NTT's Question Answering System for NTCIR-6 QAC-4

Ryuichiro Higashinaka
Hideki Isozaki
NTT Communication Science Labs.
NTT Corporation

NTT

Overview

- Our two systems:
  - NCQAW-1 and NCQAW-2
    (NTT CS Labs' QA System for 'Why' Questions)
- Features
  - Focus on 'why' questions
  - Machine learning approach for 'why' and 'how' questions
  - Pattern-based approach for 'definition' questions
- Achieve good performance on 'why' and 'definition' questions

NTT

Systems

- NCQAW-1
  - ML-based approach for ‘why’ and ‘how’
  - Pattern-based approach for ‘definition’
  - Uses SAIQA-QAC2 (our factoid QA system) for other question types
  - Question type analysis is based on rules
- NCQAW-2
  - Same as NCQAW-1 except that ‘why’ and ‘how’ questions are handled by rules

NTT

Results

Improvement by ML-based approach

- NCQAW-1

question type	all	A	B	C	D	no output
definition	24	9	0	0	11	4
other	12	1	0	0	2	9
why	38	**11**	**6**	0	18	3
how	26	1	1	1	7	16
total	100	22	7	1	38	32

Could not answer many 'how' questions because of question analysis failures

- NCQAW-2

why	38	**5**	**3**	1	25	4
how	26	1	0	1	8	16
total	100	16	3	2	46	33

NTT

‘Why’ questions

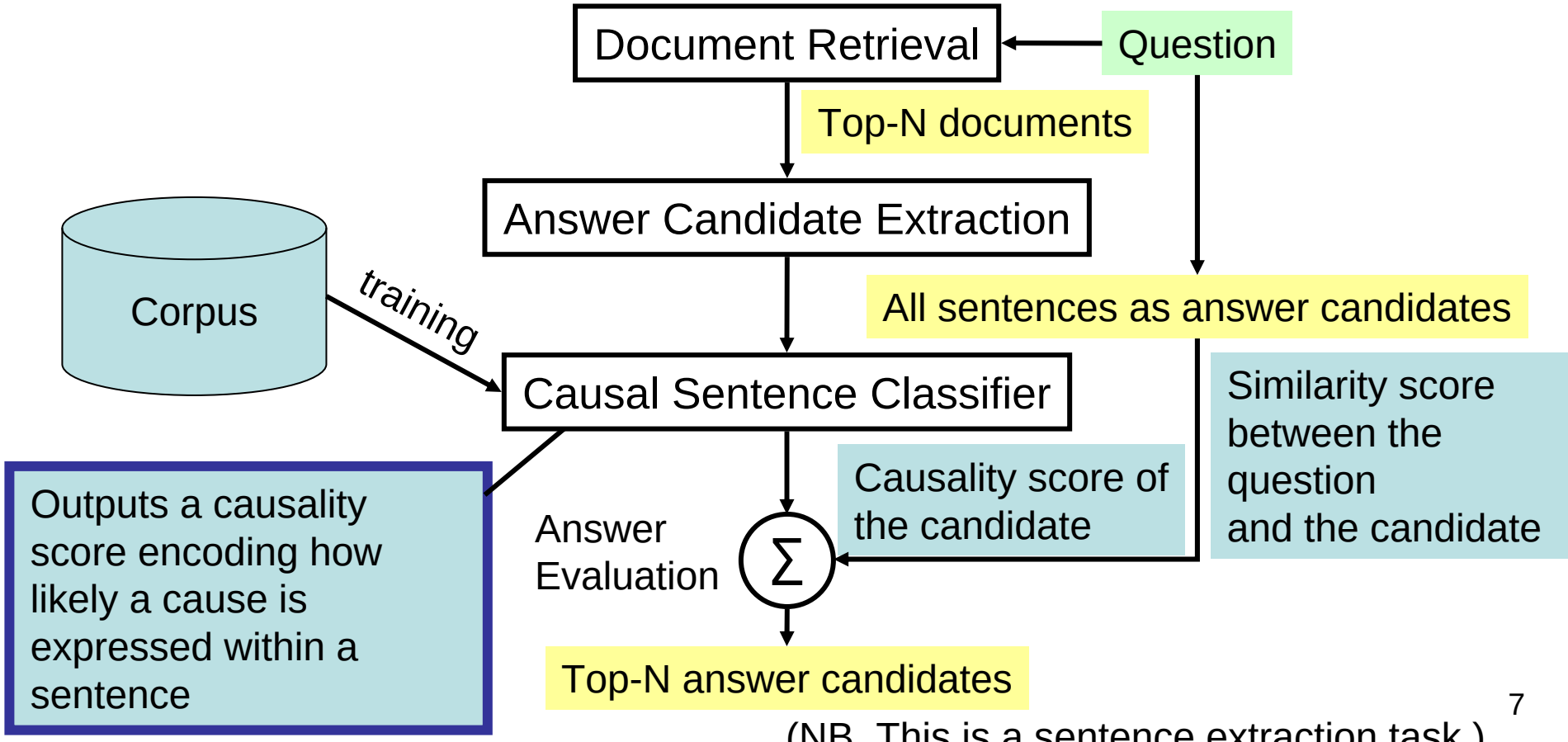
- There are few systems for answering open-domain ‘why’ questions
- Previous approach
  - Extract *causal sentences* (sentences having causal expressions) by hand-crafted rules as answer candidates
    (e.g., using cue words such as ‘tame’, ‘node’ etc.)
  - Rank the candidates by their similarity to the question
- Systems based on the approach
  - System by Morooka and Fukumoto (2006)
  - NCQAW-2

Problem

- Hand-crafted rules are costly to make
  - Cue words are not always reliable
    - Only 6-7 % of words before ‘de (by)’ are causes (Abekawa and Okumura, 2004)
  - Difficult to cover all causal expressions by hand
- Difficult to express degree of causality
  - Some expressions are more strongly expressing causality than others
  - ‘no riyuude (by reason of) ’
    vs. ‘kara (from)’, ‘tame (for)’, ‘de (by)’, etc.

Approach

- Adopt a machine learning approach to learn a *causal sentence classifier*



(NB. This is a sentence extraction task.)

NTT

Formulation

(1) $\mathrm{causal}_{\mathrm{why}}(C)$ ⟸ Causality score output by the causal sentence classifier

(2) $\mathrm{sim}_{\mathrm{why}}(S) = \sum_{w \in Q(S)} \mathrm{idf}(w)$ ⟸ Similarity score

(Sum of IDF of query terms within the candidate)

$$\mathrm{sim}'_{\mathrm{why}}(C) = 1/(1 + \exp(-\mathrm{sim}_{\mathrm{why}}(C)))$$

N.B., Similarity score is normalized by the sigmoid function

Final score of an answer candidate:

$$\mathrm{candscore}_{\mathrm{why}}(C) = \mathrm{causal}_{\mathrm{why}}(C) + \mathrm{sim}'_{\mathrm{why}}(C)$$

NTT

Causal Sentence Classifier

- Use EDR Japanese corpus for training
  - Has annotation of ‘cause’ relation
  - 8,064 sentences with ‘cause’ out of 0.2M total sents.
  - Sentences with ‘cause’ → positive examples
    without ‘cause’ → negative examples

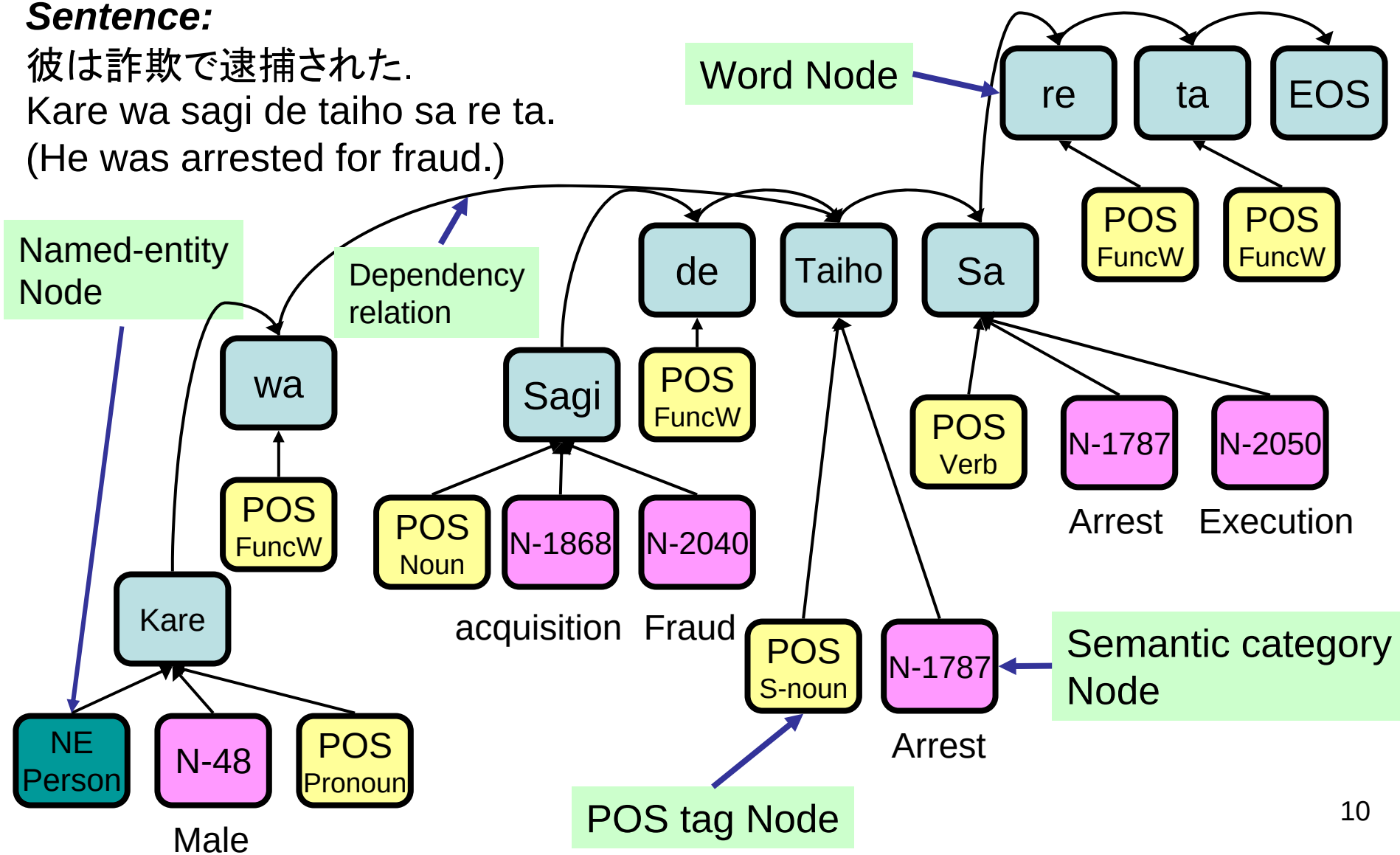
- Train a classifier by ***BACT***
  - In sentence classification tasks,
    lexical, syntactic, and semantic features are useful

    → Adopt tree feature representation of a sentence

    → BACT: a boosting algorithm for classifying trees
    (uses existence of sub-trees as weak learners)

Tree feature representation of a sentence

Sentence:
彼は詐欺で逮捕された.
Kare wa sagi de taiho sa re ta.
(He was arrested for fraud.)



NTT

Evaluation of the classifier

	rule-based	**BACT**
Accuracy	92.1%	96.0%
Precision	11.0%	**41.0%**
Recall	14.7%	4.2%
F-measure	12.6%	7.6%

(Rule-based: our implementation of Morooka and Fukumoto's causal sentence extraction rules)

- Higher precision achieved by BACT
- Precision may be important from NCQAW-1'results
- F can be raised to ca. 25% by feature engineering

NTT

Acquired Salient Patterns

rank	String-encoded subtree	α
1	EOS 。 が	0.032
2	で 疑い (by suspicion of)	0.022
3	によって (by)	0.008
4	EOS 。 が 助詞-格助詞-一般	0.007
5	ため (for)	0.007
6	による (because of)	0.006
7	により (because of)	0.006
8	ので (because of)	0.006
9	から (from, because of)	0.005
10	。	0.005
11	動詞-自立 (verb)	0.004
12	2419 [types of illness]	0.004
13	。 、 の (of)	0.003
14	、 で 助動詞 (by)	0.003
15	から こと (from the fact that ...)	0.003

α : Weights given to each pattern

- Acquired 178 patterns
- Cover most of expressions easily conceivable by humans
- Plus, include many other complex expressions
- Automatically disambiguating cue words e.g., 'de' (see below)

で 中 2639 [intermediate/middle]	-0.001
。 で 助詞-格助詞-一般) の 助詞-連体化	-0.001
の で 助詞-格助詞-一般	-0.002
は で 助詞-格助詞-一般	-0.004

NTT

Examples of Answers

- NCQAW-2 extracts sentences with a strong cue
  - QAC4-00030: What is the purpose of the green power marketing?
  - A: さらに、原子力や石油への依存度を少しでも下げるため、電源の多様化を図るのは時代の流れだろう。
- NCQAW-1 can find answers without such cues
  - QAC4-00026: Why is the movable weir needed in the Yoshino river?
  - A: 可動堰化計画は、第十堰の老朽化が進んで洪水時の障害になり、壊れる危険もある、などとして持ち上がった。

NTT

‘How’ questions

- Same process as ‘why’ questions
- ‘condition’ relation is used instead of ‘cause’ to train a sentence classifier
  - sentences having ‘condition’ may have answers within the sentence
    - Q: 聖火が消えたらどうしますか？
      (How do we cope when the Olympic flame burns out?)
    - A: 聖火が消えたら再点火します.
      (If the Olympic flame burns out, it is reignited.)

Due to question type analysis failures,
most of ‘how’ questions were not answered at all

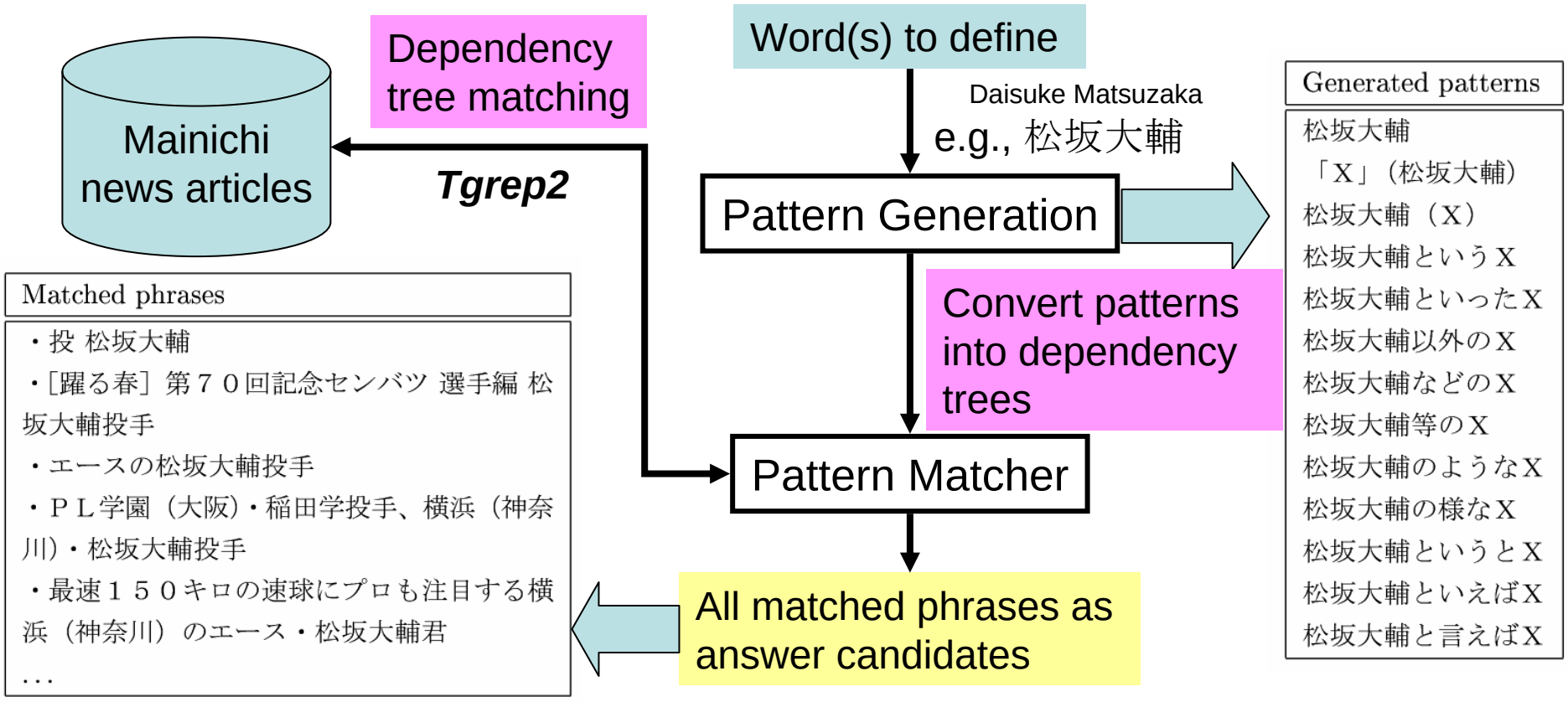
NTT

‘Definition’ questions

- There are many systems for answering ‘definition’ questions (e.g., in TREC)
- Common approach
  - Extract descriptive sentences/nuggets as answer candidates using patterns or ML
  - Rank the candidates using importance of keywords within the candidate
    (i.e., the more keywords, the better)
- We adopt a similar approach

Answer candidate extraction

- Use patterns to find descriptive phrases
- Perform dependency tree matching to obtain phrases with all their modifiers



NTT

Answer Evaluation

- Answer candidates are ranked based on importance of words within the candidates
- Importance of words:

$$\mathrm{wordscore}_{\mathrm{def}}(w) = \log(\mathrm{tf}(w; \{C_i\}))$$

(Term frequency within all answer candidates)

- Score of an answer candidate:

$$\mathrm{candscore}_{\mathrm{def}}(C) = \sum_{w \in \mathrm{CW}(C)} \mathrm{wordscore}_{\mathrm{def}}(w)$$

(Sum of wordscores within the answer candidate)

NTT

Examples of answers

- QAC4-00018: スケルトンとはどのような競技ですか。

  (What is Skelton, the competitive sport?)

  - A: 54年ぶりに冬季五輪の正式種目として復活する、そり競技のスケルトン。
  - A: うつぶせでソリに乗ってボブスレーやリュージュと同じコースを滑走する氷上競技「スケルトン」
- QAC4-00034: 「スポット傍受」とはどういったものですか。

  (What is the spot wiretapping?)

  - A: 通話内容が犯罪と関係あるか試し聴きする「スポット傍受」
  - A: 会話の冒頭（ぼうとう）を試験的に傍受する「スポット傍受」

- Adnominal/adverbial phrases by including modifiers
- Achieving good performance (9 A-rated answers/24)

NTT

Results

Improvement by ML-based approach

- NCQAW-1

question type	all	A	B	C	D	no output
definition	24	9	0	0	11	4
other	12	1	0	0	2	9
why	38	**11**	**6**	0	18	3
how	26	1	1	1	7	16
	100	22	7	1	38	32

Good performance by finding adnominal/ adverbial phrases by dep. tree matching

Could not answer many ‘how’ questions because of question analysis failures

why	38	**5**	**3**	1	25	4
how	26	1	0	1	8	16
total	100	16	3	2	46	33

Conclusion

- Our two systems:
  - NCQAW-1 and NCQAW-2
- Features
  - Focus on ‘why’ questions
  - Machine learning approach for ‘why’ and ‘how’ questions
  - Pattern-based approach for ‘definition’ questions
- Achieved good performance on ‘why’ and ‘definition’ questions