Class 11: Anaphora and co-reference resolution

Things and naming things

What is co-reference resolution?

THE prime minister has fired secretary of state Priti Patel while telling her she wishes with all her heart it was the other way around.

May confessed to Patel that although having secret meetings with the Israelis was a sackable offence, it paled in comparison to her own dire performance but “sadly nobody’s willing to pull the trigger.”

Patel said: “It was so awkward. She said ‘You don’t know how often I’ve dreamt of sitting on your side of the desk, finally being summarily dismissed for my gross incompetence.’

What is co-reference resolution?

THE prime minister has fired secretary of state Priti Patel while telling her she wishes with all her heart it was the other way around.

May confessed to Patel that although having secret meetings with the Israelis was a sackable offence, it paled in comparison to her own dire performance but “sadly nobody’s willing to pull the trigger.”

Patel said: “It was so awkward. She said ‘You don’t know how often I’ve dreamt of sitting on your side of the desk, finally being summarily dismissed for my gross incompetence.’

What is co-reference resolution?

THE prime minister has fired secretary of state Priti Patel while telling her she wishes with all her heart it was the other way around.

May confessed to Patel that although having secret meetings with the Israelis was a sackable offence, it paled in comparison to her own dire performance but “sadly nobody’s willing to pull the trigger.”

Patel said: “It was so awkward. She said ‘You don’t know how often I’ve dreamt of sitting on your side of the desk, finally being summarily dismissed for my gross incompetence.’

What is co-reference resolution?

THE prime minister has fired secretary of state Priti Patel while telling her she wishes with all her heart it was the other way around.

May confessed to Patel that although having secret meetings with the Israelis was a sackable offence, it paled in comparison to her own dire performance but “sadly nobody’s willing to pull the trigger.”

Patel said: “It was so awkward. She said ‘You don’t know how often I’ve dreamt of sitting on your side of the desk, finally being summarily dismissed for my gross incompetence.’

What is co-reference resolution?

THE prime minister has fired secretary of state Priti Patel while telling her she wishes with all her heart it was the other way around.

May confessed to Patel that although having secret meetings with the Israelis was a sackable offence, it paled in comparison to her own dire performance but "sadly nobody's willing to pull the trigger."

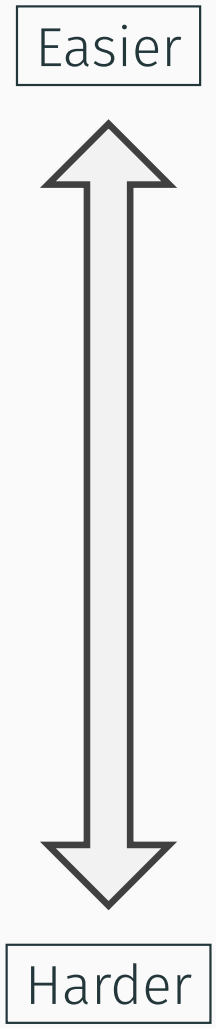
Patel said: "It was so awkward. She said 'You don't know how often I've dreamt of sitting on your side of the desk, finally being summarily dismissed for my gross incompetence.'

Noun phrases and reference

- NPs usually refer to entities in the world
- NPs may co-refer, meaning they refer to the same entity
- They may also be nested or discontinuous

Однажды Пушкин написал письмо Рабиндранату Тагору. «Дорогой далекий друг, — писал он, — я Вас не знаю, и Вы меня не знаете. Очень хотелось бы познакомиться. Всего хорошего. Саша». Когда письмо принесли, Тагор предавался самосозерцанию. Так погрузился, хоть режь его. Его жена толкала, толкала, письмо подсовывала — не видит. Он, правда, по-русски читать не умел. Так и не познакомились.

Kinds of reference



Interesting linguistics

Bound variables

She hurt *herself*
Я имею *свой* баян

Free variables

Maša read *her* book
Ей очень нравилась.

Referring expressions

Carles Puigdemont
the Catalan president
Puigdemont
Puchi
president of Catalonia
President Puigdemont

More frequent

Coreference, anaphora, cataphora

- Coreference
  - Two *mentions* (NPs) refer to the same entity
  - May be identical or completely different
- Anaphora, Cataphora
  - Interpretation is in some way dependent on an antecedent
  - Traditionally the antecedent came first, but not always the case.

Cataphora

(Oscar Wilde – The Picture of Dorian Grey)

Other types of anaphora

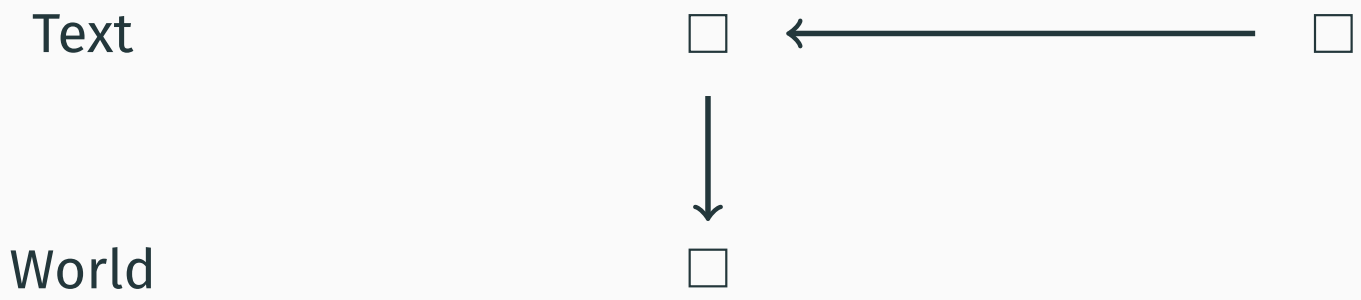
- Bridging Anaphora:
- ”Other” Anaphora:
- Non-NP Anaphora (e.g. events, propositions)

Anaphora vs. coreference

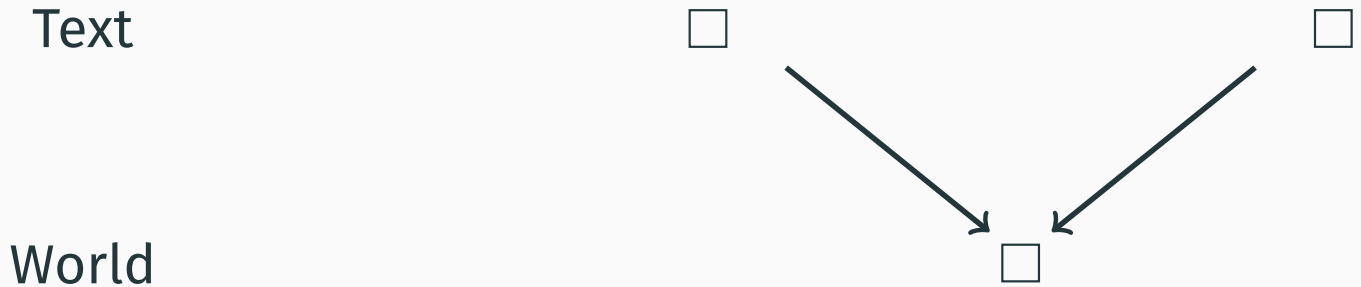
- Not all anaphoric relations are coreferential, e.g. bridging anaphora
- Multiple identical NP matches are often coreferential but not anaphoric

Two different things

Anaphora resolution



Co-reference resolution



Applications

Machine translation:

- Translating pronouns like *себя* — myself, yourself, herself, ...
- Translating from languages with no gender distinction in pronouns *här — он, она*

Text summarisation:

- [Maša]$_i$ read [Wikipedia]$_j$ with unbridled enthusiasm. [She]$_i$ spent so long reading [it]$_j$ that [she]$_i$ forget about her homework.
  - $\rightarrow$ Maša read Wikipedia and forgot about her homework.

Information extraction:

- What did Maša do ?
  - read Wikipedia
  - forgot about her homework

Pronominal anaphora resolution

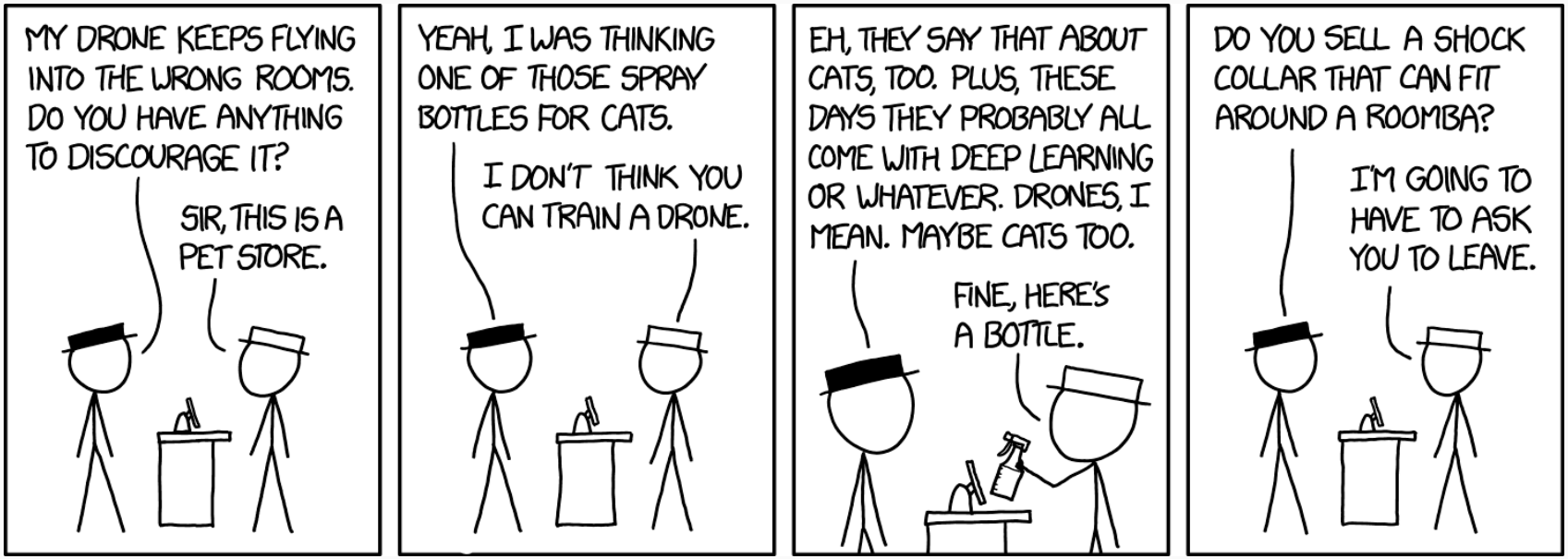
Hobbs' (1978) algorithm/1

- Simple syntax-based algorithm for 3rd person anaphoric pronouns
- Requires:
  - Constituency parser
  - Gender and number 'checker'
    - Parsers for English rarely include gender information for nouns
- Searches current and preceding sentences in a breadth-first, left-to-right manner, stops when it finds a matching NP

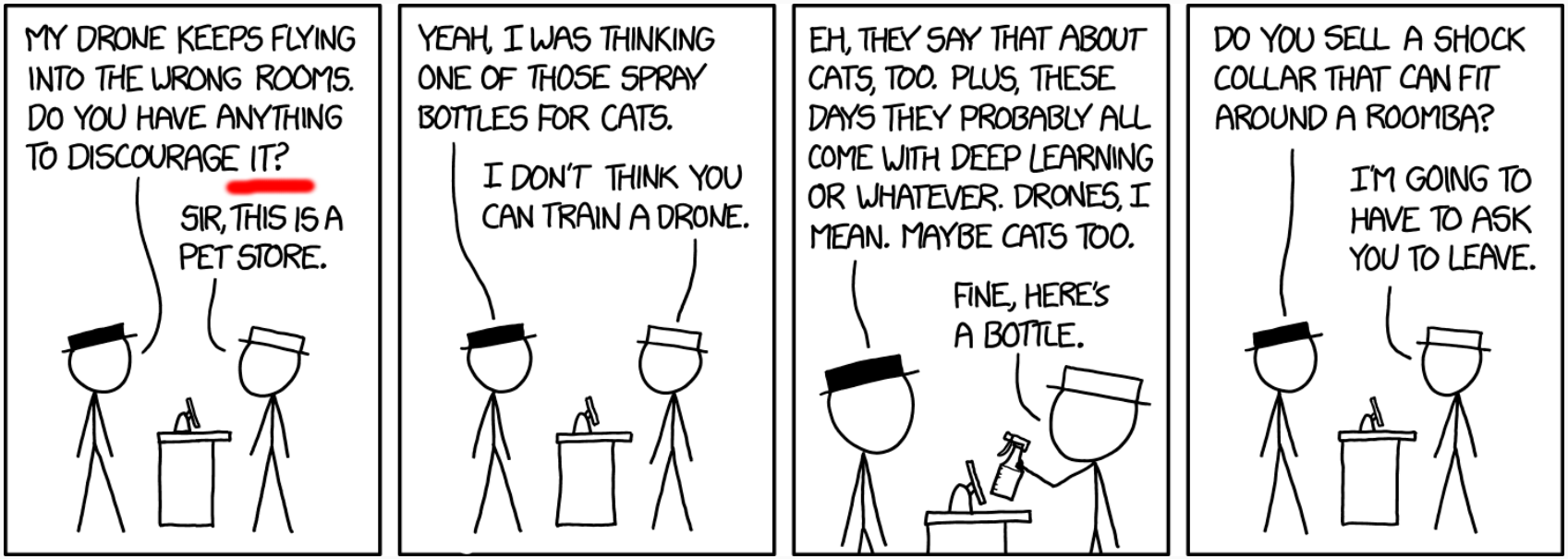
Hobbs' (1978) algorithm/2

- Right to left search in current sentence
- If not valid antecedent fine, try previous sentence
    - Left to right breadth-first search

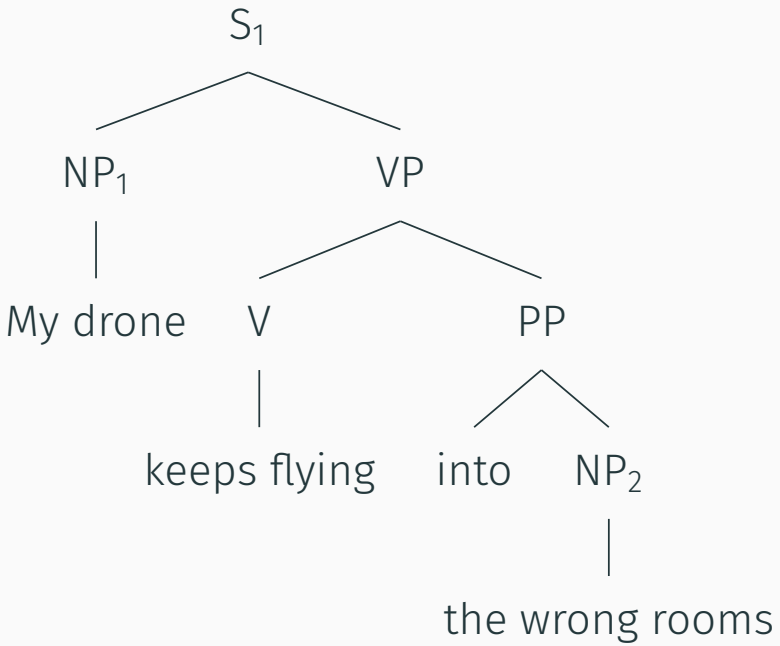
Hobbs' (1978) algorithm/3

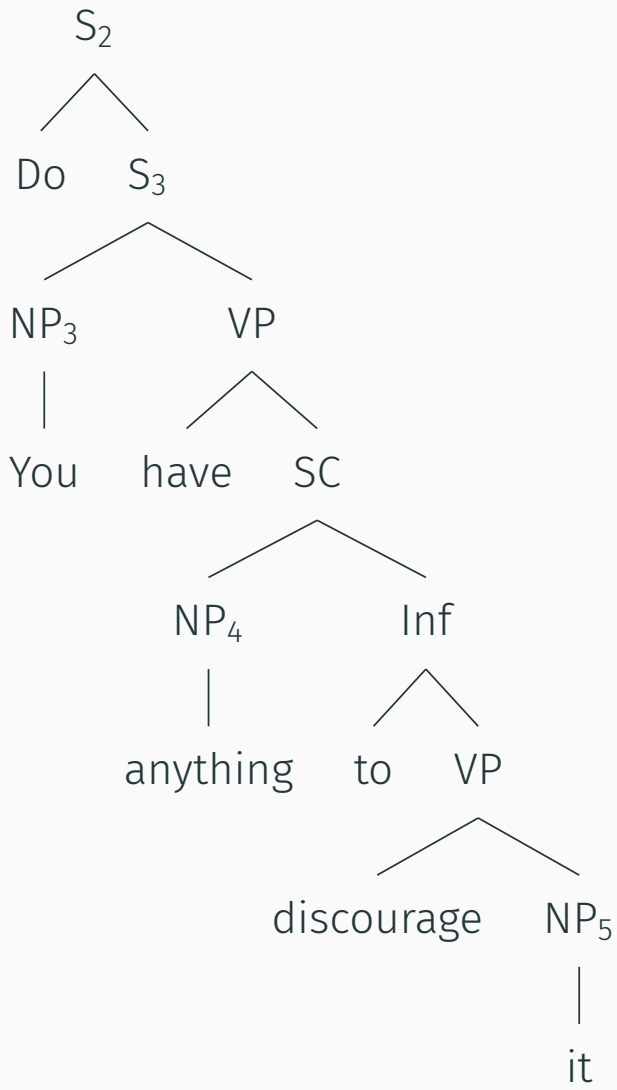


Hobbs' (1978) algorithm/3



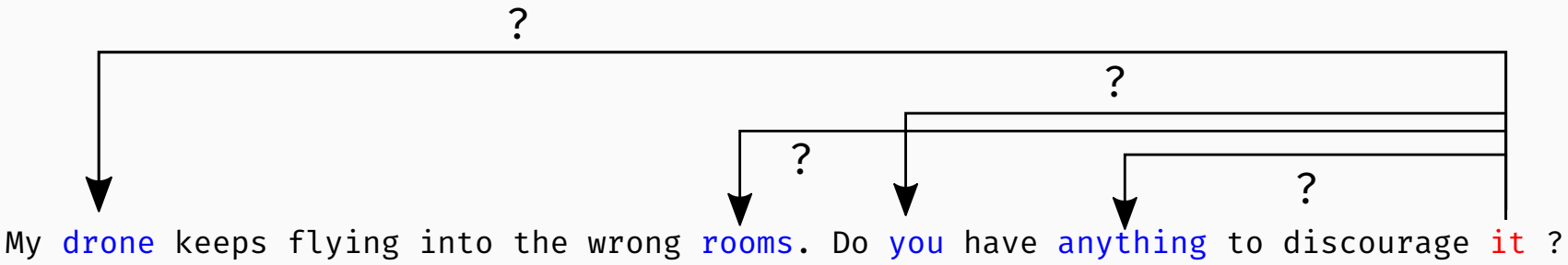
Hobbs' (1978) algorithm/4





- Start search in NP_5 in S_2
- Reject NP_4, no intervening NP
- Reject NP_3, feature mismatch
- Move to S_1
- Accept NP_1

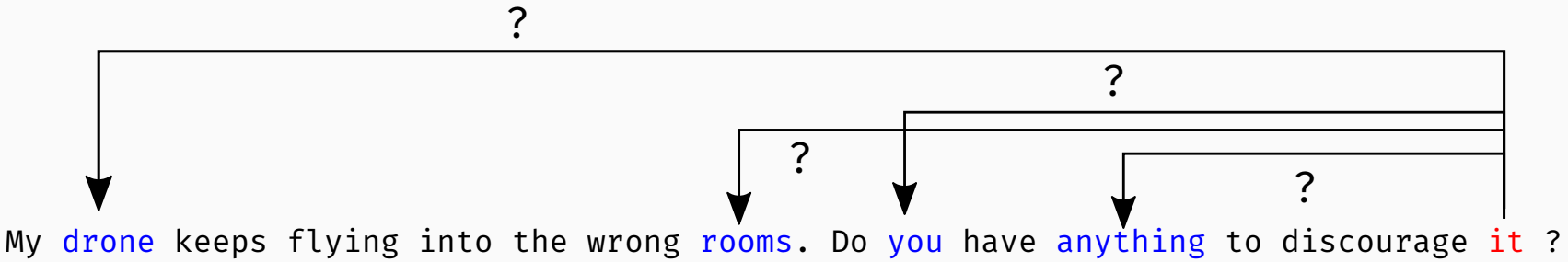
Log-linear model/1



- Supervised machine learning approach
- Requires corpus where each pronoun has been linked with its antecedent

- Extract positive and negative examples
- Train binary classifier
  - True: is co-referent
  - False: is not co-referent

Log-linear model/2



Positive example:

- (it, my drone)

Negative examples:

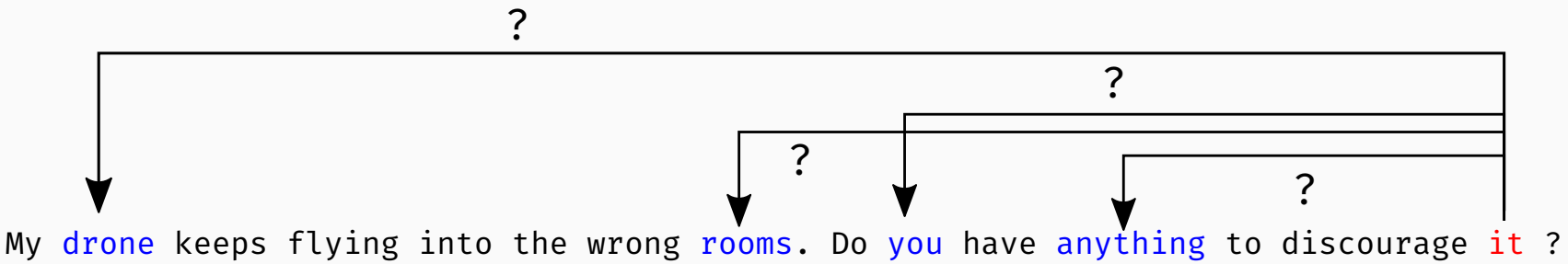
- (it, anything)
- (it, you)
- (it, the wrong rooms)

Log-linear model/3

- **strict gender** [true, false]
- **compatible gender** [true, false]
- **strict number** [true, false]
- **compatible number** [true, false]
- **sentence distance** [0, 1, 2, ...]
- **Hobbs distance** [0, 1, 2, ...]
- **grammatical role** [subject, object, ...]
- **linguistic form** [proper, def, indef, pronoun]

Can you think of some other useful features ?

Log-linear model/4



- For each pronoun,
  - For each NP we have seen so far,
    - Classify if NP is an antecedent of the pronoun

Co-reference resolution

Reference models

Однажды Пушкин написал письмо Рабиндранату Тагору. «Дорогой далекий друг, — писал он, — я Вас не знаю, и Вы меня не знаете. Очень хотелось бы познакомиться. Всего хорошего. Саша». Когда письмо принесли, Тагор предавался самосозерцанию. Так погрузился, хоть режь его. Его жена толкала, толкала, письмо подсовывала — не видит. Он, правда, по-русски читать не умел. Так и не познакомились.

Пушкин, он$_1$, я$_1$, меня$_1$

Рабиндранату Тагору, Дорогой далекий друг, Вас$_1$, Вы$_1$, Тагор, его$_1$, Он

General algorithm

Additional features

Rule-based models/1

Input is a sequence of dependency trees representing the document.

Constraint-based, rules like:

$C =< ANA, ANT, DIST, PROP >$

- ANA, ANT = constraints on the anaphor and antecedent
- DIR = direction (e.g. forward, backwards)
- DIST = how far to look (in sentences)
- PROP = should features (e.g. gender) be propagated?

Rule-based models/2

Example rules:

- Mark full-text matches as coreferent
- Two 1st person pronouns in the same quoted speech corefer
- A surname matching a previous firstname + surname corefers

Evaluation

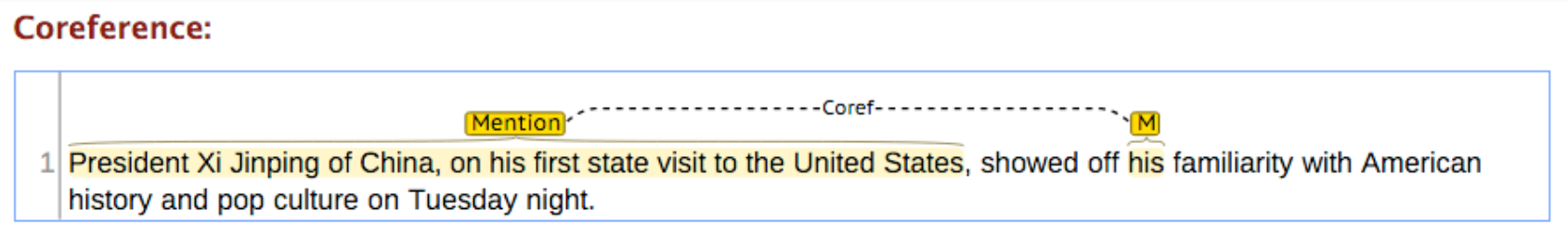
Model-Theoretic coreference scoring

Other metrics

- B^3 scoring
- CEAF

Tools and resources

Stanford CoreNLP



`https://stanfordnlp.github.io/CoreNLP/index.html`

- includes coreference for English + Chinese
- rule-based, statistical and neural models
- based on Java
- very “heavy” ($>$ 1G for code + model)

xrenner

eXternally configurable REference and Non Named Entity Recogniser

`https://github.com/amir-zeldes/xrenner`

- based on Python
- Rule-based (constraints)
- Functioning model for English
- (relatively) small footprint ($<$ 100M)

Shared tasks

SemEval 2010

RuEval-2014

Two tracks:

- Pronominal anaphora resolution (8 teams)
  - personal, possessive, relative and reflexive pronouns
  - antecedent should be in the gold-standard coreference chain
- Co-reference resolution (3 teams)

Practical

Practical

For the practical, select a short paragraph or document and write some coreference rules using Xrenner.

Further instructions:

`https://ftyers.github.io/028-komp-ling/classes/11.html`