Error Analysis of Noun Phrase Anaphora Resolution using a Machine Learning-based Approach

Ryu IIDA
Computational Linguistics Lab.
ryu-i@is.naist.jp

[COE technical presentation, 23 June 2005]

Background

- There is a huge amount of text data on the Web
- Natural Language Processing (NLP) techniques are a required component for ubiquitous computing
  - Machine Translation, Information Extraction and Question Answering
- Inter-sentential anaphora resolution is a major obstacle to their progress

Anaphora resolution

- **Anaphora resolution** is the process of determining whether two expressions in natural language refer to the same entity in the world

A federal judge in Pittsburgh issued a temporary restraining order preventing Trans World Airlines from buying additional shares of USAir Group Inc. (antecedent / anaphor)
The order, requested in a suit filed by USAir, dealt another blow to TWA's bid to buy the company for $52 a share.

- **Anaphora resolution** is classified into two processes
  1. Anaphoricity determination
  2. Antecedent identification

Anaphora resolution

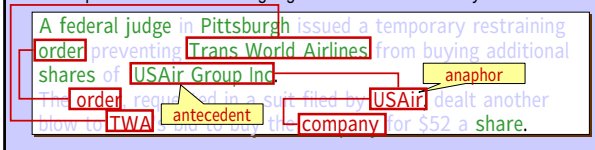
- **Anaphora resolution** is the process of determining whether two expressions in natural language refer to the same entity in the world

A federal judge in Pittsburgh issued a temporary restraining order preventing Trans World Airlines from buying additional shares of USAir Group Inc. (anaphor)
The order, requested in a suit filed by USAir, dealt another blow to TWA's bid to buy the company for $52 a share. (non-anaphor)

- **Anaphora resolution** is classified into two processes
  1. **Anaphoricity determination** is the task of classifying whether a given noun phrase (NP) is *anaphoric* or *non-anaphoric*
  2. **Antecedent identification** is the identification of the antecedent of a given anaphoric NP

Anaphora resolution

- **Anaphora resolution** is the process of determining whether two expressions in natural language refer to the same entity in the world



- **Anaphora resolution** is classified into two processes
  1. **Anaphoricity determination** is the task of classifying whether a given noun phrase (NP) is *anaphoric* or *non-anaphoric*
  2. **Antecedent identification** is process of identifying the antecedent of a given anaphoric NP

Talk outlines

1. Our previous work on anaphora resolution
2. Error analysis of Japanese NP anaphora resolution
3. Discussion and future work

Our previous work on anaphora resolution

1. Antecedent identification model
   (tournament model)

2. Anaphoricity determination model
   (pairwise classification model using the most likely antecedents)

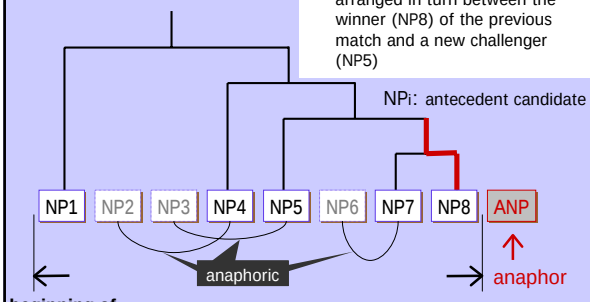
Antecedent identification model

- Tournament model (Iida , 03)
  - A model which makes pair-wise comparisons between candidates to capture the preferences of the antecedents

Tournament model

- Test Phase

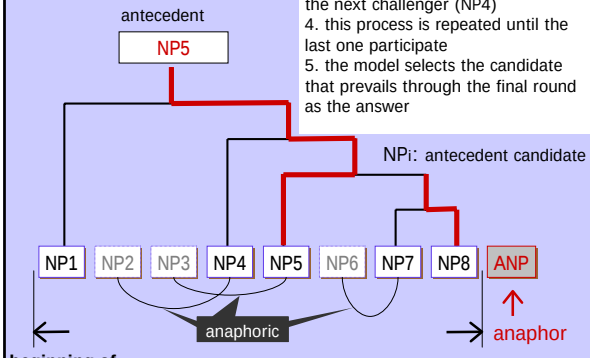
1. the first match is arranged between the nearest candidates (NP7 and NP8)
2. each of the following matches is arranged in turn between the winner (NP8) of the previous match and a new challenger (NP5)



Tournament model

- Test Phase

3. the winner is next matched against the next challenger (NP4)
4. this process is repeated until the last one participate
5. the model selects the candidate that prevails through the final round as the answer

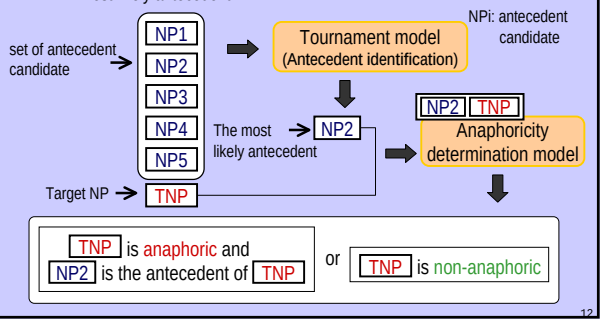


Anaphoricity determination model

- Determining anaphoricity using the most likely antecedents

Pairwise classification model

- 2-step processing:
  1. Identify the most likely antecedent candidate for a target NP
  2. Determine the anaphoricity of the target NP using the target NP and the most likely antecedent



Current performance of our anaphora resolution model

- Antecedent identification for given anaphors:
  - Precision: 87.4% (773/884)
- Anaphoricity determination
  - Recall: 67.1 % (593/884)
  - Precision: 79.5 % (593/746)
- Overall performance
  - Recall: 65.3% (577/884)
  - Precision: 77.3% (577/746)

Error analysis of Japanese NP anaphora resolution

- Investigating the source of errors manually

1. Antecedent identification
   - 111 anaphors
2. Anaphoricity determination
   - Sampled 100 anaphors
   - Sampled 100 non-anaphors

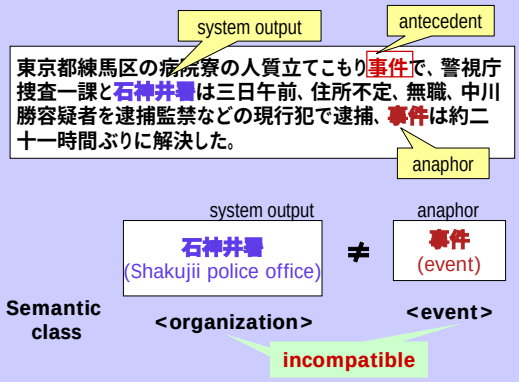
Error analysis: Antecedent identification

- Investigating examples that our system could not identify the correct antecedent for a given anaphor

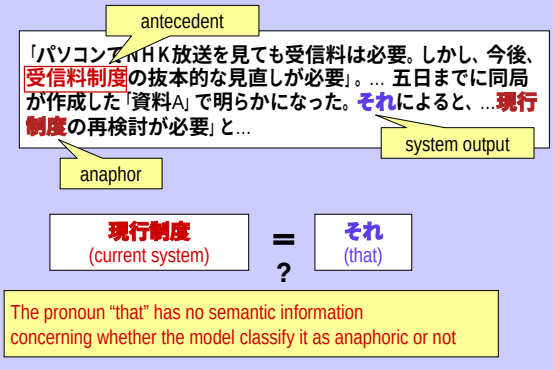
Main source of errors	Percentage (#)
(a) Semantic compatibility	36.9 % (41/111)
(b) Lack of semantic information	5.4 % (6/111)
(c) Referring to set expressions	7.2 % (8/111)
(d) Relationship between entities	9.0 % (10/111)
(e) Annotation error	7.2 % (8/111)
(f) Others	35.1 % (39/111)

(Each example is assigned to more than one source of errors)

Error (a): Semantic compatibility



Error (b): Lack of semantic information



Error analysis: Anaphoricity determination (1)

- Sampling 100 anaphors classified as non-anaphoric for given correct anaphors

Main source of errors	Percentage (#)
(a) Complete string matching	71% (71/100)
(b) Partial string matching	23% (23/100)
(c) Interpretation of bridging reference	49% (49/100)
(d) Lack of semantic information	7% (7/100)
(e) Annotation error	1% (1/100)

(Each example is assigned to more than one source of errors)

Error (a) and (b): (complete,partial) String matching

関根元容疑者は九日までに、埼玉・群馬両県警合同捜査本部の調べに対し、...三人の殺害・死体遺棄を断片的に認める供述を始めた。
殺害にはいずれも薬物を使ったことをほのめかしており、... 。関根容疑者は、川崎さんについては「薬でやった」と**殺害**を ...

殺害 (murder) =? **殺害** (murder)

It is difficult to determine an anaphoricity only using relationship between two NPs

サッカーのインタコンチネンタル選手権が六日から**サウジアラビア**の首都リヤドで始まる。出場するのは各大陸のチャンピオン、日本、デンマーク、メキシコ、アルゼンチン、ナイジェリア、**地元サウジアラビア**の6カ国。

サウジアラビア (Saudi Arabia) **地元サウジアラビア** (Saudi Arabia)

Correct anaphoric pair is classified as non-anaphoric

Error (c): Interpretation of bridging reference

- Bridging reference: referring to a target noun phrase's modifier

関根元容疑者は九日までに、埼玉・群馬両県警合同捜査本部の調べに対し、...**三人**の殺害・死体遺棄を断片的に認める供述を始めた。
殺害にはいずれも薬物を使ったことをほのめかしており、... 。
関根容疑者は、川崎さんについては「薬でやった」と**殺害**を ...

殺害 (murder) = ? **殺害** (murder)

Understanding bridging reference

(三人の)殺害 murder (of three persons) = (三人の)殺害 murder (of three persons)

Error analysis: Anaphoricity determination (2)

- Sampling 100 non-anaphors classified as anaphoric

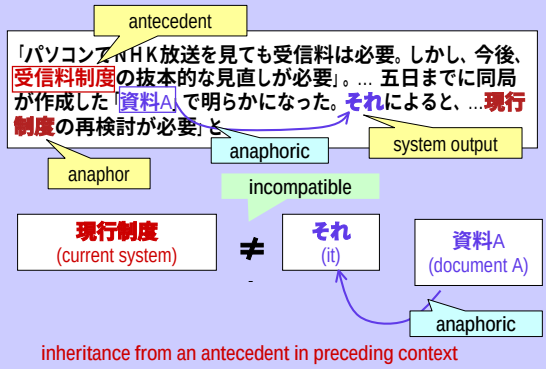
Main source of errors	Percentage (#)
(a) Relationship between entities	12% (12/100)
(b) Complete string matching	43% (43/100)
(c) Partial string matching	26% (26/100)
(d) Interpretation of bridging reference	42% (42/100)
(e) Lack of semantic information	2% (2/100)
(f) Annotation error	14% (14/100)

(Each example is assigned to more than one source of errors)

Future work

1. Improvement of semantic compatibility
   - Constructing Fine-grained semantic classes by combining existing linguistic resources
2. Antecedent identification using the preceding context information
3. Resolving bridging reference

Antecedent identification using the preceding context information



inheritance from an antecedent in preceding context

Resolving bridging reference

- Creating a corpus annotated bridging reference
  - Defining bridging reference is difficult
  - inter-annotator agreement is very low
  - → We need to set specifications for annotating corpora
- Developing a bridging reference resolution model
  - utilizing A of B (A-no-B) co-occurrence information extracted from large corpora
- Noun phrases anaphora resolution utilizing resolved bridging reference

Summary

- Explaining our framework of Noun phrases anaphora resolution
- Main source of errors:
  - Antecedent identification
  - Anaphoricity determination
- Denoting future work based on our error analysis