08 – Anaphora resolution

IA161 Advanced Techniques of Natural Language Processing

M. Medved'

NLP Centre, FI MU, Brno

November 6, 2017

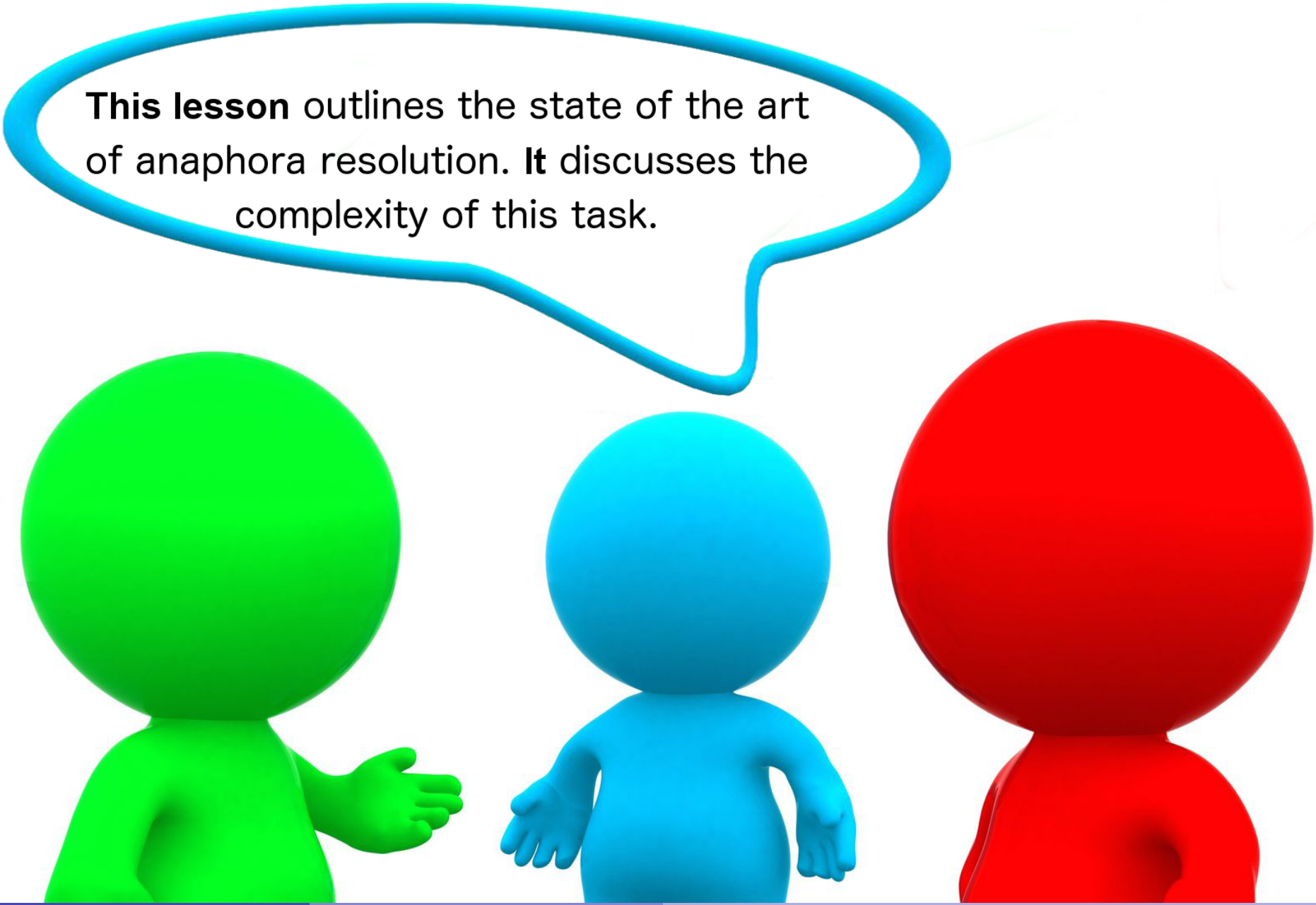
1. Linguistic fundamentals
   - Notation and terminology
   - Coreference
   - Anaphora
   - Varieties of anaphora according to the form of the anaphora
   - Types of anaphora according to the locations of the anaphora and the antecedent
   - Indirect anaphora
   - Location of anaphora
   - Location of anaphora
   - Anaphora and ambiguity

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
   - RAP algorithm
   - SAARA system – Czech

4. Recent anaphora resolution approaches
   - Main trends in recent anaphora resolution research
   - Results

Discourse


This lesson outlines the state of the art of anaphora resolution. **It** discusses the complexity of this task.

1.
Linguistic fundamentals

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 that the man 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 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 % dos 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

Anaphora and ambiguity

- problem when identifying antecedent
- level of ambiguity depends on semantics of the verb

Example

Jane told Mary she was in love.
Jane informed Mary she was in love.

2.

The process of automatic anaphora resolution

Anaphora resolution input

- 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 form set of candidates

Identification of anaphors

- identification of anaphoric pronouns (except **pleonastic it**: Paice and Husk (1987), Evans(2000, 2001))
- identification of anaphoric noun phrases (except **definite descriptions** that are not anaphoric: Bean and Riloff (1999), Munoz (2001))
- tools: morphological analyser, part-of-speech tagger, program for identifying non-anaphoric definite descriptions, parser, annotated corpora, ontology

Location of the candidates for antecedent

- search scope of candidates for antecedent
  - linear models
  - hierarchical models
- tools: full parser (sentence splitter, tokeniser,part-of-speech(POS) taggers), semantic analyser, proper name recogniser

The resolution algorithm

one the anaphors have been detected, the program try to select antecedents form 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**.

- 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**.

Example

She almost wanted Hera to know about the affair.

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 shut down **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.

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

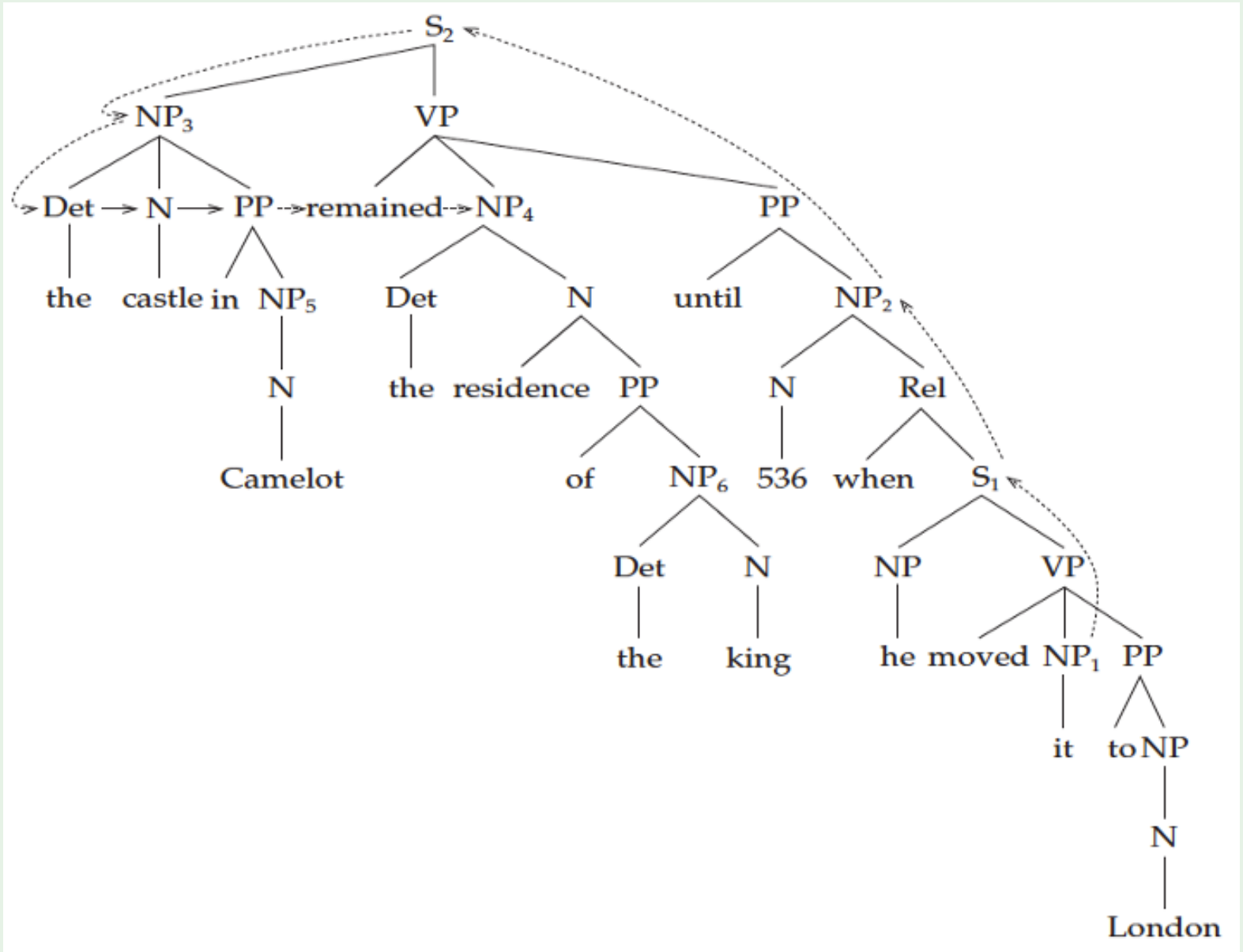
Description of Hobbs algorithm

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.4 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.)



RAP algorithm

- pronoun resolution
- account for all preferences and tendencies of anaphoric references
- based on salience factors

RAP algorithm process

- salience values of all items mentioned in the discourse so far (if any) are cut in half to account for the preference for recent antecedents. All items the salience value of which sank below 1, are removed from the model
- all NPs in the utterance are listed and classified. Determines which expressions introduce a new entity, which of them are non-referential, and which of them are left to be resolved
- salience factors are applied to the individual referring expressions found in the previous step

Salience factors (RAP)

- Sentence recency
- Subject emphasis
- Existential emphasis
- Accusative emphasis
- Indirect object and oblique complement emphasis
- Head noun emphasis
- Non-adverbial emphasis

RAP algorithm components

- **an intrasentential syntactic filter** specifying constraints on NP-pronoun coreference within a sentence
- **a morphological filter** determining non-agreement in gender, number and person
- **a procedure for identification of pleonastic pronouns**, i.e. eliminating non-referential pronoun occurrences from the discourse model
- **an anaphor binding algorithm** for resolving reciprocal and reflexive pronouns to antecedents in the same sentence
- **a procedure for computing salience parameters**, i.e. assigning each referring expression a salience value based on the respective salience factors
- **a procedure for keeping track of equivalence classes**
- **a procedure specifying preferences** for selecting the antecedent from a list of candidates

RAP novelty

- all the precise salience factor weights, have been reached after extensive experimentation and numerous re-adjustments of the individual weights
- possible to re-fit the system for a particular language and genre
- factors formulated in RAP are based mainly on syntactic concepts. Semantic features and real-world knowledge are not considered.

SAARA system – Czech

- containing re-implementations and variants of selected salience-based algorithms
- anaphora resolution algorithm:
  - BFP algorithm
  - RAP algorithm

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 (LSTM)

State-of-the-art results

- Sam et al. 2016 (MUC score): 73.42 % F1
- Cevin et al. 2016 (MUC score): 74.65 % F1

References I

Mitkov, R. (2002).
Anaphora Resolution.
Studies in Language and Linguistics. Longman.

Němčík, V. (2012).
Saara: Anaphora resolution on free text in czech.
In Aleš Horák, P. R., editor, *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN 2012*, pages 3–8, Brno. Tribun EU.