AN EMPIRICAL STUDY OF ONLINE TRANSLATION OF ARABIC LANGUAGE INTO ENGLISH LANGUAGE: IMPLICATIONS ON STUDENTS' LEARNING PURPOSES

FHA. Shibly[1], SI. Nimsith[2] and ARFS. Farwin[3]

[1] Lecturer in Information Technology, Department of Arabic Language, Faculty of Islamic Studies and Arabic Language, South Eastern University of Sri Lanka, shiblymis@gmail.com

[2] Lecturer in Accountancy & Finance, Department of Arabic Language, Faculty of Islamic Studies and Arabic Language, South Eastern University of Sri Lanka, sinimsith@gmail.com

[3] Lecturer in Arabic Linguistic & Translation, Department of Arabic Language, Faculty of Islamic Studies and Arabic Language, South Eastern University of Sri Lanka, farwinarfs@gmail.com

Introduction

Researchers in the field of natural languages have undertaken a serious effort to support manual translations by inventing machine translations. Hutchins (1986, p: 15) defines Machine Translation (MT) as "the application of computers in the translation of texts, from one natural language into another". Also known as automatic translation, MT has also been considered in the last decade as a computational linguistic phenomenon. Apparently, MT is considered as a worthwhile subject for researchers, commercial developers and users (Hovyet al. 2002). As for researchers, they need to apply their theories to find out the differences that might be made by the machines. By doing so, it will be easier for developers to detect the most problematic issues and make the implementations on the system design. Evidently, the motive of commercial developers is to attract customers to buy their products. In turn, the users, who are interested in benefitting from MT, will decide which product meets their requirements. Examples of previous researches and studies include the employment of various approaches to MT, such as studies by Marcu (2001), Richardson et al (2001), Groves (2006) and Tahir et al. (2010). Earlier researches focused on the direct approach such as the word-by-word analysis of the source language. Later on, researchers moved to the rule-based and statistical approaches. Salem (2009) is an example of this research trend. Meanwhile, there were researchers who were interested in the evaluation of MT quality since the users' demand increased for the use of machines with high levels of translation quality according to the rapid growth of technology and information. Different methods have been employed in measuring the quality of MT outputs according to different criteria outputs, such as *Fluency* and *Fidelity* (Eduard Hovy et al. 2002, p.45). Some researchers analysed MT outputs for different purposes focusing on specific features; for instance, agreement of number, and relative clauses (Flanagan, 1994). Others used the judgment of evaluators to rate whole sentences in terms of the N-point scale (White et al., 1992, 1994; Doyon et al., 1998), while others made use of the "bigram or trigram language model of ideal translation" to automatically measure the confusion which resulted from complexities in the target text (Papineni et al. 2001).

The motivation of the study in conducting Google because Google Translation has been proven, to be "the most powerful and accurate of any of the readily available machine translation tools"(Och, 2006). In the same study, a statement implies that, the developed machine translation can be achieved "without the need to understand the individual languages or their specific rules" (Och, 2006). On the other hand, Babylon is a computer dictionary and translation program for Microsoft Windows. The first version of Babylon was introduced in 1997. Within one year, in 1998, its number of users increased enormously and reached 4 million. Furthermore, in the year 2011, it became one of the most popular language translation applications. It can translate a full (text, Web page, and document) in 33 languages. It has a technical term, by including built-in dictionaries and community dictionaries.

Finally, translation quality is a concept which relates to the output of the translation, whether it is by a human or machine process. Linguists,

philosophers and scholars are continuously discussing about the applicable criteria for good translations in order to assess their quality. This study aims to determine a better MT by comparing Google and Babylon, which would be more appropriate to be used in translating Arabic news headlines into English in terms of the Hutchins and Somers criteria (viz. clarity, accuracy and style).

Methodology

The study makes use of Hutchins and Somers criteria which could be summarized as follows: The Criteria of Hutchins and Somers of Evaluation It is important to emphasize that one of the main purposes of this study is derived from the role of evaluation, as to find out what machine translation systems are able and not able to do, according to the view of misunderstandings and misconceptions of transmitted message of news headlines. The evaluation is restricted on testing the raw outputs of two machine systems, specifically Google and Babylon, in reference to the manual translation that is available by the source of the data. The testing focussed on evaluating the quality of raw outputs based on the most basic principles of machine translation evaluation rather than to focus on the operations within the potential environments of systems, as it is the task of system developers. Some of these principles are: fidelity, intelligibility, and style, which they have been reflected by Hutchins and Somers (1992).

The following represents the summary of these principles: *Fidelity* represents the accuracy of machine translation performance. It also means to what extent that the translated output has the 'same' information as the original. On the other hand, *Intelligibility* principle expresses the clarity in the translation output. In other words, it represents that the translated output should be free from obscurity, comprehensive, and understandable. The last one is *Style*, which expresses to what extent the translation has used the language, suitable to its content and purposes.

Data of the Study

There were 20 news headlines, which were randomly chosen from three different Arabic journals, namely www.daralhayat.com,www. aljazeera.net, and www.asharqalawsat.com, dating from 1st to 30th September (2013). The choice of these data is based on the availability of their human English translation.

Procedures of Analysis

The main procedures used in achieving the objectives of this research are stated below:

1. Collecting the data of the study which consist of Arabic news headlines with their English manual translated versions from online sources.
2. Each Arabic headline once will run into Google translator, and then into Babylon translator, to be translated into English.
3. The outputs of both Google and Babylon were listed in one table.
4. To fulfill the evaluation objective, the researchers had distributed a questionnaire to a group of 15 evaluators. The distributed questionnaire was based on the criteria provided by Hutchins and Somers (1992). The group of evaluators consists of 15 professionals whose second language is Arabic, and who work in different Iraqi, Egypt, Malaysian and Sri Lankan Universities, and have good English Language proficiency.

The Evaluators Assessment

This part is the most important process, which is to calculate the human judgments based on the assigned questionnaire. The present study conducted machine-translations of twenty (20) Arabic news headlines into English. The evaluators were asked to consider each Arabic headline and its machine-translated outputs to examine the three parameters which are provided in the questionnaire. The parameters consisted of three criteria: Clarity, Accuracy, and Style. Each criterion is defined according to Hutchins and Somers (1992). For each criterion there were 4 scores. There were 15 evaluators who participated in the assigned questionnaire. The average of each output was calculated based on the following statistical equation:

$$Av = \frac{X1 + X2 + X3 \ldots \ldots + X15}{n\ (evaluators)}$$

$$av = average$$
$$x = score\ of\ the\ evaluator$$
$$n = number\ of\ evaluators$$

Then, by summing up the averages of each outputs of the same parameter and dividing them by the number of outputs, we obtained the total average of each parameter according to the following equation:

$$Total\ Av = \frac{Av1 + Av2 + Av3 \ldots + Av20}{n(outputs)}$$

Av = averages of each outputs
n = number of output

Findings

The following sections will show the results of each criterion for each system. The results are based on the evaluators' assessment of the provided questionnaire, as well as the results of the preferred system in translating such data.

1. Clarity:

Clarity was the first parameter in which the participants were asked to evaluate. There were only minimal differences between the clarity of the Google and Babylon translations for each of the twenty (20) outputs of headlines. From Figure 1, it is obviously shown that both the two translators were graded with an average of 3.2 out of the highest value of 4. We can say that the evaluators assessed both the Google and Babylon outputs as being equally understandable. The score was closest to 3, which indicates that "Mostly understandable" was the answer to the question "How easily can you understand the translation?" Accordingly, the evaluators' estimation for both Google and Babylon was 80% clarity.

Figure 1. Clarity

2. Accuracy

The second parameter to be marked by the evaluators was accuracy. Referring to Figure 2, overall, Google scored higher than Babylon in terms of accuracy. Out of the highest value of 4,

Google had an average score of 3.1, whereas the combined average score of Babylon was 3.0. The assessment of the criteria indicated that both Google and Babylon were closest to the score of 3, which gave the evaluators' answer to the question, "To what extent does the translation contain the 'same' information as the source text? "It was clear that these two averages illustrated that there was a significant variation between Google and Babylon, as shown by the following rating: 77.5% for Google and 75% for Babylon. Accordingly, Google was highly regarded by the evaluators to be more accurate than Babylon, as can be seen in the following diagram:
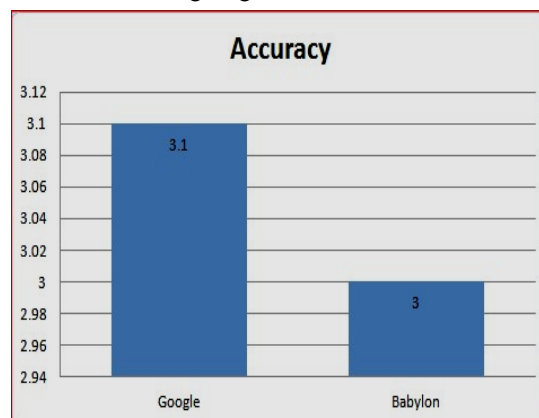

Figure 2. Accuracy

3. Style

The third parameter which the evaluators were asked to score was style. Babylon scored higher than Google, where the average for Babylon's average was 2.9 out of 4, which represented the highest rating. Google's average meanwhile was 2.8. Hence, the average of Google's style was considered as the lowest average out of the three criteria. It was apparently shown by accounting the percentage of each style average that the evaluators found that the style of the Babylon outputs was better than the style of the Google outputs. Thus, Google had 70% and Babylon had 72.5% of style. Concerning the criteria, the evaluation was based on answering the following question: "Is the language used appropriate for a software product user manual? Does it sound natural and idiomatic?" The answer revealed that Babylon somehow produced a more acceptable style in its outputs than the style of the Google outputs, as shown in the following Figure 3.
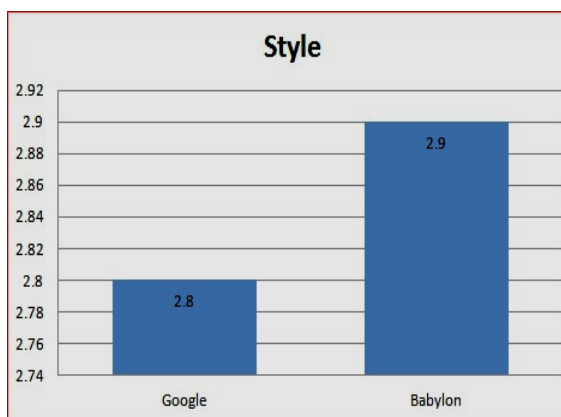
Figure 3. Style

Implications and Conclusions

Online MT can be used for the purpose of learning from school to university level because it has the characteristics of educational technologies that can help students, especially for students who want to pursue a foreign language. MT is commonly used to understand a second language text and express their ideas. MT has been shown to accelerate the translation work and very easy and time saving. MT use in translation actually shortens some steps as used in the human translation. One no longer need to search for words, flipping page after page which is certainly time consuming then write back. Instead, the software can easily translate the content and quality translation results with word choices. In the era of globalization, the dominance of such information is a value added for individuals and the organization. Information can be obtained from a variety of languages throughout the world. With the availability of MT, such information can be obtained easily and cost effective without high investment. On the other hand, if a translation done by a professional translator, translation based on a per page basis would certainly be very costly and compared to the use of MT which involves a very minimal cost.

Confidentiality is also one of the characteristics found in the nature of MT-aided translation. MT usage ensures information translated is protected whereas; the submission of documents which holds sensitive information may risk leakage if given to a human translator. The software in MT has been designed for use in universal fields. MT is very suitable for use in science, literature, languages and linguistics, and others whereas; human translation only covers specific areas of expertise.

Undoubtedly, MT has many benefits that can help university students transfer information into preferred language. It is necessary for them to be more cautious when doing advance and professional translation work since there are areas that cannot be translated as cultural aspects associated with the accuracy of meaning which cannot be produced by machine translation consistently. One can only obtain information in the form or essence of the draft document and it is not necessarily fully accurate. This is because MT is only capable of conducting literal translation of the words without understanding the actual information in context that may need to be corrected manually later. Another flip side of MT is that it cannot handle ambiguities that exist because it was created under the laws of systematic and formal rules of the language and certainly could not translate words based on experience, emotions, values, and mental outlook compared to human translation. However, online machine translation systems are continuously undergoing development, and the outputs might be improved in the near future to help students' learning more effectively.

References

1. Alsalman, S. (2004).The Effectiveness of Machine Translation. In IJAIS, International Journal of Arabic- English Studies, 5, pp.145-160
2. Abraham, I., Salim, R.(2005). A Maximum Entropy Word Aligner for Arabic-English Machine Translation. Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP), pp. 89–96, Vancouver.
3. Chalabi,A. (2001). Sakhr Web-based Arabic/English MT Engine. <www.elsnet.org/arabic2001/chalabi.pdf> [10 August 2010].
4. Farghaly, A., Senellart, J. (2003). Intuitive Coding of the Arabic Lexicon. A paper read at The MT Summit IX –Workshop: Machine translation for semitic languages, New Orleans, USA.
5. Hutchins, W.J. (1986). Machine Translation.Past, Present, Future. West Sussex, England: Ellis Horwood Limited.
6. Hutchins, J. and Somers, H. (1992).An Introduction to Machine Translation. London: Academic Press Limited
7. Papineni, K., Salim, R., Todd, W., Wei-Jing, Z.(2002). Bleu: a method for automatic evaluation of machine translation. Proceedings of the 40th annual meeting on association for computational linguistics, pages 311–318.
8. Ryding, Karin C. (2005). A Reference Grammar of Modern Standard Arabic.

9. White, J., O'Connell, T. and O'Mara, F. (1994) "The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Future Approaches". Proceedings of the 1st Conference of the Association.

10. Zughoul, M, Abu-Alshaar, A. (2005).s: A Historical Perspective Translators' Journal 50 (3): 1022 1041.