The Limits of Transability for an Arabic Machine Translation: Trials & Remedies

BELMEKKI Amine &
DJEBBARI Zakia
Université de Tlemcen

Abstract

The Arabic language is gradually gaining momentous attention in today's world due both to its socio-political importance and the challenges presented mainly by its sound patterns, highly complex word formation process of roots and the nature of its writing system.

Hence, the present paper calls for an urgent awareness of the challenges with Arabic as a source of target language in machine translation (MT), i.e., it highlights the limits of transability   for an Arabic Machine Translation (AMT) system, taking into account the interchangeability of the relationship of SL and TL texts to achieve the underlined needs and challenges.

Present decisive developments in multilingualism have brought to the forefront the growing importance of translation and the urgent need for more innovative, inventing methods of translation. Thus, this article attempts to create a platform for debates for examining critically different positions regarding the limits of **transability** within an Arabic Machine Translation system. In this regards and within the newly information age, translation, seen as a knowledge-based activity, requires much more innovative strategies that consider to a large extent the logistics of information.

The majority of professional translators are employed to satisfy the huge and growing demands for translations of scientific and technical documents, commercial and business transactions, administrative memoranda, legal documentation, instruction manuals, agricultural and medical text books, industrial patents, publicity leaflets, newspaper reports, etc. Some of these tasks is challenging and difficult in the sense that is requires accuracy and consistency. The demand for such translations is increasing at a rate far beyond the capacity of the translation profession. Hence, MT was adopted as an aid to satisfy translation needs.

The immense possibilities offered by computers in rendering the translator's work further efficient and more effective are at variance and far-reaching. The challenge at this level is far more complex for morphologically rich languages such as Arabic, which is a morphologically complex language with a large set of morphological features. These features are recognized using both concatenative (affixes and stems) and templatic (root and patterns) morphology.

Boosted available literature on machine translation (MT) has spanned several researches and approaches. It is to be noted that Machine translation (MT) seems to be the combination of information technology and applied linguistics when translating human languages such as English and Arabic. Under the umbrella of globalization and the expanding trade, plea for translation appears to grow more and more. Computer technology has been applied to technical translation in order to improve one or both of the following factors:
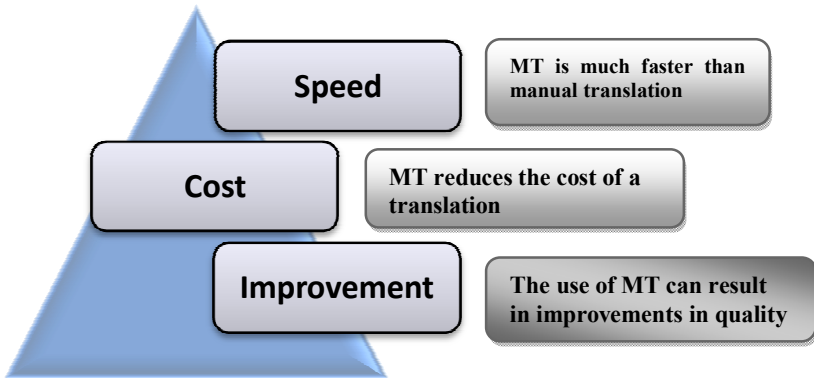
That is to say, applying machines (computers) on translation may in all probabilities help translation to be much faster than manual translation. Computer aids in translation can reduce the cost of a translation. In addition, the use of MT can result in improvements in quality, particularly in the use of consistent terminology within a text or for a particular kind of client.

Machine translation seems to be matter of serious speculation long before the existence of computers, since a great number of researchers believe that it is supposed to gained insights from studying machine translation even when it is applied to very pedestrian variety of texts, producing in this way results of a quite pedestrian kind. This may be explained by taking the fact that translation, a *sine qua non* for the comprehension of a text by a reader unfamiliar with the language in which it was first written, seem to englobe a task that obviously exercises every kind of linguistic ability, except those involved with the production and perception of speech, without requiring any other contact with the world.

Present-day translators are major beneficiaries of the technological revolution in office practice, hence, it has been estimated that the demand for translations is growing at an extraordinary rate well beyond the present or predictable capacity of the translation profession. Therefore, the appliance of computers to this task was first proposed over about forty years ago.

Just as a number of authors have stressed the positive side of various AMT systems and approaches, it seems to clearly arise considerable interest beyond the ranks of those directly concerned with developing systems or operating them.

However, it is crucial to consider the difficulty and the challenging task of translating from English to Arabic and from Arabic to English, this is may be explained by the fact that Arabic and English are distant languages from two unrelated families, hence, machine translation is, in any way, bound to face a number of intricacies in producing meaningful coherent translations between these languages, i.e., the levels of readability, in formativeness, and even grammaticality of those good translations are very high.

The two modes of translation have some common problems as well as mode-specific problems. It is trying time to diagnose and analyze these problems which may, in all probabilities assist in having lexically and grammatically well-formed translations, and in this way, the meaning communicated by the system output seems to be optimized.

Our utmost task, at this level, is in evaluating the output of MT, i.e., the transferred meaning. Therefore, semantics is a

very important facet in translation as a theory and application. The foremost difficulty in the Arabic-into-English mode is the **non-vocalization** of Arabic words, which leads to the wrong choice of TL words. Having no diacritics renders many words homographs, and thus it is difficult to determine which meaning to choose.

For example, الحر can mean *the-heat* or *the-free*, مهمة can be *task* or the feminine form of *important*, بعد can mean *to promise* or *to prepare*. Additionally, since Arabic words have no diacritics to determine their meaning, AMT systems displays confusion between a verb and a noun of the same root, or between the verb in the active and in the passive, e.g. قَتل and قتل .

Another significant problem is related to the inadequate lexicon, multiple meanings, connotation, and **collocation**. This may be explained by the existence of non-updated lexicon, i.e., the AMT system seems to ignore some basic words, this later causes mistranslation of words. This is may be the case of a number of Arabic names, which have meanings and thus would puzzle the AMT system. These can be solved by including them in the lexicon to have them transliterated instead of being translated.

Further, to put it in a nut shell, the following table may summarise one example of translation problems:

SL expresstion	MT Translation	Idiomatic Translation
شريط وثائقي	Documentary tape	Documentary
جولة سياحية	Toured tourist Touristic tour	Tour
طبيب نفساني	psychological doctor doctor the breaths	Psychiatrist

Table 1: Arabic Expressions, their MT Translations, and
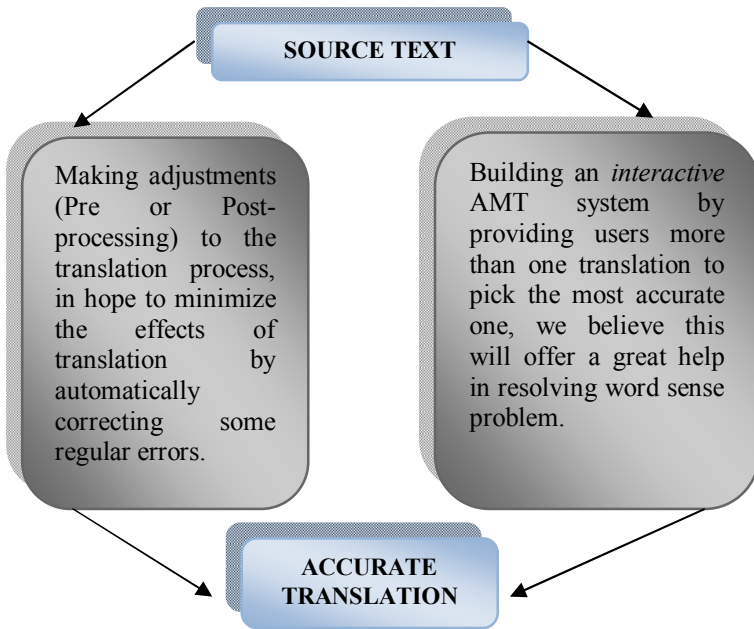Suggested Idiomatic Translations

Moreover, another noteworthy problem that needs considerable attention is words with multiple meanings. Since Arabic words may hold two or more English meanings, it is up to the system to choose, e.g. the word بطولة may mean *bravery, championship* or *heroism*. Hence, any choice may undermine the reliability of translation and the communication of the message. The proposed solution is the establishment of **lexical surroundings** in the TL which can help in excluding other translations, and the system will, automatically choose *championship* for **بطولة.**

From another stand point, there seems to be crucial to consider the fact that the real test of machine translation on Arabic is whether or not it is effective in large-scale operations, taking into account cultural encounters while translating. Through MT the linguistic marriage between the two linguistically and culturally incongruous languages can only lead to the deformation of the linguistic and rhetorical façade of the target text and will create structural damage to its

architectural beauty. On the other hand, AMT, by automatically generating a product where nothing existed before, has created a new mode of work in which a less than ideal translation can be provided for the consumer who merely needs to assimilate information and is not concerned with disseminating it.

Therefore, machine translation is not, as some believe, solved, nor is it impossible, as others still claim. It is a lively challenging and important technology, whose importance lies on a *multi-lingual* and *information-driven* world that hopefully will increase intellectually and commercially. However, a set of criticism was raised, mostly stem from the fact that the current translation quality of AMT seems to be qualifies as poor. In addition to this, AMT systems appears to be expensive to develop and their application degrades the retrieval efficiency due to the cost of the linguistic analysis, i.e., to get a so-called reliable translation, a multiple- AMT needs to incorporate a wider range of translations to choose from. To put it differently, correct or accurate translation is more likely to appear in multiple AMT systems rather than in a single AMT system.

Several other translation errors appear during translations which are of the type: wrong transliteration, wrong word sense, wrong word order and wrong pronoun. A set of remedial suggestions to overcome such problems are put as follows:

SOURCE TEXT

Making adjustments (Pre or Post-processing) to the translation process, in hope to minimize the effects of translation by automatically correcting some regular errors.

Building an *interactive* AMT system by providing users more than one translation to pick the most accurate one, we believe this will offer a great help in resolving word sense problem.

ACCURATE TRANSLATION

Conclusion

It seems to be decisive first to state certain absolute limits of transability when using an AMT system.   In total translation, translation equivalence depends on the interchangeability of the relationship of SL and TL texts to (at least some of) the same relevant features of situation-substance.

It is worth believing that working with Arabic either as a source or target language in Machine Translation (MT) a challenging inquiry for two main reasons. First, it seems quite

puzzling what the accurate representation is for Arabic words given a specific MT approach or system. Secondly, there are many MT-relevant resources for Arabic morphology, lexicography and syntax (e.g., morphological analyzers, dictionaries and treebanks) that adopt various representations that are not necessarily compatible with each other.

All, in all, one should bear in mind that natural-language translation should recognize the importance of the 'lower level' computerized translation tools. Unlike MT systems, which can only be gainfully employed under certain conditions, computerized tools offer tangible benefits in almost every area of written translation.

References

ALSALMAN, S. 2004. *The Effectiveness of Machine Translation.* In *IJAIS*, International Journal of Arabic-English Studies, 5, 145-160.

CARBONELL, J.G. & TOMITA, M. 1987 '*Knowledge Based Machine Translation, the CMU Approach'*, in S.Nirenburg (1987) *Machine Translation: Theoretical and Methodological Issues,* Cambridge: Cambridge University Press, 68–9.

ELLISTON, J.S.G. 1979 '*Computer-Aided Translation: A Business Viewpoint'*, in B.M. Snell (ed.) *Translating and the Computer,* Amsterdam: North- Holland, 149–58.

HUTTCHINS, J. & SOMERS, H. 1992 "*An Introduction to Machine Translation",* London: Academic Press.

NEWTON, J. , 1987, '*An insider's view of machine translation',* in *The Linguist,* vol. 26, no. 1:21–3.

NEWTON, J., 1992. "*Computers in Translation : A Practical Appraisal" Routledge*, London EC4P 4EE.

Google On-line Translation Page

<http://www.google.ae/language_tools?hl=en>
Systran On-line Translation Website. <www.systranet.com> .
WILKS, Y. 2009, *Machine Translation: Its Scope and Limits*,
Library of Congress.