Evaluation of Machine Translation System from Japanese to Arabic for Cross-Language Information Retrieval

Albohassan Ibrahim[1],   Saori Aida[2], and Hiroyuki Kameda[2]

1 Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology

2 School of Computer Science, Tokyo University of Technology

1. Introduction

In the world, there are various kinds of text data on the Web systems. For example, some texts are written in English and some in Japanese. This fact means cross language information retrieval (CLIR) is strongly useful for people in the street to acquire various facts and knowledge. From this point of view, we try to establish new CLIR methodology, especially between Japanese and Arabic.

This paper reports our research work to reach that research goal. More concretely to speak, we report our analysis of outputs of Japanese-Arabic machine translation systems, Google MT and Babylon MT.

2. Overview of Research procedure

We do research in the following three steps.

[Step 1] Gathering text data: In our investigation we first gathered sentences out of three levels of Japanese text on the basis of the educational curriculum for Japanese. Text are adopted from three types of linguistic corpora on the basis of three basic levels, i.e., Elementary level, Intermediate level, and Advanced level defined in the field of educational curriculum for foreign students, who study Japanese language in Japan. The total number of sentences gathered is 222.

[Step 2] Translation from Japanese to Arabic by MT systems: The text gathered are translated by Google MT and Babylon MT. Texts are input to the system manually by copy-and-paste way. The MT systems run on Windows OS on a Note PC (Toshiba DynaBook R643).

[Step 3] Analysis of output: Translated text from Japanese to Arabic are checked and evaluated manually one by one. Of course, one of the smart way is to use automatic evaluation software like Blue and etc. But in our study, we checked translated text manually, so that we really understand what is going in the machine translation systems,

according to the proverb "To see is to believe." Evaluation was done by a single evaluator, whose mother tongue is Arabic. He, of age32, can also take a command of Japanese not fluently but fairly well.

3. Results of machine translation

Results was analized from the following indices, **Gramatical correctness**, **Spelling correctness**, **meaning correctness**, and **abnormality**.

(1) Grammatical correctness: Grammatical correctness depends on the level of Japanese, i.e., elementary, intermediate, and advanced. Japanese text in the level of elementary tends to be translated more correctly than other level of Japanese. And the text of advanced level was poorly translated in the view of grammar. This suggests that previous MT systems such as Google and Babylon are well trained with use of elementary and fundamental sentences. On the other hands, text of advanced level tends to have characteristics of diversity usually.

(2) Spelling correctness: Evaluation from the view of spelling correctness also depends on the level of text. Text of elementary level is more correctly translated than those of intermediate and advanced levels.

(3) Meaning correctness: From this viewpoint of meaning correctness, the case is the same as above (1) and (2).

(4) Abnormality: Both Google translation system and Babylon translation system shows many output of abnormallyty. Output by Babylon was worse than that of Goole in our experiment.

5. Conclusion

This paper reports our preliminary analysis of outputs of Japanese-Arabic machine translation systems by Google MT and Babylon MT.

REFERENCES

[1]MAHSOUB ABDUL-SADEQ ALY (2004). "Translation Strategies of EFL Student Teachers: A Think Aloud Protocol-Based Case Study." *ERIC*, ED490356.

[2]Bader S. Dweik. and Mohammed B. Thalji. (2016). "STRATEGIES FOR TRANSLATING PROVERBS FROM ENGLISH INTO ARABIC." *Academic Research International*, **7**(2), pp.120-127.

[3]Bahaa-eddin Abulhassan Hassan. (2014). "Between English and Arabic: A Practical Course in Translation." Cambridge Scholars Publishing.

[4]Amjad, Fazel Asadi. and Farahani, Mohammad. (2013). "Problems and Strategies in English Translation of Quranic Divine Names." *International Journal of Linguistics*, **5**(1), pp.128-142.

Table 1. Sample text of elementary level

#		Grammatically correct	Spelling correct	Meaning correct	Abnormality
1	明日は土曜日です。	●	●	●	✖
2	時計は絵の左にあります。	●	●	●	✖
3	サッカーが一番面白いです。	●	●	●	✖
4	学校の食堂で食べました。	●	●	●	✖
5	試験の時隣の人の答えを見ます。	●	●	●	✖

Table 2. Sample text of intermediate level

#		Grammatically correct	Spelling Correct	Meaning correct	Abnormality
1	こうなった上は	✖	✖	▲	●
2	それを済ませた上で／よく考えた上のこと／見かけ の上では	✖	●	▲	●
3	学生である以上／約束した以上は	✖	▲	▲	✖
4	日本の歴史について／A社の業績につき／時事問題 についての話／本日は祭日につき	▲	▲	●	✖
5	中国人に違いない／彼は知っているに違いない	●	●	●	✖

Table 3. Sample text of advanced level

#		Grammatical Correct	Spelling correct	Meaning correct	Abnormality
1	穴が開く／穴を開ける	▲	▲	✖	●
2	後へ引けない／後に引けない	▲	▲	✖	●
3	意に介さず／意に介さない／意に介しない	▲	▲	✖	●
4	年が改まる／もうすぐ年が改まる。	▲	▲	✖	●
5	年を食う／本人が言うよりも年を食っているように見える。	▲	▲	✖	●