# Bi-directional Cross Language Question Answering using a Single Monolingual QA System

*Kei Shimizu* *Tomoyosi Akiba*
Department of Information and Computer Sciences, Toyohashi University of Technology
1-1-1 Hibarigaoka, Tenpaku-cho, Toyohashi-shi, 441-8580, JAPAN
akiba@cl.ics.tut.ac.jp

*Atsushi Fujii*
Graduate School of Library, Information and Media Studies, University of Tsukuba
1-2 Kasuga, Tsukuba, 305-8550, JAPAN

*Katunobu Itou*
Graduate School of Information Science, Nagoya University
1 Furo-cho, Nagoya, 464-8603, JAPAN

## Abstract

*Toward the first attempt of evaluating Cross Lingual Question Answering in NTCIR workshop 5, our team participated in the English-Japanese subtasks of both directions (E-J and J-E tasks). Our approach to construct the CLQA systems of both directions is to use only a single (Japanese) QA system. Instead of CLIR systems, we also used statistical machine translation systems of both directions. For the English-to-Japanese CLQA system, the straightforward cascade architecture, in which an E-J machine translation component is followed by a Japanese QA system, is augmented by the several supporting components that take advantage of original, therefore not degraded, information extracted from the question in English. For the Japanese-to-English CLQA system, the retrieved passages in English, which are relevant to the original Japanese question, are translated back to the passage in Japanese. Then the Japanese QA system is invoked by submitting the original Japanese question to extract the answer from the translated passages. The proposed method is effective for building a bi-directional CLQA system between the specific language pairs, in which only one language side has its monolingual QA system: this is often the case between a major language and a minor language.*

## 1 Introduction

After the past evaluations performed in NTCIR Question Answering Challenge (QAC) and Cross Language Information Retrieval (CLIR) series, Cross Lingual Question Answering (CLQA) task is started in NTCIR workshop 5, in order to evaluate the technologies for the question answering across different languages, which accept the question in some language and extract the answer from the target documents in another language.

Our approach for participating the NTCIR-5 CLQA task is to exploit the existing tools and systems. In the past NTCIR workshops, we have participated in QAC, but not in CLIR. Therefore, we have a Japanese QA system, but do not have a CLIR system. Instead of the CLIR system, we decided to use an existing (monolingual) document retrieval engine and machine translation systems, which is intended to be used for translating the question in the source language to that in the target language. We adopted a statistical approach for machine translation, because it is easy to construct the MT systems of the both directions by adopting the SMT framework if sufficient amount of parallel text corpus is available.

By using the existing tools and systems described above, we participated in the English-Japanese subtasks of both directions (E-J and J-E tasks). Since we do not have an English QA system, our Japanese QA system must be used for extracting answers in both E-J and J-E tasks.

It is straightforward to construct the English-to-

**Table 1. bilingual corpus for translation model**

| bilingual corpus | number of pairs (K) |
|---|---|
| EIJIRO | 1400 |
| Dictionary of English Collocations | 350 |
| Random House Dictionary | 180 |
| Project Sugita Genpaku | 20 |
| **total** | 1950 |

Japanese CLQA system by using a Japanese QA system. The cascade architecture is adopted, in which an E-J machine translation component is followed by a Japanese QA system. This simple architecture is augmented by the several supporting components that take advantage of extracting the fresh information extracted from the original question in English.

On the other hand, it is not straightforward to construct the Japanese-to-English by using a Japanese QA system instead of an English QA system. For our J-E CLQA system, the relevant target documents in English are translated back to the documents in Japanese, and then the Japanese QA system is invoked by submitting the original Japanese question to extract the answer from the translated documents.

## 2 Components of the CLQA system

### 2.1 Statistical Machine Translation

In our E-J and J-E CLQA systems, machine translation is used in several components. Among them, the main components, in which the MT is indispensable, are the question translation components, where a submitted question to the system is translated to the question in the target language. In the E-J system, the question in English is translated to the question in Japanese, while the submitted question in Japanese is translated to that in English.

The statistical machine translation consists of two models: the translation model and the language model. In our system, the IBM Model 4 [2] is used as our translation model. We used GIZA++ [8], the existing tools for training the translation model from a parallel corpus. Our parallel corpus is obtained from several resources listed in Table 1.

In addition to the corpora from the existing resources, about 700 pairs of parallel question fragments collected manually were also used as the training corpus for the translation model, in order to adapt the model to interrogations. For example, the pair "What is the name" and "nan to iu namae desu ka" was added to the training corpus.

The tri-gram model is used as our language model. CMU Cambridge Language Model Toolkit ver.2 [3] is used for training the model. The six years amount of Japanese newspaper articles were used for training the Japanese language model, while the two years amount of English newspaper articles were used for training the English language model.

The ISI Rewrite Decoder [5] is used for decoding. The weight for the translation model and the language model was set to be equal (0.5 for both) for decoding.

### 2.2 Japanese Question Answering

The question answering system participated in NTCIR QAC2 was used for our CLQA system as it was. The detailed description about the system is found in [1].

## 3 English-to-Japanese CLQA system

### 3.1 Overview

The process of our English-to-Japanese (E-J) CLQA system consists of the following steps.

1. Translate an English question sentence submitted to the system into the Japanese question sentence by using the statistical machine translation.
2. Estimate the question type from the original English question.
3. Extract the question focus from the original English question and translate the extracted English focus into Japanese.
4. Invoke the Japanese question answering system by submitting the translated Japanese question sentence, the estimated question type, and the translated Japanese question focus, then obtain the Japanese answer strings.

Figure 1 illustrates the configuration of our E-J CLQA system.

### 3.2 Question Type Estimation

Our Japanese QA system utilizes 6 question types to restrict the answer type: *PERSON*, *LOCATION*, *ORGANIZATION*, *OTHER NAME EXPRESSION*, *DATE*, and *OTHER NUMERICAL EXPRESSION*. The English question is analyzed to estimate its question type. Basically, the type can be estimated from the English question words: what, which, who, whose, when, where and how. The process of the estimation is as follows.

Firstly, the English question word is extracted if it exists. Since the question word is also used as a relative pronoun, we introduce the heuristics that select a

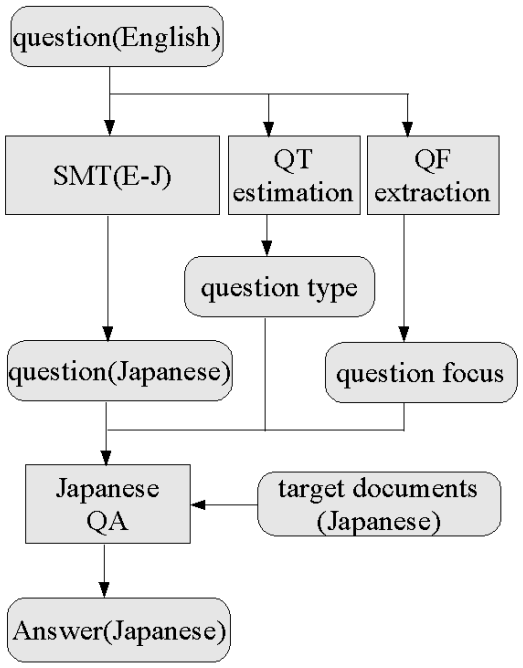


**Figure 1. E-J CLQA system**

question word according to the following order as the preference.

1. The first word in the question.
2. The second word in the question.
3. The word immediately after a comma.
4. The other words.

If the second and third heuristics selects more than one word, we prefer the word appeared first in the question. For example, from the question *"In this case, where is the criminal person who steals the jewel?"*, the word "where" is extracted as the question words.

Secondly, the question type is basically decided according to the question word extracted. The correspondence between the question word and the question type is listed as follows.

- what: ALL
- which: ALL
- who: PERSON
- whose: PERSON
- when: DATE
- where: LOCATION
- how: QUANTITY

We also utilize the question focus, explained in Section 3.3, to decide the question type. For example, if the question word is *"what"* and the question focus is *"country"*, the question type *LOCATION* is estimated. We constructed such rules manually. The example of the rules is listed in Table 2.

**Table 2. The head word of QF and the corresponding QT.**

| QF | changed QT |
|---|---|
| persent | QUANTITY |
| number | |
| rate | |
| amount | |
| population | |
| country | LOCATION |
| prefecture | |
| location | |
| city | |
| organization | ORGANIZATION |

### 3.3 Extracting Question Focus

In TREC and NTCIR, a question often contains the word or phrase that directly express the semantic category for the correct answer. For example, the question *"kokumin eiyo shou wo jushou shita eiga kantoku wa dare desu ka?"* (Who is the film director received the national honorary prize?) implies that the answer should be an instance of *"eiga kantoku"* (film director). In another example, the question *"ZIP no kioku youryou wa ikutsu desu ka?"* (What is the capacity of ZIP?) implies that the answer should be a numerical expression followed by a unit expression, such as "mega byte" and "giga byte" implicated by the word *"kioku youryou"*(memory capacity). We shall call these words (or phrases) representing the semantic categories for the correct answers "Question Focus (QF)".

A QF provides useful information to find correct answers. Moldovan et al. [7] utilized the QFs for answering "what" type questions. Ittycheriah et al. [6] preferred the answers that had hypernym or hyponym relationship in WordNet with the QF. Prager and Chu-Caroll [9] focused on answering "What is X?" questions. The WordNet was consulted for the extracted QF and the hypernyms were considered as the answer candidates of the definitional what-is question.

Our Japanese QA system also utilizes the QF for answer evaluation [1]. Therefore, extracting the QF from the original English question seems to promise for improving the CLQA performance.

Our current implementation of QF extraction is simple. The last word of the word sequence immediately after the question word is extracted as a QF.

Then, the extracted QF in English is translated into Japanese by using our machine translation system. The translated Japanese string is used as the Japanese QF and submitted to the Japanese QA system along with the translated Japanese question.

## 4 Japanese-to-English CLQA system

The process of our Japanese-to-English (J-E) CLQA system consists of the following steps.

1. Translate a Japanese question sentence submitted to the system into the English question sentence by using the statistical machine translation.
2. Using the English question as the query, English document retrieval is performed to extract the passages relevant to the query from the English target documents.
3. The top ranked $N$ passages are translated back into Japanese by using the statistical machine translation. The word alignment between the English and Japanese passage is also obtained.
4. Invoke the Japanese QA system by submitting the original Japanese question to extract the answer from the translated passages. The QA system returns the answer strings and their position in the passages.
5. Using the word alignment between the English and Japanese passage, the English answers are obtained from the Japanese answers.

Figure 2 illustrates the configuration of J-E CLQA system.

Note that this step 5 is not necessary for the practical use of the CLQA system. According to the task definition of NTCIR-5 CLQA evaluation, the answer must be answered in the target language. This definition does not match to the practical use of CLQA, in which the answer is better to be returned in the source language, but is adopted in order to simplify the task as the first attempt of the evaluation. However for our strategy, our system is forced to execute the unnecessary additional process.

## 5 Experiments

### 5.1 E-J CLQA system

Experimental evaluation was conducted for our E-J CLQA system. The QAC2 subtask 1 test collection was used for the evaluation, because the corresponding English questions were distributed from the organizers. We submitted the English questions to the E-J CLQA system and evaluated the returned answers with

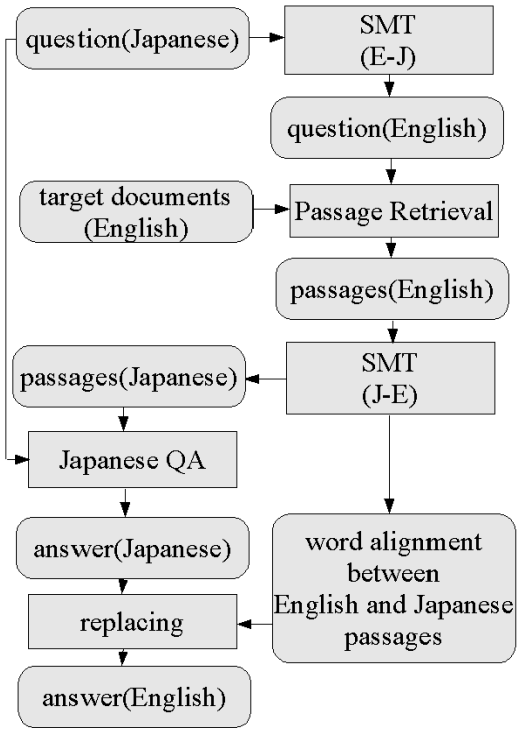


**Figure 2. J-E CLQA system**

respect to the answer set of QAC2 subtask 1. Mean Reciprocal Rank (MRR) was used for the evaluation measure, which is used for QAC2 subtask1 evaluation [4]. The evaluation was performed from the two points of view: we investigated the effects of the machine translation performance and the effects of the two optional processes, i.e. the question type estimation and the question focus extraction.

#### 5.1.1 Effects of Machine Translation Performance

We compared three systems that differ only in the machine translation method used for translating an English question to Japanese. For all three systems, we used neither the question type estimation nor the question focus extraction.

The first system used our statistical machine translation (referred as *Statistical MT*), which is just the system used for our participation in CLQA formalrun. The second system used the online machine translation system available through the Web (referred as *Internet MT*). The third system was just our Japanese QA system and the corresponding Japanese questions were submitted instead of English questions. This corresponds to the system with perfect machine translation (referred as *Ideal MT*).

The results are shown in Table 3. This indicates that the statistical MT used for our system has very

**Table 3. Effects of machine translation methods (E-J task).**

| MT method | MRR |
|---|---|
| Statistical MT | 0.020 |
| Internet MT | 0.163 |
| Ideal MT | 0.486 |

**Table 4. Effects of QT estimation and QF extraction (E-J task).**

| QTQF | MRR |
|---|---|
| *baseline* | 0.163 |
| *+QT* | 0.127 |
| *+QF* | 0.164 |
| *+QT, +QF* | 0.080 |

low performance and is quite unreliable. The considerable improvement will be indispensable. Therefore the following experiment was conducted by using Internet MT.

#### 5.1.2 Effects of Question Type Estimation and the Question Focus Extraction

For the second experiment, we investigated the effects of the two optional processes to support the CLQA system, that are the question type (QT) estimation and the question focus (QF) extraction. We compared four systems: the baseline system that does not use both QT estimation and QF estimation (referred as *baseline*), the system that uses only QT estimation (referred as *+QT*), the system that uses only QF extraction (referred as *+QF*), and the system that uses both QT estimation and QF extraction (referred as *+QT +QF*).

The results are shown in Table 4. The use of QT estimation degraded the performance. Since our Japanese QA system has its own QT estimation mechanism, we think that the current implementation of the QT estimation from an English question is somehow problematic. On the other hand, the use of the QF extraction slightly improved the performance. The combination of QT estimation and QF extraction degraded the performance largely. We think that this is also the problem of the current implementation.

### 5.2 Results with respect to CLQA

Experimental results with respect to CLQA are shown in table 5. We investigated two grades of judgment. In the judgment referred as **R**, we considered the answer correct if both its string and its supported document ID agree with the correct answer set provided by the CLQA organizers. In the judgment referred as **R+U**, we considered the answer correct if its string agrees with the correct answer set regardless of its supported document.

Seeing **R**, the results shows that using Internet MT instead of Statistical MT decreases the performance, which is reverse of the results in QAC2 (Table 3). However, seeing **R+U**, Internet MT outperforms Statistical MT, which is the same with the results in QAC2. We manually investigated several unsupported answers obtained by using the Japanese QA system, and found that 18 out of 20 answers selected at random were correctly supported. It seems to indicate that the current correct answer set is incomplete.

## 6 Conclusion

We participated in the English-Japanese subtasks of both directions (E-J and J-E tasks). Our approach to construct the CLQA systems of both directions was to use only a single (Japanese) QA system. For the E-J CLQA system, the straightforward cascade architecture, in which an E-J machine translation component was followed by a Japanese QA system, was augmented by the two supporting processes, i.e. question type estimation and question focus estimation. For the J-E CLQA system, the retrieved passages in English were translated back to the passage in Japanese. Then the Japanese QA system was invoked by submitting the original Japanese question to extract the answer from the translated passages. The experimental results showed that the current system should be improved in several points. However, we think the method proposed in this paper is effective for building a bi-directional CLQA system between language pairs in which only one language side has its monolingual QA system: this is often the case between a major language and a minor language.

## References

[1] T. Akiba, A. Fujii, and K. Itou. Question answering using "common sense" and utility maximization principle. In *Proceedings of The Fourth NTCIR Workshop*, 2004. `http://research.nii.go.jp/ntcir/workshop/OnlineProceedings4/QAC/NTCIR4-QAC-AkibaT.pdf`.

[2] P. F. Brown, S. A. D. Pietra, V. J. D. Pietra, and R. L. Mercer. The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics*, 18(4):263–311, 1993.

[3] P. R. Clarkson and R. Rosenfeld. Statistical language modeling using the CMU-cambrdge toolkit. In *Proceedings of European Conference on Speech Communication and Technology*, pages 2707–2710, 1997.

[4] J. Fukumoto, T. Kato, and F. Masui. Question answering challenge for five ranked answers and list answers – overview of NTCIR 4 QAC2 subtask 1 and 2 –. In *Proceedings of 4th NTCIR workshop*, 2004.

[5] U. Germann, M. Jahr, K. Knight, D. Marcu, and K. Yamada. *ISI Rewrite Decoder User's Man-*

**Table 5. Experimental results with respect to CLQA.**

| | | **R** | | | **R+U** | | |
|---|---|---|---|---|---|---|---|
| direction | MT method | Acc | Top 5 Acc | MRR | Acc | Top 5 Acc | MRR |
| E-J | Statistical MT | 0.0550 | 0.0650 | 0.0588 | 0.0650 | 0.115 | 0.0838 |
| | Statistical MT +*QT* +*QF* | 0.0550 | 0.0650 | 0.0588 | 0.0650 | 0.115 | 0.0846 |
| | Internet MT | 0.0300 | 0.0550 | 0.0381 | 0.0750 | 0.170 | 0.108 |
| | Ideal MT | 0.110 | 0.250 | 0.162 | 0.175 | 0.425 | 0.265 |
| J-E | Statistical MT | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | Internet MT | 0.000 | 0.000 | 0.000 | 0.000 | 0.0025 | 0.005 |

*ual*. http://www.isi.edu/natural-language/software/decoder/manual.html.

[6] A. Ittycheriah and S. Roukos. IBM's statistical question answering system – TREC-10. In *Proceedings of TREC-10*, 2001.

[7] D. Moldovan, S. Harabagiu, M. Pasca, R. Mihalcea, R. Goodrum, R. Girju, and V. Rus. LASSO: A tool for surfing the answer net. In *Proceedings of TREC-8*, pages 65–73, 1999.

[8] F. J. Och. *GIZA++: Training of statistical translation model*. http://www-i6.informatik.rwth-aachen.de/Colleagues/och/software/GIZA++.html.

[9] J. Prager and J. Chu-Carroll. Answering what-is question by virtual annotation. In *Proceedings of Human Language Technology Conference*, pages 26–30, 2001.