Answering Contextual Questions Based on the Cohesion with Knowledge ⋆

Tatsunori Mori, Shinpei Kawaguchi, and Madoka Ishioroshi

Graduate School of Environment and Information Sciences
Yokohama National University
79-7 Tokiwadai, Hodogaya, Yokohama 240-8501, Japan
{mori,kawaguchi,ishioroshi}@forest.eis.ynu.ac.jp

Abstract. In this paper, we propose a Japanese question-answering (QA) system to answer contextual questions using a Japanese non-contextual QA system. The contextual questions usually contain reference expressions to refer to previous questions and their answers. We address the reference resolution in contextual questions by finding the interpretation of references so as to maximize the cohesion with knowledge. We utilize the appropriateness of the answer candidate obtained from the non-contextual QA system as the degree of the cohesion. The experimental results show that the proposed method is effective to disambiguate the interpretation of contextual questions.

1 Introduction

In recent years, *contextual question-answering (QA) systems* have gained attention as a new technology to access information. In this paper, we propose a method to construct a Japanese contextual QA system using an existing Japanese non-contextual QA system[1]. Although a contextual question generally contains reference expressions[2], we expect that the non-contextual QA system will be able to find answers for such a question if reference expressions are appropriately completed along with their antecedents.

The completion of a question may be performed in the following steps: (1) detect reference expressions, and then (2) find an antecedent for each reference expression. However, there are ambiguities in these steps, and we may have multiple interpretations, namely, multiple *completed question candidates*, for one question. In the research area of discourse understanding, there exist many studies of the reference resolution in terms of *the cohesion with the context*. The

⋆ This study was partially supported by Grant-in-Aid for Scientific Research on Priority Areas (No.18049031) and Scientific Research (C) (No.17500062) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

[1] We define *a contextual question* as "a question that may have references to the context, i.e., previously asked questions and their answers."

[2] In this paper, we use the term "reference expressions" to refer to not only reference expressions themselves, such as demonstratives and pronouns, but also ellipses like zero pronouns. Zero pronouns are ellipses of obligatory case elements.

centering theory is one of the most widely used methods [1]. This type of reference resolution attempts to find an optimal interpretation so as to maximize the cohesion between a newly introduced sentence and the context. Such a method would definitely work in many cases, but it does not provide the method to resolve the ambiguity in the step (1).

In this paper, we propose another approach. It is the reference resolution in terms of *the cohesion with knowledge*. It is based on the fact that the QA system can refer to not only the context of the dialogue but also the knowledge base (i.e. a large text document collection). It is noteworthy that "answering a question" can be regarded as finding an object, i.e., an answer, whose context in the knowledge base is coherent with the question. Therefore, the cohesion with knowledge may be one of the promising criteria for finding the best interpretation of the question. Here, we hypothesize that the degree of cohesion with knowledge is analogous to *the appropriateness of the answer candidate* for each completed question candidate. The completed question candidate with the most appropriate answer can be accordingly considered as the best interpretation of the (original) question.

2 Related work

2.1 Contextual question answering

The contextual QA was introduced as a subtask of the QA track in TREC 2001. However, Voorhees [2] summed up the evaluation as follows: "the first question in a series defined a small enough subset of documents such that the results were dominated by whether the system could answer the particular type of the current question, rather than by the system's ability to track context." For this reason, this task was excluded from all subsequent TRECs. On the other hand, a context task has been employed as a subtask of QAC in NTCIR, which is a series of evaluation workshops organized by the National Institute of Informatics, Japan. Kato et al. [3] summarized the lessons from the context task of TREC QA as follows: (1) the number of questions in a series is relatively small, and (2) there is no topic shift in a series. They prepared the test sets for NTCIR QAC according to the lessons, that is, (1) a series is relatively long, about seven questions (QAC3), and (2) two types of series are introduced, namely, *the gathering type* and *the browsing type*. A question series of the gathering type contains questions that are related to one topic. On the other hand, in a series of the browsing type, the topic varies as the dialogue progresses.

The approaches of the systems participating in the context tasks in the NTCIR QAC are mainly based on the cohesion with the context. In general, the approaches are classified into two types. The first type is based on the effort involved in the document/passage retrieval. It expands the query submitted to the IR system with the words/phrases that appeared in the previously asked questions[4]. The second type of approach is based on the completion of questions by resolving reference expressions[5]. One completed question is submitted to a non-contextual QA system. The method that we propose in this paper is similar to the second approach. However, our approach is based on the cohesion with knowledge as well as the cohesion with the context.

2.2 Reference resolution

Detection of reference expressions: The detection of zero pronouns is particularly very important and is studied from various viewpoints. One of the most widely used methods is detection using a case-frame dictionary. A case-frame dictionary is used to find unoccupied cases in a sentence.
Identification of antecedents: Nariyama [6] proposed a modified version of the centering theory[1] in order to resolve Japanese zero pronouns. It utilizes a "*salient referent list (SRL)*" that pools all overt case elements that have appeared up to the sentence in question. If a new case element appears with a case marker identical to that of another case element already existing in the SRL, the new case element takes its place because of recency. In an SRL, the case elements are listed in the following order of salience: Topic $>$ Nominative $>$ Dative $>$ Accusative $>$ Others. A zero pronoun is resolved by selecting the most salient case element in the SRL.

3 Proposed method

Figure 1 shows the overview of the proposed method. The method generates an answer list for each question in a given series by using the following procedure. It should be noted that the non-contextual QA system can perform list-type question answering, as described in Section 3.3. It is the task in which a system is requested to enumerate all correct answers.

Input A new question and a list of antecedent candidates. The list is initialized to an empty list for the first question.
Output An answer list and an updated antecedent candidate list.
Procedure

1. Detect reference expressions including zero pronouns in the new question using a case frame dictionary, and then generate question candidates with the zero pronouns. Generally we obtain multiple candidates because a single verb may have multiple case-frame entries of case frame.
2. Find antecedent candidates for reference expressions according to a selected strategy for completing reference expressions. We proposed three strategies: *CRE-C*, a strategy based on a modified version of the SRL-based centering theory; *CRE-H*, a heuristic strategy in which the characteristics of a series of questions are taken into account; and *CRE-A*, a strategy that adopts all possible noun phrases.
3. Generate all possible completed question candidates using the results of Step 2. Then, select the M-best completed question candidates according to the semantic consistency in the reference resolution.

4. Submit completed question candidates to the non-contextual QA system, and obtain a list of answer candidates for each question candidate. Each answer candidate in the list is associated with its answer score. From the answer scores, calculate the appropriateness of the list. We propose two measures of appropriateness: *AM-D*, a measure defined in terms of the distribution of scores of answer candidates; and *AM-M*, the maximum score of the answer candidates in the list.
5. Provide the most appropriate answer list as the final output.
6. Using the completed question candidate that provides the most appropriate answer list, update the list of antecedent candidates according to a selected strategy for completing the reference expressions, and output the list for the next question.

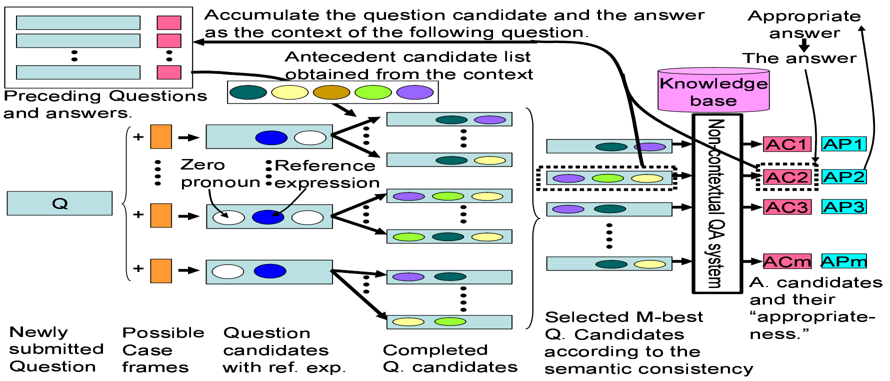

Fig. 1. Overview of the proposed method

3.1 Example

Before we explain the details of each step, we describe the flow of the procedure using the following example series of questions. In this example, we adopt Strategy CRE-C and Measure AM-D as the strategy for completing reference expressions and the appropriateness measure of the answer lists, respectively.

(1) a.

Shizuoka-sutajiamu "ECOPA"-no	kokeraotoshi-de	Shimizu-S-Pulse-no	
Shizuoka-studium "ECOPA"-REL	opening-AT	Shimizu-S-Pulse-REL	
taisen-aite-wa	doko	de-su-ka	
matched-team-TOP	where/what	BE-POL-INTERROG	

Which team played a match against the Shimizu S-Pulse at the opening game of Shizuoka Stadium "ECOPA"?

b.

sono-shuryoku-senshu-ni	dare-ga	i-masu-ka
its-leading-player-AS	who-NOM	exist-POL-INTERROG

Who is its leading player?

Since Question (1a) is the first in the series, it is submitted to a non-contextual QA system without any modification, and obtained the following answer list. The system also assigns semantic categories[3] to each answer in the list according to a thesaurus.

(2) a. {"Iwata" (CITY, STATION)}[4].

In Step 6 of the procedure, the system updates the list of antecedent candidates. When CRE-C is adopted, the list would be the following SRL obtained from the first question and its answer list[5]:

(3) [top:{"Iwata" (CITY, STATION), "Shimizu-S-Pulse-no taisen-aite" (OPPONENT, COMPANION, OTHER PERSON)}, nom:{}, dat:{}, acc:{}, other:{}].

As shown in this example, if the interrogative is the expression to be stored in the SRL, we replace it with the answer list. Each expression in the SRL is also assigned semantic categories.

Next, the system receives the second question (1b). Since the question has a context, the system tries to detect the reference expressions, as described in Section 2.2. First, the system checks whether the question contains demonstratives or pronouns, and it finds a demonstrative adjective called "sono." Next, the system tries to detect zero pronouns by using the following steps: (i) look up the verb of the question in a case-frame dictionary in order to obtain case frames, and (ii) detect unoccupied cases in the question as zero pronouns. The system also obtains the information of semantic categories for unoccupied cases from the case frames. In the case of Question (1b), the system obtains the following case frame for the verb "iru" (exist)[6], which satisfies the selectional restriction in terms of the semantic category, and detects that there are no zero pronouns.

(4) NP1(PLACE, PERSON)-ni NP2(PERSON)-ga iru

The antecedent candidates for the demonstrative adjective are obtained from SRL according to the order in the list. Since there are two candidates at the top of the list, we have the following two completed question candidates:

(5) a. Shimizu-S-Pulse-no taisen-aite-no shuryoku-senshu-ni
Shimizu-S-Pulse-REL matched-team-REL leading-player-AS
dare-ga i-masu-ka
who-NOM exist-POL-INTERROG
Who is the leading player of the team that played against the Shimizu S-Pulse?

b. Iwata-no shuryoku-senshu-ni dare-ga i-masu-ka
Iwata-REL leading-player-AS who-NOM exist-POL-INTERROG
Who is the leading player of Iwata?

The system selects the M-best completed question candidates according to the semantic consistency in the reference resolution, as described in Section 3.7, and submits them to the non-contextual QA system in order to obtain an answer list and a value of the appropriateness of the list for each question candidate. In

[3] Semantic categories are shown in parentheses
[4] The full name of the team is "Jubilo Iwata," and "Iwata" is the name of city.
[5] In the Japanese sentence "NP1 wa NP2 da," the copula "da" represents that NP2 is equivalent to NP1. We accordingly treat not only NP1 but also NP2 as the topic.
[6] The expression "iru" is the root form of the verb "i" in Question (1b)

the example, we have the following results: {"NANAMI Hiroshi[7]","YAMADA Kosuke[8]"} and 0.019 for Question candidate (5a), and {"NANAMI Hiroshi"} and 0.031 for (5b). Since we hypothesize that the question candidate whose answer list has the highest value of appropriateness is the best interpretation in terms of the cohesion with knowledge, the system outputs the latter answer list as the answer for Question (1b).

After the process of the second question, the system updates the SRL and proceeds to the processing of the next question.

3.2 Non-contextual Japanese QA system

The non-contextual Japanese QA system used is a Japanese real-time QA system based on the study by Mori[7]. Mori reported that the MRR (mean reciprocal rank) of the system is 0.516 for the test set of NTCIR-3 QAC2. It treats each morpheme in retrieved documents as a seed of an answer candidate and assigns it a score that represents the appropriateness for an answer. Under the assumption that the morpheme is the answer, the score is calculated as the degree of matching between the question and the sentence where the morpheme appears. In his current implementation, the score is a linear combination of the following sub-scores: the number of shared character bigrams, the number of shared words, the degree of case matching, the degree of matching between dependency structures, and the degree of matching between the NE type of the morpheme and the type of question.

3.3 List-type QA processing in non-contextual Japanese QA

We also introduce a list-type QA processing proposed by Ishioroshi et al. [8]. They assume that the distribution of answer scores contains a mixture of two normal distributions, $\phi_p(x; \mu_p, \sigma_p)$ and $\phi_n(x; \mu_n, \sigma_n)$, i.e., those of the correct answers and incorrect answers, where μ and σ are the average and the standard deviation, respectively. Under these assumptions, the correct answers may be separated from the mixture of the distributions by using the EM algorithm. Figure 2 shows an example of the score distribution in the case that the score distribution of the correct answers is separable from that of the wrong answers.

3.4 Detecting of reference expressions

Our method treats the three types of reference expressions — (i) demonstratives and pronouns, (ii) zero pronouns, and (iii) ellipsis of the adnominal modifier "NP_1-NO" in a noun phrase "NP_1-NO NP_2 (NP_2 of NP_1)."

The detection of the reference expressions of type (i) is not difficult because they appear explicitly. With regard to type (ii), we employ an existing method

[7] A correct answer.

[8] A wrong answer.

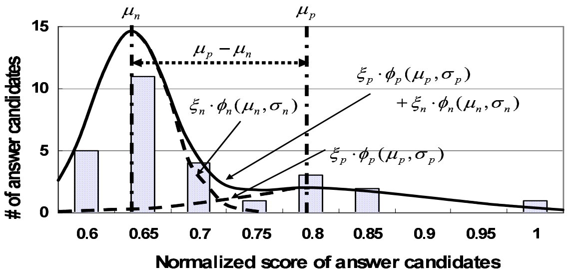

Fig. 2. An example distribution of answer scores

based on a case-frame dictionary as described in Section 2.2. We use "nihon-go goi taikei" (a Japanese lexicon) as the case-frame dictionary. For type (iii), we adopt Takei's method [9].

If no reference expression is detected in a question, the system assumes that a topic phrase is omitted in the question and introduces a zero topic phrase in order to force the question to have a relation with the context.

3.5 Finding antecedent candidates of reference expressions

Strategy CRE-C: This strategy is based on a modified version of Nariyama's SRL-based centering theory. The list of antecedent candidates is SRL itself. This method is different from Nariyama's method in the following manner: (a) demonstratives and pronouns are resolved before zero pronouns, and (b) a zero topic phrase may refer to all possible case elements in SRL.

Strategy CRE-H and Strategy CRE-A: For these strategies, the list of antecedent candidates is maintained as described in Section 3.9. The system may select any element from the list for each reference expression.

3.6 Narrowing down antecedent candidates using selectional restriction

According to the selectional restriction, for each reference expression in the current question, the system filters out inappropriate candidates in the antecedent candidates that are obtained by one of the strategies described in Section 3.5. With respect to Strategy CRE-C, the system selects the candidate that is at the highest rank in SRL among the appropriate candidates. The selectional restriction is based on the similarity $sim(r, a)$ between the semantic categories of the antecedent candidates and reference expressions, a and r, defined in the thesaurus("nihon-go goi taikei"). The similarity is calculated by the following equation [10]:

$$sim(r, a) = \begin{cases} \frac{2 \times L_{ra}}{l_r + l_a} & \text{if } a \not\prec r \\ 1 & \text{if } a \prec r \end{cases} \quad (1)$$

where l_r and l_a are the depths of the categories r and a in the thesaurus respectively, and L_{ra} is the depth of the lowest common ancestor of r and a. The symbol "$\prec$" represents the subsumption relation. We determine a threshold value of the similarity Th_{sim}, and filter out each antecedent candidate whose similarity is less than the threshold value.

3.7 Generating completed question candidates and narrowing them down

By completing each reference expression in the current question with all the possible antecedent candidates, the system generates all the possible candidates of the completed question. However, this process may generate many question candidates, and the non-contextual QA systems may take a very long time to process them. Therefore, we introduce a measure for a completed sentence in terms of *the degree of consistency in the reference resolution*, and select the M-best question candidates by using the measure. We defined the degree $C(Q)$ of a question candidate Q in Equation (2).

$$C(Q) = \sum_{<r_i, a_i> \in resolv(Q)} \frac{c_1(r_i, a_i)}{|resolv(Q)|} \tag{2}$$

$$c_1(r, a) = \begin{cases} 1 & \text{if } a \prec r \wedge NE(a) \\ 1.5 & \text{if } a \prec r \wedge \neg NE(a) \\ sim(r, a) & \text{if } a \not\prec r \end{cases}$$

where $resolv(Q)$ is the set of pairs of a reference expression and its antecedent in the question Q, and $NE(a)$ is true if a is a named entity.

3.8 Finding the most appropriate answer list by using the non-contextual QA system

The selected question candidates are submitted to the non-contextual QA system. The completed question candidate with the most appropriate answer may be considered as the best interpretation of the original question. We propose the following two methods as the appropriateness measure.

Measure AM-D: The appropriateness of an answer candidate list is assumed to be measured by $\mu_p - \mu_n$ in Figure 2. Some of the candidates of question completion may not be consistent with the knowledge base. In such cases, the scores of highly ranked answer candidates are not very high and have almost the same distribution as that of the lower ranked candidates. Conversely, if the value $\mu_p - \mu_n$ is relatively large, we can expect that an appropriate list is obtained.

Measure AM-M: The appropriateness of an answer candidate list is assumed to be measured by the maximum score of the answer candidates in the list. It is based on the fact that the score of an answer is calculated according to the coherence between the question and the context of the answer.

3.9 Updating the list of antecedent candidates

By using the completed question candidate that provides the most appropriate list of answer candidates, the system updates the list of antecedent candidates according to a selected strategy for completing the reference expressions.

Strategy CRE-C: The list is updated in the same manner as the SRL. The difference is the treatment of the interrogatives. An interrogative is replaced with its answer list before the SRL is updated.

Strategy CRE-H: The list of antecedent candidates is maintained so as to have the following elements: (a) all the case elements in the current completed question, (b) all the phrases in the answer list of the current completed question, and (c) topic phrases, that is, all the case elements in the first question. Here, it should be noted that the system extracts the case elements not from the original current question but from the *completed* current question. The case elements in the questions before the current question may be retained in the completed current question if the question continues to refer to them.

Strategy CRE-A: We just adopt all the case elements in all the completed questions thus far.

4 Experimental results

We evaluate the proposed systems in terms of the accuracy of the question answering by using the test set of NTCIR-5 QAC3 [11]. The test set comprises 50 series and 360 questions. In these series, 35 series (253 questions) are of the gathering type and 15 series (107 questions) are of the browsing type. It should be noted that the systems are not allowed to use the series type for answering the questions. The document collection as the knowledge source consists of all (Japanese) articles in the Mainichi Shimbun newspaper and Yomiuri Shimbun newspaper published in 2000 and 2001. The parameters are tuned with the test set of NTCIR-4 QAC2. The threshold value for the selectional restriction Th_{sim} is 0.54, and the number M of completed question candidates to be selected is 20. For measuring the accuracy of the list-type question answering, we use the mean of the modified F measure $MMF1$ defined by Kato et al. [11].

We also prepare systems that do not use cohesion with knowledge ("No-CK" in the following table) as baseline systems. These systems adopt the completed answer candidate that has the largest value of the degree of consistency in the reference resolution, which was described in Section 3.7. As an upper limit of accuracy, we also evaluate the non-contextual QA system with questions whose reference expressions are manually resolved.

The experimental results are shown in Table 1.

5 Discussion

5.1 Overall performance of question answering

With regard to the effect of the introduction of cohesion with knowledge, as shown in Table 1, both AM-D and AM-M, which utilize the cohesion with knowl-

Table 1. Evaluation of Question Answering in MMF1

	(a) All Series			(b) Gathering Type			(c) Browsing Type		
	AM-D	AM-M	No-CK	AM-D	AM-M	No-CK	AM-D	AM-M	No-CK
CRE-C	**0.174**	0.166	0.164	0.177	0.178	0.168	**0.166**	0.137	0.146
CRE-H	0.147	0.147	0.136	0.142	0.168	0.141	0.134	0.097	0.123
CRE-A	0.169	0.157	0.133	**0.180**	0.166	0.146	0.146	0.135	0.105
Manually resolved	0.242								

edge, outperform the baseline No-CK, which is only based on the degree of consistency in the reference resolution. In particular, the appropriateness measure AM-D works well for all strategies for completing the reference expressions.

With regard to the differences between the strategies for completing the reference expressions, Strategy CRE-C exhibits a better performance than the others. However, the difference between CRE-C and CRE-A is not significant when the measure AM-D is used.

In order to investigate the situation in further detail, let us compare (b) and (c) in Table 1. The systems based on Strategy CRE-C are stable over both types of series. The combination of CRE-C and AM-D is more robust even when topic shifts occur. The systems with the measure AM-M seem to have a better performance in the gathering type series. The reason for this is as follows. Because of the composition of the answer score, the score of a longer question tends to be larger than that of a shorter question. Consequently, the use of the measure promotes the selection of the case frames that have a larger number of case elements. As a result, the system tends to select question candidates that are more cohesive with the context. However, the systems easily fail to track the shift of topic, as shown in Table 1 (c). The strategies CRE-H and CRE-A have good performance for the series of the gathering type, but are not good at the series of the browsing type because they could not track topic shifts properly.

5.2 Failure analysis

A detailed analysis of success and failure is summarized in Table 2. In this table, "Success" implies that the generated answer list contains at least one correct answer. The other cases are "Failure." Among the success cases, there are many cases where the reference resolution fails but the system successfully finds the answers for the question. This implies that the introduction of expressions of the context into the current question has a positive effect on the performance of question answering even if the accuracy of the reference resolution is insufficient. One of the reasons for this is that these newly introduced expressions may work well in the early stages of question answering, such as document/passage retrieval.

The main reason for the failure lies in the stage of updating the list of antecedent candidates in the CRE-C. The failure is caused by, at least, the following reasons: (1) failure in finding correct answers for some previous questions, (2) failure in finding the appropriate antecedents for the reference expressions in the list of antecedent candidates.

Table 2. Detailed analysis of success and failure

Method	Success		Failure				
	Res. OK Ans. OK	Res. NG Ans. OK	Ante. NG	Q. Gen. NG	Q. Sel. NG	Ans. NG	Others
CRE-C + AM-D	11.0% (34)	13.5% (42)	26.5% (82)	21.0% (65)	9.7% (30)	18.1% (56)	0.3% (1)
CRE-C + AM-M	9.4% (29)	13.2% (41)	29.4% (91)	20.3% (63)	11.0% (34)	16.5% (51)	0.3% (1)
CRE-C + No-CK	9.4% (29)	11.9% (37)	29.4% (91)	21.0% (65)	10.6% (33)	17.4% (54)	0.3% (1)
CRE-A + AM-D	11.3% (35)	11.3% (35)	17.4% (54)	14.5% (45)	29.4% (91)	15.2% (47)	1.0% (3)
CRE-A + AM-M	10.6% (33)	8.7% (27)	18.1% (56)	13.5% (42)	37.1% (115)	11.0% (34)	1.0% (3)
CRE-A + No-CK	6.8% (21)	9.0% (28)	19.4% (60)	15.2% (47)	35.2% (109)	13.5% (42)	1.0% (3)

Res.: reference resolution
Ans.: question answering by the non-contextual system
Ante.: appropriate antecedent
Q. Gen.: generation of completed question candidates
Q. Sel.: selection of an appropriate question candidate

The number of failures in the updating stage in CRE-A is relatively low because the restriction on the list of antecedent candidates is relatively weak and the list may have more candidates than CRE-C. On the other hand, there are many failures in the stage involving the selection of an appropriate question candidate. Since the cohesion with knowledge is taken into account in this stage, the ratio of failures in this stage is closely related to the effectiveness of the cohesion with knowledge. However, we can *not* jump to the conclusion that the proposed method is not effective because the method based on the cohesion with knowledge works correctly only when the non-contextual QA system can find at least one correct answer in the knowledge base.

In order to estimate the ability of disambiguation based on the cohesion with knowledge more accurately, we investigate the accuracy of answering the questions that satisfy the following conditions: (a) each of them has multiple completed question candidates, and at least one candidate is a proper interpretation in the given context, and (b) the non-contextual QA system can find at least one correct answer for at least one of the completed answer candidates with the correct interpretation. The result shown in Table 3 implies that the use of cohesion with knowledge significantly improves the accuracy. By comparing Table 3 with Table 1, we also find that the accuracy may be improved when the non-contextual QA system is able to find appropriate answers.

Table 3. MMF1 values when the cohesion with knowledge may work correctly

	AM-D	AM-M	No-CK
CRE-C	0.478	0.437	0.399
CRE-A	0.436	0.374	0.334

6 Conclusion

In this paper, we introduced the notion of "*cohesion with knowledge*" and on its basis, proposed a question-answering system to answer contextual questions using a non-contextual QA system. Experimental results showed that the system works effectively under the combination of the strategy based on the SRL-based centering theory, i.e., CRE-C and the appropriateness measure of an answer that is defined in terms of the score distribution of the answer candidates, i.e., AM-D.

According to the failure analysis, the main reason for the failure was that the appropriate antecedents of the reference expressions in the current question do not appear in the list of antecedent candidates. Therefore, further improvement in the non-contextual QA system is required.

References

1. Walker, M., Iida, M., Cote, S.: Japanese discourse and the process of centering. Computational Linguistics **20**(2) (1994) 193–232
2. Voorhees, E.M.: Overview of the TREC 2001 question answering track. In: Proceedings of the tenth Text Retrieval Conferene (TREC 2001). (2001)
3. Kato, T., Fukumoto, J., Masui, F., Kando, N.: Are open-domain question answering technologies useful for information access dialogues? ACM Transactions on Asian Language Information Processing (TALIP) **4**(3) (2005) 243–262
4. Murata, Y., Akiba, T., Fujii, A., Itou, K.: Question answering experiments at NTCIR-5: Acquisition of answer evaluation patterns and context processing using passage retrieval. In: Proceedings of the Fifth NTCIR Workshop Meeting. (2005)
5. Matsuda, M., Fukumoto, J.: Answering questions of IAD task using reference resolution of follow-up questions. In: Proceedings of the Fifth NTCIR Workshop Meeting. (2005)
6. Nariyama, S.: Grammar for ellipsis resolution in Japanese. In: Proceedings of the 9th International Conference on Theoretical and Methodological Issues in Machine Translation. (2002) 135–145
7. Mori, T.: Japanese question-answering system using A^* search and its improvement. ACM Transactions on Asian Language Information Processing (TALIP) **4**(3) (2005) 280–304
8. Ishioroshi, M., Mori, T.: A method of list-type question-answering based on the distribution of answer score generated by a ranking-type Q/A system. SIG Technical Reports 2005-NL-169, Information Processing Society of Japan (2005) (in Japanese).
9. Yamura-Takei, M.: Approaches to zero adnominal recognition. In: Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL2003) Student Research Workshop, ACL (2003) 87–94
10. Kawahara, D., Kurohashi, S.: Zero pronoun resolution based on automatically constructed case frames and structural preference of antecedents. In: Proceedings of the 1st International Joint Conference on Natural Language Processing. (2004) 334–341
11. Kato, T., Fukumoto, J., Masui, F.: An Overview of NTCIR-5 QAC3. In: Proceedings of the Fifth NTCIR Workshop Meeting. (2005)