MTMT in QALab-3: World History Essay Question Answering System that Utilizes Textbooks and Open Knowledge Bases

Takaaki Matsumoto
Carnegie Mellon University
SOC Corporation
takaaki.m@gmail.com

Teruko Mitamura
Carnegie Mellon University
teruko@cs.cmu.edu

ABSTRACT

This paper introduces the system and its evaluation for answering world history essay questions by utilizing linked open data which assists machine translation. Since the target questions are the world history subject of the entrance examination of the University of Tokyo, most answers can be found in the Japanese world history textbooks. However, an equivalent content of high-quality English translation of the Japanese world history textbooks is not available. Therefore, we try to translate those textbooks utilizing linked open data, and using source language knowledge resource of which content is not equivalent with the target knowledge resource.

The evaluation result indicates that the proposed system shows the best ROUGE-1 scores of all the end-to-end submissions [13]. The result of this paper concludes followings. 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. 3) Adding source language open knowledge resource would help even if its content is not equivalent to the target knowledge resources.

Keywords

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

Team Name

MTMT

Subtasks

English subtask (Phase 2: End-to-End for Essay Questions)

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 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 a 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][12][13]. In the QA Lab, there are three types of questions; multiple-choice, term (factoid), and essay. 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 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 Section 4, the evaluation result of the proposed system is compared with other submissions for the NTCIR-13 QA Lab-3.

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 [12][10], 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. [9][11]. 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 used as “morning” generally. 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 shows some nouns (especially, compound noun) were mistranslated by the latest neural transition, and Wikidata.org translated them perfectly. Therefore, 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 is greater than two, the translated results are ambiguous, and they are not utilized.

The second strategy is to avoid mistranslation. This approach 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, the morphological analysis (MeCab [6]) is applied, and a 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 four characters, and found only approx. 100 mistranslations in the results. Finally, 6,962 Japanese terms and their English translations were retrieved. Also, approx. 2,000 Japanese-English translation pairs 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 the commercial machine translation. First, all terms that match with the bilingual corpus in the whole Japanese text are replaced by English terms, and then a Japanese-English mixed text is generated. After that, the text is translated by the neural machine translation API. In this paper, we used the 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 the Microsoft Bing Translator into:

“Also according to the Institute of the Christian region.”

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

Table 1: Translation Examples

Japanese Term	Google Translate (Apr. 2017)	Bing Translator (Apr. 2017)	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 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 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 “longest 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, the exact match has precedence over searching in our algorithm, and the ambiguity problem does not happen if the exact match succeeds.

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 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 other 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 very efficient for the machine translation of the textbooks to translate rare

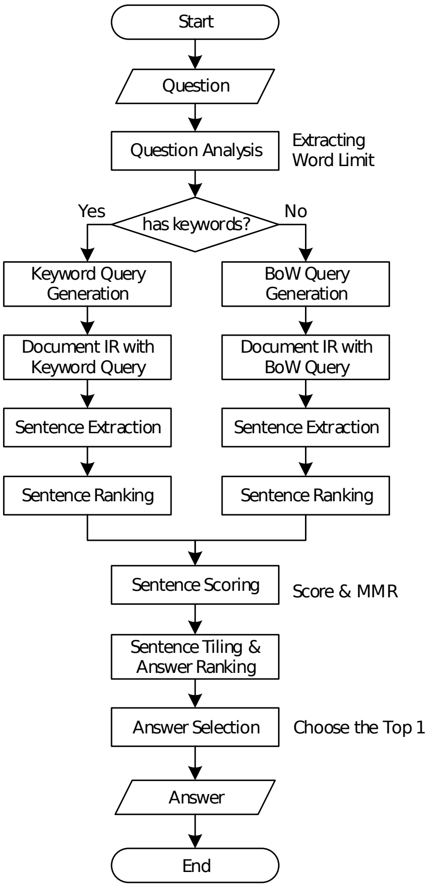

Figure 1: System Flowchart.

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 online, the textbook of Boundless.com is an 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 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 words (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 the order of the extracted sentences. The score of an essay candidate is the 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 [8]. 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

We submitted two results (MTMT1 and MTMT2). The MTMT1 employs the proposed knowledge resources and the sentence scoring method. The knowledge bases are the translated textbooks with linked open data and neural machine translation technique, and the open world history textbook. The MTMT2 is almost the same as the baseline system (some minor bugs were fixed) with the five machine translated Japanese world history textbooks which were translated in 2015 with Google translate.

Table 3 shows the ROUGE-1 and ROUGE-2 evaluation results of all the systems submitted for the NTCIR 13 QA Lab-3 phase-2 end-to-end run [13]. The official phase-2 dataset contains five long/complex and 22 short/simple essay questions [3]. The evaluation is done by human experts, ROUGE

Table 3: End-to-end Evaluation Results in ROUGE-1 and ROUGE-2 [13]

System	ROUGE-1 Mean (Case)	ROUGE-2 Mean (Case)	ROUGE-1 Mean (Stem)	ROUGE-2 Mean (Stem)	ROUGE-1 Mean (Stopword)	ROUGE-2 Mean (Stopword)
CMUQA1	0.1231	0.0159	0.1257	0.0164	0.0347	0.0065
CMUQA2	0.1251	0.0160	0.1278	0.0166	0.0371	0.0066
CMUQA3	0.0768	0.0105	0.0968	0.0184	0.0594	**0.0129**
Forst1	0.1057	0.0086	0.1057	0.0086	0.0183	0.0022
IMTKU1	0.1187	0.0133	0.1239	0.0156	0.0457	0.0013
IMTKU2	0.0065	0.0000	0.0101	0.0001	0.0126	0.0001
MTMT1	**0.1717**	0.0180	**0.1841**	**0.0207**	**0.0615**	0.0054
MTMT2	0.1546	**0.0190**	0.1630	0.0197	0.0465	0.0020

Table 4: Short and Long Essays Comparison [13]. The grammatical errors mean the unusual spelling, and the extra and unnecessary words and phrases. The semantic errors include sentences which cannot or hardly understood in the context or has any inconsistency inside itself, and the misinterpretation about facts. The number of missing keywords (KW) shows how many given keywords which must be included in the essay did not exist in the system answer.

	Short Essay		Long Essay					
System	ROUGE-1 Mean (Stopword)	ROUGE-2 Mean (Stopword)	ROUGE-1 Mean (Stopword)	ROUGE-2 Mean (Stopword)	Expert Score Mean	Num. of Grammatical Err. Mean	Num. of Semantic Err. Mean	Num. of Missing KW Mean
CMUQA1	0.0152	0.0054	0.1204	0.0115	0 (-26.2)	7.8	0.2	3.6
CMUQA2	0.0173	0.0054	0.1241	0.0119				
CMUQA3	**0.0442**	**0.0131**	0.1263	0.0119				
Forst1	0.0225	0.0027	0.0000	0.0000				
IMTKU1	0.0262	0.0000	0.1315	0.0069	0 (-29)	6.0	0	4.6
IMTKU2	0.0124	0.0000	0.0132	0.0004				
MTMT1	0.0370	0.0023	**0.1692**	**0.0190**	0 (-36.4)	10.8	1.8	4.4
MTMT2	0.0255	0.0000	0.1387	0.0107				
DGLab1 SumExP			0.1428	0.0139	0 (-22)	6.8	0.4	3.0
DGLab2 SumExP			0.1458	0.0138				

method and Pyramid method [3]. In this paper, ROUGE-1 and 2, uni-grams and bi-grams to compare the essay to a set of gold-standard essays, are used for evaluation.

Table 3 indicates that our proposed system (MTMT1) showed the best ROUGE-1 means in all (Case, Stem and Stopword) end-to-end evaluations. Since for all ROUGE means except ROUGE-2 mean (Case) the MTMT1 indicated better performance compared with the MTMT2, the proposed method is considered to be effective. Also, the fact that the CMUQA2 employed the same scoring method as MTMT1 but had less ROUGE means the usefulness of the knowledge resources of the proposed method. The reason for the good ROUGE-1 means of the MTMT1 can be attributed to the proper entity names of the knowledge resources and the similarity measurement in the sentence scoring process.

4.1 Error Analysis

Table 4 shows the comparison of the short and long essay ROUGE-1 and ROUGE-2 means. It indicates that the proposed system (MTMT1) showed the second best ROUGE-1 and 2 (Stopword) means in the short essays and the best ROUGE averages in the long essays. As for the expert score, all submissions were zero points. However, when we relax the scoring to negative values, the MTMT1 was worst of all due to grammatical and semantic errors. The number of missing keywords (KW) is slightly higher than those of other systems.

Due to the two steps translation of the proposed method, some grammatical and semantic errors occurred. For example, in the system answer of the G792W10, the following text was marked as a sentence which cannot be understood in the context by the human expert.

> Thus in strong economic nationalism, internationalism after World War momentum is gone left.

This ungrammatical sentence can be regarded as the side effect of the proposed two steps translation. The original Japanese sentence and its correct translation which is simply obtained by the Bing Translator are following:

> こうして各国で経済ナショナリズムが強まり，大戦後の国際協調の気運は消えさった。
> Thus, economic nationalism strengthened in each country, and the momentum of international co-operation after the war disappeared.

In the Wikidata.org, there are no entries of 経済ナショナリズム and 国際協調, but still one search results for them, respectively. We obtain "economic nationalism" and "internationalism." "economic nationalism" is the correct translation for 経済ナショナリズム. However, the word "internationalism" is similar but incorrect translation; it is a word for 国際協調主義, not 国際協調.

Hence, the system generates a Japanese-English mixed text as follows:

> こうして各国で economic nationalism が強まり，
> 大戦後の internationalism の気運は消えさった。

The system applies the Bing Translator to the text, and the above ungrammatical sentence appeared. The machine translation for this kind of Japanese-English mixed sentence sometimes fails. However, the latest version of the Bing Translator can translate this Japanese-English Mixed sentence grammatically as follows:

> The economic nationalism strengthened in each country, and the momentum of internationalism after the war disappeared.

Therefore, when we try the proposed translation process again with the latest neural machine translation, we would get a better result.

Another possible way to avoid ungrammatical translation is to control the LOD pre-translation based on the term frequency. Since the neural machine translation sometimes fails to translate Japanese-English mixed sentences, by suppressing too easy pre-translations such as country name, which is considered to have a high term frequency, the mistranslation will be reduced.

5. CONCLUSIONS

In this paper, the methodology and its evaluation results for essay question answering for a narrow domain by utilizing linked open data were discussed. The proposed method translates narrow domain knowledge resources (Japanese world history textbooks) by utilizing Wikidata. The evaluation results indicated that the proposed method showed the best performance compared with other end-to-end submissions [13].

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. Improving the grammaticality of the translated text is a future task.

Acknowledgment

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

APPENDIX

A. LINKED OPEN DATA ASSISTED MACHINE TRANSLATION EXAMPLE

Firstly, we extract a text from Japanese world history textbooks. Then the bilingual world history term corpus is applied to the text as follows:

> 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, the neural machine translation (in this paper, we used the 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 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 the 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 names of the 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.

B. REFERENCES

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

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

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

[4] M. Day, C. Tsai, W. Chuang, J. Lin, H. Chang, T. Sun, Y. Tsai, Y. Chiang, C. Han, W. Chen, Y. Tsai, Y. Lin, Y. Lin, Y. Guo, C. Chien, and C. Lee. 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*, 2016.

[5] A. Kawazoe, Y. Miyao, T. Matsuzaki, H. Yokono, and N. Arai. 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*, pages 42–50. Springer, 2013.

[6] T. Kudo, K. Yamamoto, and Y. Matsumoto. Applying conditional random fields to japanese morphological analysis. In *EMNLP*, volume 4, pages 230–237, 2004.

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

[8] J. Pennington, R. Socher, and C. D. Manning. Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532–1543, 2014.

[9] K. Sakamoto. *FelisCatus Zero-multilingual*, 2017. https://github.com/ktr-skmt/FelisCatusZero-multilingual.

[10] K. Sakamoto, M. Ishioroshi, H. Matsui, T. Jin, F. Wada, S. Nakayama, H. Shibuki, T. Mori, and N. Kando. 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*, 2016.

[11] K. Sakamoto, T. Matsumoto, M. Ishioroshi, H. Shibuki, T. Mori, N. Kando, and T. Mitamura. 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*, 2017.

[12] H. Shibuki, K. Sakamoto, M. Ishioroshi, A. Fujita, Y. Kano, T. Mitamura, T. Mori, and N. Kando. 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*, 2016.

[13] H. Shibuki, K. Sakamoto, M. Ishioroshi, Y. Kano, T. Mitamura, T. Mori, and N. Kando. Overview of the NTCIR-13 QA lab-3 task. In *Proceedings of the 13th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, December 5-8, 2017*, 2017.

[14] H. Shibuki, K. Sakamoto, Y. Kano, T. Mitamura, M. Ishioroshi, K. Y. Itakura, D. Wang, T. Mori, and N. Kando. 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*, 2014.