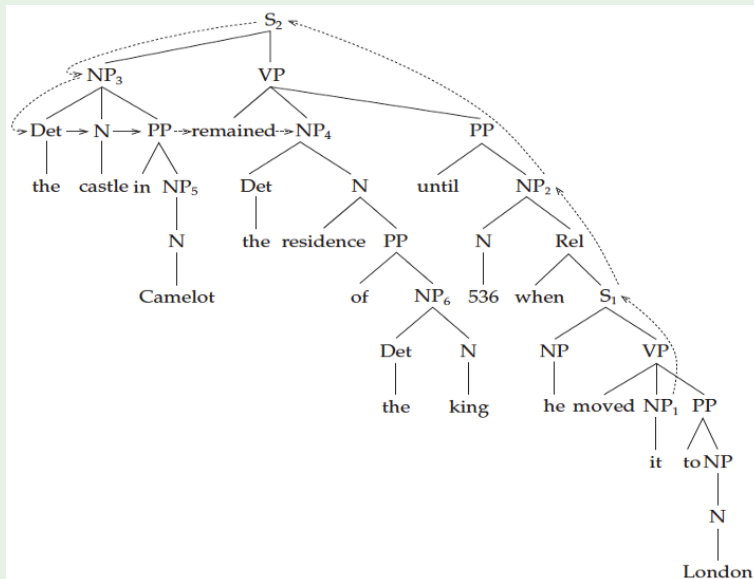
# Description of Hobbs algorithm

- 1 Begin at the NP node immediately dominating the pronoun in the parse tree of the sentence S.
- 2 Go up the tree to the first NP or S node encountered. Call this node X, and call the path used to reach it p.
- 3 Traverse all branches below node X to the left of path p in a left-to-right, breadth-first fashion. Propose as the antecedent any NP node encountered that has an NP or S node between it and X.
- 4 If the node X is the highest S node in the sentence, traverse the surface parse trees of previous sentences in the text in order of recency, the most recent first; each tree is traversed in a left-to-right, breadth-first manner, and when an NP node is encountered, it is proposed as antecedent. If X is not the highest node in the sentence, proceed to step 5.

## Description of Hobbs algorithm: continue

- 5 From node X, go up the tree to the first NP or S node encountered. Call this node X and call the path traversed to reach it p.
- 6 If X is an NP node and if the path p to X did not pass through the N-bar node that X immediately dominates, propose X as the antecedent.
- 7 Traverse all branches below the node X to the left of path p in a left-to-right, breadth-first manner. Propose any NP node encountered as the antecedent.
- 8 If X is S node, traverse all branches of node X to the right of path p in a left-to-right, breadth-first manner, but do not go below any NP or S node encountered. Propose any NP node encountered as the antecedent.
- 9 Go to step 4.

Example (The castle in Camelot remained the residence of the king until 536 when he moved it to London.)



- based on two submodules:
  - ▶ a rule-based mentions-detection module which uses SpaCy's tagger, parser and NER annotations to identify a set of potential coreference mentions
  - ▶ a feed-forward neural-network which compute a coreference score for each pair of potential mentions

## Neuralcoref (extracting mentions)

Yes, **I** noticed that **many friends**, around **me** received **it**.  
**It** seems that **almost everyone** received **this SMS**.

- Each mention can co-refer with a various number of previous mentions

Mention	Potential antecedents						
I	Ø						
many friends	Ø	I					
me	Ø	I	many friends				
it	Ø	I	many friends	me			
It	Ø	I	many friends	me	it		
almost everyone	Ø	I	many friends	me	it	It	
this SMS	Ø	I	many friends	me	it	It	almost everyone

# Neuralcoref (extracting mentions' features)

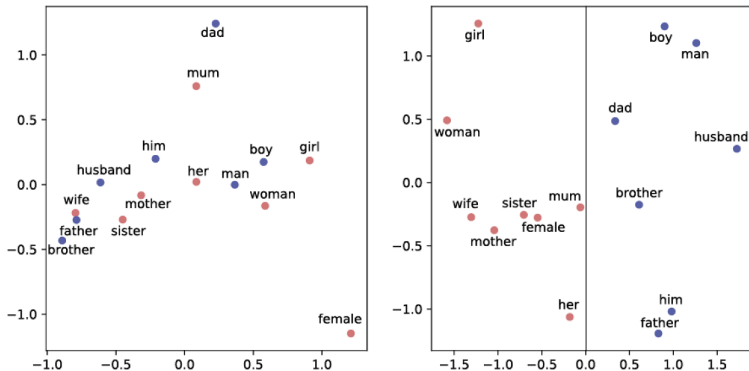
- For each mention and mention pair a set of features is extracted

Category	Mention features							
Type	Type (noun/...)		Nested ?		Doc (talk/news/...)			
Location / size	Location (int)		Length (int)					
Word indices	Head	0	Last	+1	+2	-1	-2	Root head
Span vectors	Mention	Before	After	Sentence		Doc		

Category	Mention-pair features		
Speakers	Same ?	Speaker name in mention/antecedent ?	
String match	Exact match ?	Relaxed match ?	Head match ?
Locations	Distance	Sentence distance	Overlapping ?

# Neuralcoref (training the module)

- pre-trained word vectors that contains lots of features (OntoNotes 5.0 dataset)





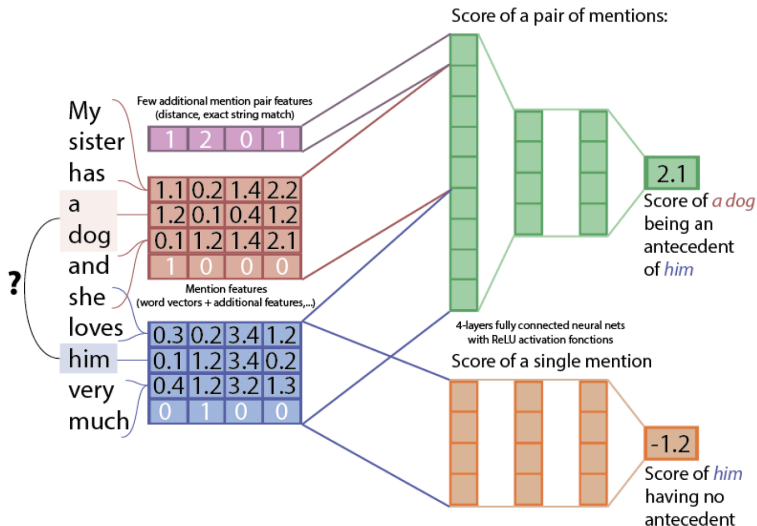
# Neuralcoref (training the module)

## • Training data

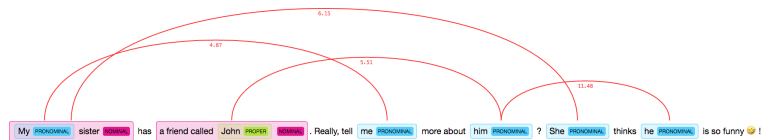
```
1 #begin document (bc/cctv/00/cctv_0005); part 003
2
3 bc/cctv/00/cctv_0005 3 0 Yes UH (TOP(S(INTJ+)) Wang_shilin * (ARGM-DIS*) *
4 bc/cctv/00/cctv_0005 3 1 , , Wang_shilin * *
5 bc/cctv/00/cctv_0005 3 2 I PRP (NP*) Wang_shilin * (ARG0*) (12)
6 bc/cctv/00/cctv_0005 3 3 noticed VBD (VP* notice 01 1 Wang_shilin * (V*)
7 bc/cctv/00/cctv_0005 3 4 that IN (SBAR* Wang_shilin * (ARG1*)
8 bc/cctv/00/cctv_0005 3 5 many JJ (S(NP(NP*) Wang_shilin * (ARG0*)
9 bc/cctv/00/cctv_0005 3 6 friends NNS * Wang_shilin *
10 bc/cctv/00/cctv_0005 3 7 , , Wang_shilin * *
11 bc/cctv/00/cctv_0005 3 8 around IN (PP* Wang_shilin *
12 bc/cctv/00/cctv_0005 3 9 me PRP (NP*)) Wang_shilin * * (12)
13 bc/cctv/00/cctv_0005 3 10 received VBD (VP* receive 01 1 Wang_shilin * (V*)
14 bc/cctv/00/cctv_0005 3 11 it PRP (NP*)) Wang_shilin * * (ARG1*) (119)
15 bc/cctv/00/cctv_0005 3 12 , , Wang_shilin * *
16
17 bc/cctv/00/cctv_0005 3 0 It PRP (TOP(S(NP*) Wang_shilin * *
18 bc/cctv/00/cctv_0005 3 1 seems VBZ (VP* seem 01 1 Wang_shilin * (V*)
19 bc/cctv/00/cctv_0005 3 2 that IN (SBAR* Wang_shilin * (ARG1*)
20 bc/cctv/00/cctv_0005 3 3 almost RB (S(NP* Wang_shilin * * (ARG0*)
21 bc/cctv/00/cctv_0005 3 4 everyone NN *) Wang_shilin * * *)
22 bc/cctv/00/cctv_0005 3 5 received VBD (VP* receive 01 1 Wang_shilin * (V*)
23 bc/cctv/00/cctv_0005 3 6 this DT (NP*) Wang_shilin * * (ARG1*) (119)
24 bc/cctv/00/cctv_0005 3 7 SMS NN *) Wang_shilin * * *) (119)
25 bc/cctv/00/cctv_0005 3 8 , , *) Wang_shilin * *
26
27 #end document
```

# Neuralcoref (training the module)

- Model visualization



# Neuralcoref (Example output)



- Online demo: <https://huggingface.co/coref/>

# 4.

## Recent anaphora resolution approaches





# Main trends in recent anaphora resolution research

- knowledge-poor approach: cheaper and more reliable corpus-based NLP tools
- corpora:
  - ▶ co-occurrence rules
  - ▶ training decision trees
  - ▶ identify anaphor–antecedent pairs
- neural networks trained on annotated corpora (BERT)

## State-of-the-art results on GAP dataset

- ProBERT (2019), Gendered Ambiguous Pronouns Shared Task: Boosting Model Confidence by Evidence Pooling : F1 92.5 %
- Attree (2019), multi-head attention and feedforward (FFN) 92.5 %
- Liu (2019) FFNN and Bidirectional LSTM : 89.26 %

# References I

-  Kocijan, V., Camburu, O.-M., Cretu, A.-M., Yordanov, Y., Blunsom, P., and Lukasiewicz, T. (2019).  
Wikicrem: A large unsupervised corpus for coreference resolution.
-  Lata, K., Singh, P., and Dutta, K. (2020).  
A comprehensive review on feature set used for anaphora resolution.  
*Artificial Intelligence Review*, pages 1–90.
-  Liu, F., Zettlemoyer, L., and Eisenstein, J. (2019).  
The referential reader: A recurrent entity network for anaphora resolution.  
*CoRR*, abs/1902.01541.
-  Mitkov, R. (2002).  
*Anaphora Resolution*.  
Studies in Language and Linguistics. Longman.

## References II



Toshniwal, S., Ettinger, A., Gimpel, K., and Livescu, K. (2020).  
Petra: A sparsely supervised memory model for people tracking.