

Answering Questions of IAD Task using Reference Resolution of Follow-up Questions

Jun'ichi Fukumoto
Ritsumeikan University

QACIAD task

- One question set consists of a series of related questions in Japanese.
- There are some anaphoric expressions including zero anaphora in all questions except the first question of each series.

Example of a series of questions

- Series 30002 (English translation)
 - When did Asahi breweries Ltd. start selling their low-malt beer?
 - What is the brand name?
 - How much did it cost?
 - What brands of low-malt beer were already on the market at that time?
 - Which company had the largest share?
 - How much low-malt beer was sold compared to regular beer?
 - Which company made it originally?

Some approaches to IAD task

- Extract answers of follow-up questions from documents retrieved using clues of the first questions (Sasaki 2002)
- Use of clues extracted from a target question and its previous questions for document retrieval (Murata 2002)

Our approach

- Resolve reference including zero anaphora of follow-up questions to complete the follow-up questions
- Applying the completed question to our main QA system (core QA)

Q1:アメリカの大統領はだれですか。
(Who is the president of USA?)

Q2:フランスはだれですか。
(Who is **Φ** of France?)



Q1:アメリカの大統領はだれですか。
(Who is the president of USA?)

Q2:フランス**の大統領**はだれですか。
(Who is **the president of** France?)




core QA system

Three types of ellipsis patterns

- We analyzed 319 questions of QAC1 and QAC2.
- We classified ellipsis patterns into 3 types:

 Ellipsis using pronoun

 Ellipsis of obligatory case elements of verb

 Ellipsis of modifier or modificand in a follow-up question

Ellipsis using pronoun (pattern1)

- A pronoun in a follow-up question refers to its antecedent in the previous question.

アメリカの大統領は誰です

(Who is the president of America?)

そこが独立したのはいつですか。

(When did it become independent?)

Ellipsis using pronoun (pattern1)

- A pronoun in a follow-up question refers to an answer of the previous question.

アメリカの大統領は誰ですか。<answer>
(Who is the president of America?)

彼の出身地はどこですか。
(Where is his birth place?)

Ellipsis of obligatory case elements of verb (pattern2)

- There is a zero anaphora in a follow-up question.
- The zero anaphora is obligatory case element of verb in the follow-up question.

アメリカの大統領は誰ですか。(ブッシュ)
 (Who is the president of America?) (Bush)

(Φ :agent は) いつ (Φ :goal に) 就任しました
 (When did (he) inaugurate (as President)?)

Ellipsis of modifier or modificand (pattern3)

- There is a zero anaphora in a follow-up question
- The zero anaphora is modifier of the follow-up question

アメリカの大統領は誰ですか。
(Who is the president of America?)

(Φ :*modifier* の) 国務長官はだれですか。
(Where is a minister of state (*of America*)?)




Ellipsis of modifier or modificand (pattern3)

- There is a zero anaphora in a follow-up question
- The zero anaphora is modificand of the follow-up question

アメリカの大統領は誰ですか。
(Who is the president of America?)

フランス (Φ :*modificand* の) はだれですか。
(Who is (Φ of) France?)

Overview of our resolution mechanism of ellipsis

-  Estimate ellipsis pattern (pattern 1, 2 or 3)
-  Estimate kinds of the omitted word according to its ellipsis pattern
 - **Pattern1:** Replace pronoun with a word
 - **Pattern2:** Fill up obligatory case frame
 - **Pattern3:** Find antecedent using co-occurrence information
-  Decide the succeeded word of the previous question

Use of EDR dictionary

- EDR Japanese Word Dictionary
 - Japanese word – concept code
- EDR Concept Dictionary
 - Hierarchy of concept codes
- EDR Co-occurrence Dictionary
 - Japanese co-occurrence data of various modifications

Ellipsis handling in pattern1

- Estimate antecedent types using information of pronoun
- Choose appropriate word from the previous question

Ex.)

○ country × position × human

アメリカの大統領は誰ですか。(ブッシュ)

(Who is the president of America?) (Bush)

そこが独立したのはいつですか。

(When did it becomes independent?)

Organization

Location

Ellipsis handling in pattern2

- Recognize from information of omitted obligatory case of the verb
- We use case information of EDR dictionary

アメリカの大統領は誰ですか。

(Who is the president of America?)

Ex.)

(Φ :agent は)いつ(Φ :goal に)就任しました

(When did (he) inaugurate (as it)?)

agent of 就任する → human

goal of 就任する → position

Selection of the antecedent (pattern2)

■ Same as pattern1

Ex.) × country ○ position ○ human

アメリカの大統領は誰ですか。(ブッシュ)

(Who is the president of America?) (Bush)

(Φ :agent は)いつ(Φ :goal に)就任しましたか。

(When did (he) inaugurate (as it)?)

Recognition of antecedent type (pattern3)

- Search words which co-occurred with the topicalized element of follow-up question
- We use co-occurrence data of EDR dictionary

Ex.) アメリカの大統領は誰ですか。
(Who is the president of America?)
(Φ: *modifier* の)首都はどこですか。
(searched data) (Where is the capital (of the country)?)

フランスの首都
(capital of France)
中国の首都
(capital of China)
ロシアの首都
(capital of Russia)

フランス (France)
中国 (China) → country
ロシア (Russia)
Type: country

Selection of the antecedent (pattern3)

- Same as pattern1

Ex.)

○ country × position × human

アメリカの大統領は誰ですか。(ブッシュ)
(Who is the president of America?) (Bush)

(Φ: modifier の) 国務長官はだれですか。
(Where is a minister of state (of America)?)

Evaluation

- Evaluation data
 - 310 pairs of questions
 - A pair of questions consists of a follow-up question of Formal Run and its just previous question of Reference Run
- Correctness have been judged by human
- When the antecedent is the answer of previous question, we have used “<ANS>”

Evaluation data

F-Run data

R-Run data

Q1-1 → R1-1 (same as Q1-1)

Q1-2 → R1-2 (no ellipsis)

Q1-3 → R1-3 (no ellipsis)

Q1-4 → R1-4 (no ellipsis)

Results

- **Reference of pronoun (pattern1)**
 - System classified 88 of 310 questions in this pattern
 - The classification was 100% correct
 - Ellipsis handling succeeded in 13.6% (12 questions)
- **An obligatory case element is zero anaphora (pattern2)**
 - System classified 158 of 310 questions in this pattern
 - The classification was 66.5% correct
 - Ellipsis handling succeeded in 7.6% (8 questions)
- **Modifier or modificand is zero anaphora (pattern3)**
 - System classified 64 of 310 questions in this pattern
 - The classification was 68.8% correct
 - Ellipsis handling succeeded in 9.1% (4 questions)

Classification of failure reasons

- Failure of classification of reference pattern
 - System used wrong verbs . . . 29
 - All obligatory cases of verb was filled and other element was omitted . . . 22
- Failure of recognition of antecedent type
 - The verb information was not in the EDR dictionary . . . 35
 - Lack of rule for reference of pronoun . . . 17
- Failure of selection of antecedent
 - System failed to recognize type of candidate of antecedent of previous question . . . 79
 - System failed to decide to range of taken word . . . 21

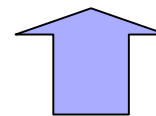
Failure reasons (1)

■ Verb case elements:

- Use of Japanese delexical verb such as “いる (iru)”, “なる (naru)” causes case element filling.
- No information in dictionary

■ Morphological analysis failure

- “そこで (sokode)” should be “そこ (soko)” and “で (de)”.



pronoun

Failure reasons (2)

- Lack of pronoun patterns
 - この宇宙ステーション (**this** space station)
- Case information handling error
- Lack of co-occurrence information
- Passive verb expression
- Multiple candidates

Example of failure case

- System failed to recognize that the answer of first question was “team name”.

Q1: 静岡スタジアム「エコパ」のこけら落としで清水エスパルスと対戦したのはどこですか。(ジュビロ磐田)

(What team is the team which have a game with *Shimizu S-PULSE* at the first game of *Shizuoka Stadium “ESUPA”*?)

Q2: そのチームの設立はいつですか。

(When the team was formed?)

Q2': 清水エスパルスの設立はいつですか。

✗ (When the *Shimizu S-PULSE* was formed?)

Conclusions

- Classification of reference pattern into three types and their reference resolution method
- Performance was not good in our current implementation.
- Major failing reason was lack of word information.



Thank you.