Problems of Arabic-English Machine Translation: Evaluation of Google's Online Machine Translation System

KHALDI Anissa
Université de Tlemcen

Abstract
The present article discusses problems of translation from Arabic to English using the online machine translation system of Google. As Arabic and English are distant languages from two unrelated families, machine translation is bound to face many problems in producing coherent translations between these languages. This article is confined to diagnosing and analysing problems related to lexicon and syntax. The aim is to highlight the danger of relying blindly on Google's online machine translation.

1. Introduction

Ever since peoples started travelling around the world, and experiencing the need for communication and trade, humankind has realized the need for understanding each other's language. This need became more urgent with the increased trade between different nations. In recent decades, the phenomenon of globalization has created more pressure in terms of needs to communicate. Since English has become the world's primary language of business negotiations; academic conferences and scientific research; and hence, communication, the need for translation has dramatically increased. In parallel to this, human search for a less expensive and less time consuming translation has led to the invention of machine translation. The term machine translation (MT) refers to use of

computer software to translate a text from one natural language to another. The present article attempts at discussing this type of technology. It first traces the history of MT. Then, the different strategies and paradigms are tackled. The final section is devoted only to some the problems that may arise from using MT in translating from Arabic to English.

2. A short History of MT

During the 1950s and before the end of 1960s there had been many enthusiastic attempts, sponsored by some governments, which aimed at developing MT. For example, the Georgetown experiment in 1954 involved fully automatic translation of more than sixty Russian sentences into English. The experiment was a great success, and sufficiently impressive to stimulate massive funding of MT in the United States and to inspire the establishment of MT projects throughout the world. The earliest systems consisted primarily of large bilingual dictionaries. However, disillusion grew as researchers encountered semantic barriers for which they saw no straightforward solutions. There were some operational systems – the Mark II system (developed by Washington University), and the Georgetown University system at the US Atomic Energy Authority – but the quality of output was disappointing. By 1964, the US government sponsors had become increasingly concerned at the lack of progress; they set up the Automatic Language Processing Advisory Committee (ALPAC). The ALPAC report (1966) concluded that MT was slower, less accurate and twice as expensive as human translation. The report caused a major reduction in U.S research and development efforts in the area of MT. It, however, recommended that tools be developed to aid translators — automatic dictionaries, for example — and that

some research in computational linguistics should continue to be supported (Hutchins 2009). The report had the same impact on research and development effort into machine translation in the Soviet Union and the United Kingdom. However, research did continue in Canada, in France and in Germany. The late 1970s and during 1980s witnessed the formulation of EUROTRA project by the European Commission to provide MT of all the member nations' languages. The project was motivated by one of the founding principles of the EU: that all citizens had the right to read any and all legal acts of the Commission in their own native language. Since the 1980s onwards, the desire for more foreign markets has stimulated research in MT, in U.S.A.

3. Strategies of MT

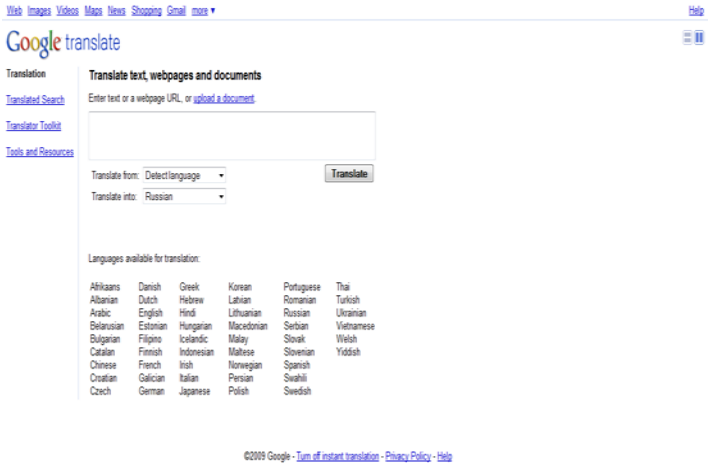
There are two strategies of MT: direct, transfer (Hutchins1995). In the direct strategy, each word in the source language is linked to its equivalent in the target language. However, there is a unidirectional correlation, for example from English to Arabic but not the other way round. The transfer strategy is currently the widely used method in MT. The source text is first analyzed by the help of a dictionary of the source language. This is called the analysis stage. Then, the transfer stage changes the results of the analysis stage and produces the linguistic and structural equivalents between the two languages. A bilingual dictionary is used at this stage. The third stage is the generation stage which produces the target text based on linguistic data of the source language by using a target language dictionary.

4. Paradigms of MT

Whilst strategies of MT refer to the actual processing design, paradigms refer to the informational components that aid the processing design. There are three main paradigms, Rules Based Machine Translation (RBMT), Statistics based (EBMT and SBMT), and Hybrid System. Rules based techniques rely on using linguistic rules for the source and target languages, and on trying to translate the meaning through dictionaries and grammatical parsing. The second main method is statistical. This method uses a body of pre-existing translation and then compares source strings to see if an existing translation exists. It Consists of EBMT (Example-Based Machine Translation: extraction of phrases for recombination) and SBMT (Statistics-Based Machine Translation: statistical translation model, based on word frequency). These systems analyze a large number of previously created bilingual sentence pairs to establish which words or expressions in one language are most frequently matched with words or expressions in the other. Hybrid System uses a combination of the above two techniques.

5. Google Translate: Lexical and Syntactic Problems in Arabic –English Translation

Google Translate is a service provided by Google. It is based on the statistical machine translation, and can translate 35 different languages to each other, forming 595 language pairs.



Google translate can translate from Arabic into English. However, the output resulting from such translation may contain some mistakes, and hence, incoherent translation between the source text and the target one. This section will explore some of these problems at the level of lexicon and syntax (Izwaini 2006).

Concerning problems related to lexis, one may include what follows:

- Some content words may be dropped and hence, they are not found the output. This is due to a technical problem related to the inability of the MT to recognize some words;
- Another problem is that words in Arabic may have to or more overlapping meanings in English. Hence, the MT system may not opt for the right word. For instance: قمة has different meanings in English: (top, climax, summit, peak). There need

to be a consideration of the surrounding context in order to be able to opt for the appropriate translation;
- In other cases, homographs can result in mistranslations. For instance:لها (to her/it) can be translated into fun.

As for syntax**,** sometimes sentences are treated as individual words, which results in almost meaningless output. Furthermore, Some Arabic coordinators may cause problems when they are combined with other words may change the meaning of the source text. Example: when ف is added to قد we get the English translation 'lost'. Finally, sentences in Arabic may be combined as SVO (subject, verb, object) or VSO (verb, subject, object). In comparison to the second order, the first one usually leads to correct translation as it corresponds to the English word order. However, when the source text has a VSO order a syntactic problem may arise, since the TM system takes the original order as it is, translating, thus, only individual words.

6. Conclusion

Translation is by no means an easy task. This process involves decoding the meaning of the source text and then re-encoding it in the target language. Google translate may provide some advantages in terms of low price, quick translation, confidentiality(i.e., translating personal messages that one does not want other people to read). However, sometimes this service may bring about incoherent output. Therefore, human intervention is needed through post-editing, that is revising and correcting the output of the MT by humans.

Abbreviations:
ALPAC: Automatic Language Processing Advisory Committee
MT : Machine Translation
US : United States of America
USA: United States of America
RBMT: Rules Based Machine Translation
SBMT: Statistics based Machine Translation

References

ALPAC (1966). Language and machines: computers in translation and linguistics. A report by the Automatic Language Processing Advisory Committee. National Academy of Sciences, Washington, DC.

HUTCHINS, J.W.(1995): Machine Translation : A brief History. In E.F.K.Koerner and R.E.Asher(eds), Concise history of the language sciences: from the Sumerians to the cognitivists(431-445). Oxford: Pergamon Press.

HUTCHINS, J. (2009): The history of machine translation in anutshell.http://ourworld.compuserve.com/homepages/WJHutchins

IZWAINI, S. (2006): Problems of Arabic Machine Translation. Proceedings of the International Conference on the Challenge of Arabic for NLP/MT. The British Computer Society (BSC), London 23rd Oct ,118-148.