Applying Linked Open Data to Machine Translation for Cross-lingual Question Answering

Takaaki Matsumoto
Carnegie Mellon University
5404 GHC 5000 Forbes Ave.
Pittsburgh, Pennsylvania, USA 15213
SOC Corporation
3-16-17, Takanawa, Minato-ku
Tokyo, Japan 108-0074
takaaki.m@gmail.com

Teruko Mitamura
Carnegie Mellon University
6711 GHC 5000 Forbes Ave.
Pittsburgh, Pennsylvania, USA 15213
teruko@cs.cmu.edu

ABSTRACT

This paper addresses the methodology and its evaluation for answering cross-lingual essay questions by utilizing linked open data which assists machine translation. The question answering (QA) system studied in this paper generates English essays for the world history subject of the entrance examination of University of Tokyo. Most answers can be found in the Japanese world history textbooks. However, equivalent content of high quality English translation of the Japanese world history textbooks are not available. Therefore, we try to translate those textbooks utilizing linked open data, and to use source language knowledge resource of which content is not equivalent with the target knowledge resource.

The evaluation result indicates that the proposed method shows better performance compared with the baseline method [10] and the previous research [4]. The result of the proposed system is almost equivalent to the well designed Wikipedia based system [8].

The result of this paper concludes that 1) simple neural translation of knowledge resource does not work for domain specific cross-lingual question answering, 2) linked open data is effective to find correct translation for difficult terms in machine translation process, and 3) adding source language open knowledge resource would help even if its content is not equivalent with the target knowledge resources.

KEYWORDS

Question answering, linked open data, NTCIR-13, Wikidata

1 INTRODUCTION

Question Answering (QA) research has been done for a long time, and their successes are widely found in factoid and multiple-choice questions. However, essay question answering, which is often found in a real-world situation, is considered to be one of the most difficult QA tasks, because it is often related to a multi-document summarization task.

It is essential to have knowledge resources to solve essay QA tasks. Some domains, for example law, patent, business, and so on are highly dependent on a language or a culture, and effective knowledge resources disproportionately exist from language to language. For example, answering English essay question about Japanese business custom is not an easy task. There are three ways to solve this kind of cross-lingual QA; 1) applying machine translation to question and answer, and solving the QA task in the target language, 2) translating the target knowledge resources into the source language by machine, and solving the QA task in the source language, and 3) solving the QA in the source language using a large scale open-domain knowledge resource of the source language, hence it is a mono-lingual QA. The first option is the simplest way. However, two times machine translations, source question to target question and source answer to target answer, may reduce the translation accuracy. The second option can be a useful approach if the knowledge resources are not very large. The third option does not contain machine translation. However, since a large scale open-domain knowledge resource like Wikipedia is high signal to noise ratio, retrieving correct answer is difficult. This paper employs the second option, because the knowledge resource size of the target task is small enough.

The NTCIR-13 QA Lab is a challenge to solve the Japanese university entrance examinations (on world history) in English [3][14][13] . In the QA Lab, there are three types of questions; multiple-choice, term (factoid), and essay question. The essay questions of QA Lab are selected from the past world history examinations of University of Tokyo, Japan. University of Tokyo entrance examination is considered to be one of the most difficult examinations in Japan, and generally questions are based on the Japanese high school textbooks.

In the task, there are two types of essays; 1) short/simple essay and 2) complex/long essay. A short/simple essay question expects a short answer, which is usually a single sentence (15-60 words). Many of these questions may contain a factoid question as part of the answer. A complex/long essay question requires a longer answer, which consists of multiple sentences (225-270 words). It usually contains a longer introductory paragraph and it also contains a list of 4-9 keywords that are required to be used in the essay.

In this paper, we focus on the essay question answering for world history subject in the NTCIR-13 QA Lab-3 in English. We describe the previous challenges and performance difference between closed and open knowledge bases (Section 2), the methodology to utilize linked open data for the task in English (Section 3), results and discussions of the proposed method (Section 4), and conclude the paper (Section 5). In the Section 4, the evaluation result of the proposed system is compared with not only the baseline but also an another QA system that uses a large scale open domain knowledge base, which is mentioned as the third option in above.

2 PREVIOUS RESEARCH AND BASELINE

In the NTCIR-12 QA Lab-2 (2016) [1], Phase-1, both English and Japanese essay tasks were evaluated. The best ROUGE-1 [7] scores were quite different; the best Japanese system had approx. 0.3 [13][11], while the English system had 0.0326 [4]. This was only 1/10 of that of Japanese. One of the reasons for low scores in English can be a language barrier, because the entrance examination is based on the Japanese world history high school textbooks and no English version of them were available.

For the baseline system of this study, we use a multilingual essay question answering system developed by Sakamoto et al. [10][12]. In the baseline system, the knowledge resources they used are machine translated texts of five Japanese world history textbooks and one Japanese world history glossary published from Tokyo Shoseki and Yamakawa. The translation was attempted in 2015 with Google translate, in which the statistical translation technique was used.

3 PROPOSED METHOD

As described above, one of the most different things between Japanese and English tasks in NTCIR QA Lab was the availability of the knowledge resources. Japanese teams could use five Japanese high school textbooks, while English teams mainly used Wikipedia. In this section, we propose an essay generating system for cross-lingual question answering task that utilizes linked open data for machine translation of the knowledge resource.

3.1 Improving of Machine Translation of Native Textbooks using Linked Open Data

The proposed method attempts to improve machine translation quality of Japanese textbooks. We use a linked open data to find correct translation.

A preliminary study of Japanese exams indicated that the Japanese textbooks cover more than 80% of the questions of University of Tokyo entrance examinations. However, machine translated textbooks by Google Translate in 2015 lack many important terms and produce errors. For example, ササン朝 (Sasanian Empire) was translated as "sasan morning," because the Japanese character 朝 means both "dynasty" and "morning," and generally uses as "morning." The latest neural translation technology might be able to improve translation quality. However, we found that some nouns are mistranslated in the neural translation as follows (Table 1). Table 1 clearly shows some nouns (especially, compound noun) were mistranslated by the latest neural transition, and Wikidata.org translated them perfectly. Therefore, in order to translate difficult but important terms, we created a bilingual world history term corpus by utilizing linked open data (LOD).

3.1.1 Bilingual World History Term Corpus. In order to find the correct English translation in the Wikidata.org and build a bilingual world history term corpus, two strategies were adopted; 1) exact match or only one, 2) longer match.

The objective of the first strategy is to generate the bilingual corpus with very high precision and adequate recall. A candidate Japanese term found in the Japanese world history glossary is firstly tried exact match in Wikidata.org. If it matches, the translation word is retrieved. If it does not match exactly, then the word is searched, and if the number of search results is only one, the translation word is retrieved. If the number of the search result are greater than two, the translated results are ambiguous and they are not utilized.

The second strategy is to avoid mistranslation. This strategy would help to retrieve compound nouns correctly. Assume that the following Japanese passage in the glossary:

またキリスト教綱要によれば

(Also according to the Institutes of the Christian Religion).

Firstly, morphological analysis (MeCab [6]) is applied and tokenized text is obtained.

また|キリスト|教|綱要|に|よれ|ば

(CONJ | NP | suffix | N | case marker | V | CONJ particle).

Then, the linked open data assists translation. Translation starts with a noun or proper noun, and ends if the next word is neither a noun nor some exceptions (suffix or some symbols). At first, また, which means "also," is neither a proper noun or a noun, and therefore また is ignored. キリスト is a proper noun and the translation starts. The Wikidata.org has an exact match result of "Christ." The next word 教 is a suffix and the translation continues. キリスト教 is also found in Wikidata.org and the translation of "Christianity" is retrieved. 綱要 is also a noun and キリスト教綱要 is found in Wikidata.org and its translation of "Institutes of the Christian Religion" is saved. The next word に is a case marker, so the translation process stops. Finally, the longest translation "Institutes of the Christian Religion" word is retrieved correctly as the translation of キリスト教綱要.

By using this technique, the bilingual world history translation corpus was generated. Since the results were large, we could not examine all the results. However, we sampled the results and found that most long terms are correct and some short terms were wrong. We checked all terms of which length is less than 4 characters, and found only approx. 100 mistranslations in the results. Finally, 6,962 Japanese terms and their English translations were retrieved. In addition, approx. 2,000 English words were added from the world history ontology [5].

3.1.2 Translating Japanese World History Textbooks. The Japanese textbooks are translated in two steps; firstly by the bilingual world history term corpus described 3.1.1 and secondly by commercial translation API (Microsoft Bing Translator). At first, all terms that match with the bilingual corpus in the whole Japanese text are replaced into English terms and then a Japanese-English mixed text is generated. After that, it is translated by commercial neural translation API. In this paper, we used Microsoft Bing Translator since it translated some world history related nouns better than Google Translate as shown in Table 1. For example, a Japanese passage:

" またキリスト教綱要によれば "

is firstly translated into Japanese-English mixed text:

" また Institute of the Christian Religion によれば. "

Then, the text is translated by Microsoft Bing Translator into:

" Also according to the Institute of the Christian region. "

This is a better translation than Google Translate, " According to Christianity requirements. "An example of this process is shown in the appendix.

Table 1: Translation Examples

Japanese Term	Google Translate (2017)	Bing Translator	Wikidata	Correct Translation
林則徐	Hayashi Noriro	the zexu	Lin Zexu	Lin Zexu
欽差大臣	Minister of Ginza	Minister of the Qin	Imperial Commissioner	Imperial Commissioner
キリスト教綱要	Christianity requirements	Christian elements	Institute of the Christian Religion	Institute of the Christian Religion

3.1.3 Discussion. The proposed method has two strategies, 1) exact match or only one, and 2) longer match, to build the bilingual world history term corpus. They might be seem not to be effective to solve critical issues that may arise in the translation process, because the "exact match or only one" strategy can be regarded as avoiding of the ambiguity problem. However, based on our observations and assumptions of the translation problems of the world history textbooks, we think that the proposed strategies are effective even.

Firstly, we found that most of the mistranslating terms in the Japanese world history textbooks are very difficult and rare nouns. They are the names of a person, country, dynasty, war, treaty, and so on. Those terms are often found unambiguous ways. Some wars or treaties have alias names. However, since we can write down only one name in the answer in general and alias name is not often asked, translation to the alias name is not necessary.

Secondly, the combination of the "exact match or only one" and the second strategy of the "longer match" often helps to solve ambiguity problems. Let's look at the example of オスマン帝国は (in English, Ottoman empire is). By the morphological analysis of the MeCab, we obtain a chain of morphemes of オスマン/帝国/は (NP/N/Particle). The system tries exact match of the first word オスマン in Wikidata.org. However, it is ambiguous and has no exact match. Then, because of no exact match, searching in Wikidata.org is attempted. We have many search results, Ottoman Empire, Osman I, Ottoman Dynasty, Ottoman Turkish, and so on. These translations can be correct if only the word of オスマン is given. This kind of ambiguity can be solved by contexts. However, we have the another noun of 帝国, which succeeds to the オスマン. The compound noun of オスマン帝国 gets the exact match of the Ottoman Empire. We still have many search results for オスマン帝国, if searching in Wikidata.org is attempted. However, exact match has precedence over searching in our algorithm, and the ambiguity problem does not happen if the exact match is succeeded.

Searching in Wikidata.org makes sense when the term has alias names, including orthographic variants. As we pointed before, we have some aliases for word history terms. Especially, Japanese has Romanization and it often generates many similar aliases. For example, "Sasanian Empire" is represented as ササン朝 in the textbooks we used, but, the de-facto translation is considered to be サーサーン朝, which uses to macrons (there are many orthographic variants for foreign originated terms in Japanese Katakana). Hence, ササン朝 fails exact match in Wikidata.org because it only checks the title of the article. However, the articles in Wikidata.org contains alias field and we can find "Sasanian Empire" when we use the search of ササン朝. Another example for this problem is じゃがいも飢饉 (Great Irish Famine). Since じゃがいも飢饉 is a common noun compound, Google translate mistranslates "Potato famine," which is translations of じゃがいも and 飢饉. However, Wikidate.org can find correct translation for not only the de-facto term of じゃがいも飢饉 but also its alias name of アイルランド大飢饉 (Great Irish Famine). We can say the proposed strategies can handle the translation problem of the orthographic variants or alias names of the source language (Japanese) correctly.

Table 2: Comparison between LOD assisted Machine Translation and Simple Machine Translation

	Number of words translated by LOD	LOD Failure and Bing Success	Bing Failure and LOD Success
Sample 1	33	1	2
Sample 2	40	3	10
Sample 3	22	1	3
Sample 4	21	1	9
Sample 5	42	3	7

Another discussion for the proposed method can be words that are not in the Wikidata.org are not usable (as mentioned in 3.1.1). We used the language link data of the Wikidata.org which is equivalent with the inter-language link of the Wikipedia articles to find correct translation. Some articles of the Wikipedia are deep-rooted in the culture and tradition and few language links can be found, and some words are clearly not in Wikipedia. However, since the question answering task in this paper deals with the world history subject of a university entrance examination, we think that the coverage of the Wikidata.org is considered to be enough.

We analyzed 5 sample articles of a textbook, which becomes approx. 250 words in English after translation (the original articles have about 500 characters in Japanese). We counted the number of words translated by the bilingual world history term corpus (LOD assisted machine translation), and checked their translation quality. Table 2 shows the result. In all five sampled articles, approximately from 20 to 40 words of each article were translated from Japanese to English using the bilingual world history term corpus. A few (from 1 to 3) words of each article were found to be mistranslated. About the half of them could be translated correctly if the Bing Translator is used directly, but the another words cannot be translated by both of the corpus (Wikidata) and Bing Translator. When we directly applied Bing Translator to the sample articles, we had many mistranslations for the words that were translated by the bilingual corpus correctly. This result indicates that the pre-translation by the proposed bilingual world history term corpus is

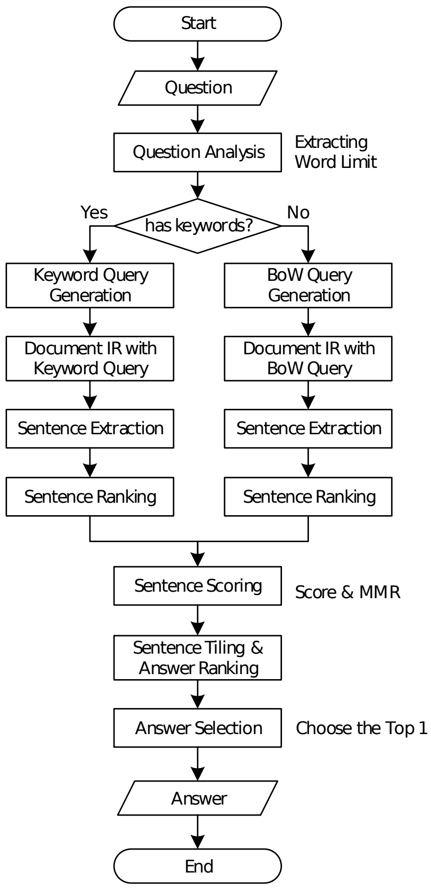

Figure 1: System Flowchart.

very effective for the machine translation of the textbooks to translate rare nouns correctly. On the other hand, we found some effects of the pre-translation process. Some sentences can lose coherency, and the translation quality of some words improves or worsens. These analyses are future research.

3.2 Additional Domain Specific Open Knowledge

Since the translation of Japanese textbooks is done by machine translation, mistranslations are inevitable. Therefore, we add one public English world history textbook from Boundless.com [2]. While some public English world history textbooks are available in PDF format in on-line, the textbook of Boundless.com is a HTML based and easy to use for natural language processing task.

3.3 System Description

Fig. 1 shows the system flowchart of this method. The system flow is following.

(1) At first, the question data is given in XML format.
(2) The question data is analyzed by the question analysis module, and the maximum answer length is obtained.
(3) The system has a different IR strategies for question type. If the question has keywords that are required to be used in the essay, the question is a complex/long essay. Otherwise, the question is regarded as a short/simple essay.
(4) Query data for IR is generated. For long essays, the keywords in the question are used. For short essays, the bag of word (BoW) of the question sentences are adopted.
(5) Using the query, documents (set of passages) are retrieved from the knowledge resources.
(6) Sentences are ranked by the IR scores.
(7) Sentences scoring module gives a score which indicates the relevance or entailment for the question to the extracted sentences.
(8) Scored tiling module generates essays by changing order of the extracted sentences. The score of an essay candidate is summation of the sentence scores in the essay.
(9) The top 1 score essay is chosen as the answer.
(10) The answer XML data is generated.

The baseline system uses following scoring method by default:

$$\text{Score} = \frac{k_m}{m} \tag{1}$$

where k_m is the number of keywords in the sentence, and m is the number of words of the sentence. All keywords and words of the sentence are stemmed. Stop words and punctuations are removed before calculation.

Eq.1 measures the density of the keywords in a sentence. However, not always the given keywords and words in the sentence match exactly. Some words of the answer sentence could be similar to the given keywords. Hence, word level similarity between retrieved or given keywords and an extracted sentence is calculated as follows:

$$\text{Score} = \sum_{i=1}^{m} \frac{\max(w_i \cdot k_1, w_i \cdot k_2, ...w_i \cdot k_n)}{\log m} \tag{2}$$

where, m is the number of words in the sentence except stop words and punctuations, n is the number of keywords, w_i is the $i-$th word vector of the sentence, and k_j is the $j-$th keyword vector. Word embedding is given by GloVe [9]. Using the score, answer candidates are generated and their scores are also given by just summation of the sentence score. Finally, the top 1 essay is selected as an answer and answer XML file is outputted.

4 RESULT AND DISCUSSION

The proposed methods are evaluated using the NTCIR-13 QA Lab-3 official phase-1 dataset, which contains 5 long/complex and 22 short/simple essay questions and ground truths [3]. In the QA Lab, evaluation is done by human experts, ROUGE method and Pyramid method [3][13]. In this paper, ROUGE-1 and 2, unigrams and bigrams to compare the essay to a set of gold-standard essays, are used for evaluation. Sample questions, gold standards and system answers are shown in the appendix.

Table 3 shows the ROUGE-1 and ROUGE-2 evaluation results of the baseline and prosed method for the dataset.

The system A shows the combination of the baseline system and the baseline knowledge resources (machine translated Japanese textbooks using Google Translate in 2015). The system B shows the combination of the baseline system and the neural machine

Table 3: End-to-end Evaluation Result of Each System

System	Evaluation Method	Number of Questions	Mean	Max	Median	Min	Variance	Standard Deviation
(A) Baseline	ROUGE-1	27	0.063	0.244	0	0	0.007	0.081
	ROUGE-2	27	0.009	0.067	0	0	0.000	0.018
(B) Baseline	ROUGE-1	27	0.056	0.260	0	0	0.010	0.077
+NMT	ROUGE-2	27	0.004	0.054	0	0	0.000	0.011
(C) Baseline	ROUGE-1	27	0.081	0.375	0.054	0	0.010	0.100
+LNMT	ROUGE-2	27	0.011	0.064	0	0	0.000	0.021
(D) Baseline	ROUGE-1	27	0.076	0.225	0.063	0	0.010	0.075
+LNMT + WS	ROUGE-2	27	0.012	0.118	0	0	0.001	0.027
(E) Baseline	ROUGE-1	27	0.128	0.485	0.105	0	0.015	0.122
+LNMT +WS +ET	ROUGE-2	27	0.028	0.176	0	0	0.003	0.050
(F) Wikipedia-based	ROUGE-1	27	0.123	0.320	0.1	0	0.008	0.088
	ROUGE-2	27	0.025	0.167	0	0	0.002	0.042

translated (NMT) textbooks, and it is worse than that of the baseline system and baseline knowledge resource. One of the reasons of the difference is the mistranslation of some rare terms, as pointed out in the section 3. The system C shows the combination of the baseline system and linked open data (Wikidata) which assisted neural machine translated textbook (LNMT). When LOD assisted neural machine translated textbooks are used, the score was improved. Since the ROUGE-1 is based on unigrams to compare to the gold-standard, correct words in an answer existed is very important. In addition, the LOD assisted translation can give correct English entity names. Therefore, system C improved the ROUGE score effectively.

The system D and E adopt word similarity based sentence scoring (WS). The system D gets almost the same ROUGE-1 and ROUGE-2 means compared with those of the system C. However, when the English textbook (ET) is added to the knowledge resource (system E), it has the best ROUGE-1 and ROUGE-2 means. It should be noted that the number of the question is only 27 (5 long essays and 22 short essays). Since the NTCIR QA Lab uses the real past entrance examination of University of Tokyo, the provided data was very small. The performance differences are not statistically significant when the standard deviations are considered.

System F is the reference system developed for the same task (NTCIR-13 QA Lab-3) [8] which uses whole English Wikipedia as the knowledge resource. It employs carefully designed keyword weighting for document retrieval and sentence extraction to overcome the high signal to noise ratio of the whole Wikipedia. The proposed system in this paper has almost equal ROUGE-1 and 2 means to the system F. In addition, even though it should be noted that the results of the proposed system and the previous research cannot be simply compared because of the different questions, the best ROUGE-1 mean of the proposed system is about four times larger than that of the previous study that also uses Wikipedia (0.0326 in ROUGE-1 mean) [4].

In summary, the reasons for the better ROUGE-1 and ROUGE-2 means of the proposed method compared with that of the baseline are attributed to the accurate named entities of the knowledge resources and the similarity measurement in the sentence scoring process.

Table 4: Short and Long Essays

System	All Essay ROUGE-1 Mean	Short Essay ROUGE-1 Mean	Long Essay ROUGE-1 Mean
(A) Baseline	0.063	0.032	0.202
(E) Baseline +LNMT +WS + ET	0.128	0.114	0.190

4.1 Comparison of the Short and Long Essays

Table 4 shows the comparison of the short and long essay ROUGE-1 means. It clearly indicates that the performance improvement of the proposed methods, compared with the baseline comes from short essay.

The short essay ROUGE-1 means of both proposed methods are almost half or less than those of long essays. One of the reasons of this gap between short and long essay ROUGE mean can be attributed to the short essay question answering scheme. As described in the section 1, the answer of the short essay question often contains factoid answers as a part of the essay (i.e. “In 30 English words or less, indicate the name of this Merovingian dynasty king and explain what kind of religion he converted to.”). Since the QA systems studied in this paper generate essays by BoW search based on the question, the answer of the factoid part is often unsolved. In addition, from the aspect of the probability, getting ROUGE-1 score in a long essay is easier than short essay. Generally long essay contains 5-10 sentences, and if one of them matches to the part of the gold standard, the system answer can get non-zero score. However, in short essays, the answer usually have only one sentence. Therefore, long essay answer has approx. 5-10 times larger chance to get positive ROUGE score than short essay.

5 CONCLUSIONS

In this paper, the methodology and its evaluation results for essay question answering for a narrow domain by utilizing linked open data was discussed. The proposed method translates narrow domain knowledge resources (Japanese world history textbooks) by utilizing Wikidata. The evaluation result indicated that the proposed method showed better performance compared with the baseline method [10] and the previous research [4]. The result of the proposed system was almost equivalent to the well designed Wikipedia based system [8].

The result of this paper concludes that 1) simple neural translation of knowledge resource does not work for domain specific cross-lingual question answering, 2) linked open data is effective to find correct translation for difficult terms in machine translation process, and 3) adding source language open knowledge resource would help even if its content is not equivalent with the target knowledge resources.

A LINKED OPEN DATA ASSISTED MACHINE TRANSLATION EXAMPLE

At first, we extract text from Japanese world history textbooks as follows:

> イギリスで増大しつづける中国茶（紅茶）の消費に対して，イギリス東インド会社はしだいに銀による支払いが追いつかなくなっていた。そこで，１８世紀末から，イギリスはインドでアヘンの専売制を始め，専売による財源の増加とアヘンを中国に売却することによって，茶の支払いにあてようとした。１８３９年，アヘン弛禁派をおさえ，厳禁派の林則徐が欽差大臣として広州に派遣され，アヘン密輸問題の解決にあたった。彼は外国商人がもつアヘンを没収してそれを廃棄した。アヘン貿易商人はこれに強く反発し，イギリス議会ではグラッドストンらによる“恥ずべき戦争”という反対にあったが，９票差で戦争を決定した。イギリス軍は沿岸の各地で清軍をやぶりながら北上し，１８４２年清朝と南京条約を結んだ。イギリスにつづき，１８４４年にアメリカ（望厦条約）とフランス（黄埔条約）も同様な条約を結び，清朝に条約の完全履行をせまった。しかし，華夷思想（中華思想）にもとづく朝貢外交の様式と異なるこの条約外交は，広州へのイギリス人の入城に対して地方官僚がそれを拒否するなど多くの摩擦を発生させることとなった。

Then, the bilingual world history term corpus is applied to the text:

> United Kingdom で増大しつづける China 茶（紅茶）の消費に対して，East India Company はしだいに銀による支払いが追いつかなくなっていた。そこで，１８世紀末から，United Kingdom は India でアヘンの専売制を始め，専売による財源の増加とアヘンを China に売却することによって，茶の支払いにあてようとした。１８３９年，アヘン弛禁派をおさえ，厳禁派の Lin Zexu が Imperial Commissioner として Guangzhou に派遣され，アヘン密輸問題の解決にあたった。彼は外国商人がもつアヘンを没収してそれを廃棄した。アヘン貿易商人はこれに強く反発し，United Kingdom 議会では Gladstone らによる“恥ずべき戦争”という反対にあったが，９票差で戦争を決定した。British Armed Forces は沿岸の各地で清軍をやぶりながら北上し，１８４２年清朝と Treaty of Nanjing を結んだ。United Kingdom につづき，１８４４年に United States of America（Treaty of Wanghia）と France （Treaty of Whampoa）も同様な条約を結び，清朝に条約の完全履行をせまった。しかし，Sinocentrism （Sinocentrism）にもとづく朝貢外交の様式と異なるこの条約外交は，Guangzhou への British people の入城に対して地方官僚がそれを拒否するなど多くの摩擦を発生させることとなった。

Finally, machine translation (in this paper, we used Microsoft Bing Translator) is applied:

> The East India Company was gradually unable to keep up with silver payments for the consumption of a growing *China tea (tea)* in United Kingdom. So, since the end of the eighteenth century, United Kingdom began to pay for tea by starting with the monopoly system of opium in India and by selling opium to China. In 1839, the prohibition of opium Sag, Lin Zexu was dispatched to Guangzhou as Imperial Commissioner to solve the problem of opium smuggling. He confiscated the opium of the foreign merchant and disposed of it. The opium trade merchant rallied strongly against this, and the *United Kingdom Congress* was opposed to the "shameful war" by Gladstone, but the war was decided by nine votes. British armed Forces north while veils the Qing army in various parts of the coast, and tied the *Senyi morning* and Treaty of Nanjing in 1842. United Kingdom, United States of America (Treaty of Wanghia) and France (Treaty of Whampoa) signed a similar treaty to the Qing Dynasty in 1844 to fully implement the Treaty. However, this treaty diplomacy which differs from the style of tribute diplomacy based on Sinocentrism (Sinocentrism) has caused a lot of friction that local bureaucrats refuse it for the entry of British people to Guangzhou.

In this machine translated text, there are three translation errors. “China tea,” “United Kingdom Congress” and “Senyi mornig” should be “Chinese tea” “British Parliament” and “Qing Dynasty,” respectively.

However, if we use machine translation only, there are six mistranslations; “the zexu,” “Minister of the Qin,” “Senyi morning,” “the Treaty,” “Huang Tai Treaty,” and “Chinese thought,” as follows:

> The British East India Company was gradually unable to keep up with silver payments for consumption of growing Chinese tea. So, since the end of the eighteenth century, the British tried to pay for the tea by starting the monopoly system of opium in India, increasing the financial resources and selling opium to China. In 1839, the prohibition of opium-sag, and

> *the zexu* of the Forbidden faction was dispatched to Guangzhou as *Minister of the Qin*, and the settlement of the opium smuggling problem was resolved. He confiscated the opium of the foreign merchant and disposed of it. The opium trade merchant rallied strongly against this, and the British Parliament was opposed to the ' shameful war ' by the Gladstone, but the war was decided by nine votes. The British Army veils the Qing army in various parts of the coast, and it tied the Nanjing Treaty with *Senyi morning* in 1842. In 1844, the United States (*the Treaty*) and France (*Huang Tai Treaty*) signed a similar treaty to the United Kingdom, and the Qing Dynasty concluded the full implementation of the Treaty. However, this treaty diplomacy, which differs from the style of tribute diplomacy based on *Chinese thought*, has caused a lot of friction, such as local bureaucrats refusing to enter the British into Guangzhou.

Compared with the linked open data assisted translated text, the mistranslations in this text are serious. For example, the name of treaty or person name are vanished or wrong. Since the names of treaty, person, dynasty, and so on often appear as the required keywords in answer or the important keywords for document retrieval in the question, losing this kind of terms can cause a serious problem.

ACKNOWLEDGMENTS

The authors thank Mr. Kotaro Sakamoto and Prof. Eric Nyberg for assistance with the development of the essay QA software.

REFERENCES

[1] 2016. *NTCIR-12 QA Lab-2*. http://research.nii.ac.jp/qalab/qalab2/index.html.

[2] 2017. *Boundless.com*. https://www.boundless.com/.

[3] 2017. *NTCIR-13 QA Lab-3*. http://research.nii.ac.jp/qalab/.

[4] Min-Yuh Day, Cheng-Chia Tsai, Wei-Chun Chuang, Jin-Kun Lin, Hsiu-Yuan Chang, Tzu-Jui Sun, Yuan-Jie Tsai, Yi-Heng Chiang, Cheng-Zhi Han, Wei-Ming Chen, Yun-Da Tsai, Yi-Jing Lin, Yue-Da Lin, Yu-Ming Guo, Ching-Yuan Chien, and Cheng-Hung Lee. 2016. IMTKU Question Answering System for World History Exams at NTCIR-12 QA Lab2. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, June 7-10, 2016.*

[5] Ai Kawazoe, Yusuke Miyao, Takuya Matsuzaki, Hikaru Yokono, and Noriko Arai. 2013. World history ontology for reasoning truth/falsehood of sentences: Event classification to fill in the gaps between knowledge resources and natural language texts. In *JSAI International Symposium on Artificial Intelligence*. Springer, 42–50.

[6] Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying Conditional Random Fields to Japanese Morphological Analysis.. In *EMNLP*, Vol. 4. 230–237.

[7] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop*, Vol. 8. Barcelona, Spain.

[8] Takaaki Matsumoto, Francesco Ciannella, Fadi Botros, Evan Chan, Cheng-Ta Chung, Keyang Xu, Tian Tian, and Teruko Mitamura. 2017. Wikipedia Based Essay Question Answering System for University Entrance Examination. In *Proceedings of Open Knowledge Base and Question Answering Workshop at SIGIR.* (to appear).

[9] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global Vectors for Word Representation.. In *EMNLP*, Vol. 14. 1532–1543.

[10] Kotaro Sakamoto. 2017. *FelisCatus Zero-multilingual*. https://github.com/ktr-skmt/FelisCatusZero-multilingual.

[11] Kotaro Sakamoto, Madoka Ishioroshi, Hyogo Matsui, Takahisa Jin, Fuyuki Wada, Shu Nakayama, Hideyuki Shibuki, Tatsunori Mori, and Noriko Kando. 2016. Forst: Question Answering System for Second-stage Examinations at NTCIR-12 QA Lab-2 Task. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, June 7-10, 2016.*

[12] Kotaro Sakamoto, Takaaki Matsumoto, Madoka Ishioroshi, Hideyuki Shibuki, Tatsunori Mori, Noriko Kando, and Teruko Mitamura. 2017. FelisCatusZero: A world history essay question answering for the University of Tokyo ' s entrance exam. In *Proceedings of Open Knowledge Base and Question Answering Workshop at SIGIR.* (to appear).

[13] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, and Noriko Kando. 2016. Overview of the NTCIR-12 QA Lab-2 Task. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, June 7-10, 2016.*

[14] Hideyuki Shibuki, Kotaro Sakamoto, Yoshinobu Kano, Teruko Mitamura, Madoka Ishioroshi, Kelly Y. Itakura, Di Wang, Tatsunori Mori, and Noriko Kando. 2014. Overview of the NTCIR-11 QA-Lab Task. In *Proceedings of the 11th NTCIR Conference on Evaluation of Information Access Technologies, NTCIR-11, National Center of Sciences, Tokyo, Japan, December 9-12, 2014.*