The Middle East International Journal
for Social Sciences (MEIJSS)
e-ISSN: 2682-8766
Vol 6, No 1 Jun. (2024):1-8

ARABIC MACHINE TRANSLATION: CHALLENGES AND PERSPECTIVES

Ayad Mohammed Saleh

General Directorate of Education in Nineveh

Abstract

The current research deals with the challenges that we face in machine translation. Where it has been tackled to the beginnings of applications in this field, the subject of translation in general and what it is, and how it develops in some countries, because the structure and rules of the Arabic language are different from the rules and structures of the English language, and some problems in the structures when translated into Arabic, in terms of verbs and their pronounces, as some additions affect translation. For example, when analyzing data, errors occur in the entry of information because of the variety of forms and meanings of words in Arabic, the differences between pronouns and pronouns, and the differences between pronouns and pronouns, which confuse these pronouns and make it difficult to translate with the required accuracy, and the structure of sentences varies from Arabic to English in terms of words, verb, subject, and object.

In Arabic, on one hand, there are so-called noun phrases that do not contain a verb; such phrases may indicate a general time that is difficult to interpret in English, so there is a need to address such matters. On the other hand , however, there are differences between English and Arabic: in Arabic, the script is right to left , while in English it is the reverse, In addition to that human translation is not available all the time, and that the human element may become bored, tired and costly, and to overc ome all these difficulties we need a machine translation system, so it was important to develop this system. For all these reasons, the translation system needs to be developed in many ways. A team of different technical, scientific, economic, and cognitive disciplines could be organized to capture the most important vocabulary that could be needed.

Keywords: Machine Translation, Problems, Multiple Meaning, collocation.

1.Introduction

There are over 6,800 languages in the world today, each of these languages is most effective and crucial way of communicating, expressing feeling, transferring knowledge and imposing power over others. Therefore, there was

and is still very important to negotiate or interact by using these languages efficiently and clearly in order not to lose their above functions. And as an X language may be used over or across its home i.e. among non-native speaker, the need for translation (humane or non-humane) become very necessary.

In spite of trying to create most fluent and accurate translation by human being, still many drawbacks or problems with a human translator appear like shortage, getting tired and bored, expensive or charging fee, not available everywhere, Translators can die and inconsistent i.e. different translation in different times. So, to overcome these limitations of human translators, we need machine translation systems. And since every "machine" has in one way or in another some of drawbacks or challenges.

Aim of Study

The goal of the present study is to highlight the most important challenges and weaknesses in machine translation, and to try to develop some appropriate proposals to solve these problems in order to serve the most accurate and appropriate translation process.

2. History of Machine of Translation

The use of machine to translate text from one language to another has long been a dream. However, one can trace back the origin and history of rising what is now known as Machine Translation. According to Dupont, Quinn (2018), the origin of machine translation can be traced back to 9th century by a work for an Arab cryptographer Al-Kindi who invented systems for language translation involving cryptanalysis, frequency analysis and probability and statistics that are used now in current machine translation. Later on, the notion (machine translation) appeared in 17th century by René Descartes (1629), when he demonstrated of "universal language" of same ideas with different tongues sharing one symbol. The first machine translation was invented by Georges Artsrouni for his bilingual dictionary made of paper tape in the mid of 1930s. After that, Peter Troyanskii submitted a system of a bilingual dictionary and method for dealing with grammatical roles between languages, based on the grammatical system of Esperanto. The field of "machine translation" began in Warren Weaver's Memorandum on Translation (1949). In 1955, some researches appeared in Japan and Russia and in 1956, there was first MT conference in London. Association for Machine Translation and Computational Linguistics was formed in the U.S. in 1962 and National Academy of Sciences formed the Automatic Language Processing Advisory Committee (ALPAC) to study MT (1964) and researchers joined the field. (Täuschel:2008), (Arnold et al:1996) and (Hutchins:1978)

3. What is the Machine Translation?

Machine translation is a computer application that automatically converts content from one language (source) to another (target). Arnold et al. (1996) define machine translation as "a subfield of computational linguistics that investigates the use of computer programs to translate text or speech from one natural language to another." He adds that MT is a field that draws on philosophies and procedures of linguistic knowledge, translation theory and computer science. Tawchel (2008) argues that machine translation is "a specialized software system developed to translate from one human language to another." In addition, he points out that MTs are not machines as he thought, but programs that run on machines known as computers. It is a subfield from artificial intelligence to a large field of computer science

4. The State of Arabic Machine Translation

When comparing Arabic machine translation to its American, European, and Asian counterparts, it seems that it is still in its infancy. However, this area has recently seen momentum in the production of commercial software as follows:

In Virginia, USA, Apptek has begun developing an AMT system called "Transphere", an English-to-Arabic machine translation system that can translate general and specialized domain texts. The I program currently runs under UNIX and Windows, in Paris, France, CIMOS designed a system called Nakheel for translation into various languages. Currently, it is used to translate from English to Arabic and French and from French to both Arabic and English. It runs under MS Windows, in the United Kingdom, the institution invented the "Arabic Translator", which is referred to as the first professional English-Arabic computer-based translation system. The same company designs Al-Wafa, which is a miniature version of the Arab Dieter. In Riyadh, Saudi Arabia, scientific is a system translation between English and Arabic. (Zantout and Gisum: 1999)

5. Challenges for Arabic Machine Translation

MT is a challenging task since it contains various statistical models, machine learning models and other techniques which helps to improve MTs. However, generally speaking, there are two main challenges encounter AMT. Linguistic and technical aspects. In which, the linguistic challenges are related to the nature of Arabic language like its richness in morphology. While the technical challenges are related to the computation process of MT. However, Arabic is a very rich language and belongs to Semitic languages, therefore, in translating between Arabic and other languages using MT, many problems arise that make it difficult to understand or depending upon it (MT) whether in work or everyday situation. It seems important to mention some of characteristics that make Arabic distinctive from other languages before digging deeply into challenges of AMT. For Habash (2007:263) “Arabic has a very rich morphology characterized by a combination of templatic and affixational morphemes, complex morphological rules, and a rich feature system”. However, the following are the most prominent challenges that face Arabic machine translation:

5.1. Problems of Lexis

The main problem in translation via MT is the non-vocalization of Arabic words, inadequate lexicon, multiple meanings, connotation, and collocation which lead to the wrong choice of target language words.

A. **Non-vocalization** is the cause of poor translation in the target language. With no diacritics in homogeneous words, it would be difficult to determine their meaning in the target language. For example, systems mix a verb and a noun from the same root, or a verb with an active and passive verb, as in "قُتل" and "قتل". (Izwini, 2006) (Zakraoui et al, 2021:161452) (Alkhatib and Shaalan, 2018:153)

B. Inadequate Lexicon :Lack of definitive, future prefix,

A preposition or pronoun associated with basic forms of Arabic words or the lack of proper names for people, cities, countries, organizations, nationalities and even some simple phrases such as "too much" leads to a mistranslation or translation. To improve translation, removing the attached item or adding it to other flag names helps the system recognize and translate it correctly. (Izwiny, 2006) (Guissum and Zantout, 2005:305)

- e.g: Skin discolouration and pigmentation is just one of the effects of too much exposure to the harsh Gulf climate.

- جلد [ديسكلوورأيشن] وصبغ فقط واحدة من التأثيرات من [توو موش] انكشاف إلى القاسية خليج مناخ.

Similarly, target language terms and idioms must be included in the lexicon with equivalence in the output as one unit; Otherwise, translation can be misleading. "Bird flu" and "Come in" and "Bird flu" in!" Translated literally to إنفلونزا عصفور and في تعال . Here punctuation may guide the system in such cases.

C. Multiple Meaning, Connotation, and Collocation

Another problem is words with multiple meanings. Many Arabic words have overlapping meanings compared to English or other languages. In this case, the systems need to specify which one to choose for example, the Arabic word "قمة" equals the English words "climber, peak, summit, summit ". Choosing any of these meanings may harm the intended message. To solve this problem, account should be taken of the lexical medium and translations in the source language as well as the lexical elements of the produced target language and its frequency (Gissom and Zantut, 2005.305) (Zalawi et al. 2021:161451) (Khatib and Shalin, 2018:152)

Connotation is another problem that needs consideration. In some cases, the target language words have connotations which do not match with those of the source languages. For example, a probable input "من يخلو" in Arabic has a negative connotation when it comes with "المصداقية" (credibility). However, the output "*free of*" gives it a positive connotation as in:

- السلطة الفلسطينية: خطاب شارون يخلو من المصداقية

(the-authority the-Palestinian: speech Sharon lack from the-credibility)

The Palestinian Authority: Sharon's speech is free of the credibility. (Izwaini,2006)

Collocation is another problem that is available in AMT. Some lexical items must be translated in relation to those words that co-occur with them frequently. For instance, "مركز تجاري" is translated literally into (commercial center) which has completely different meaning, whereas its real meaning in relation to collocation in the target language is (shopping center or shopping mall). (Izwaini, 2006)

5.2. Problems of Grammar and Syntax

Grammar and syntax are problems face AMT especially when treating words individually as blocks organized one after another. Doing so, lead to awkward and meaningless translations. For example,

- سنسافر الى الفلبين غدا(we-will-travel to the Philippines tomorrow)

Is translated into (Sensavr to the Philippines tomorrow) in relation to the syntactic order of the target language. (Izwaini,2006)

A. **Word Order**. Arabic sentences are either nominal (SVO) or verbal (VSO) according to the syntactic order of the sentences. The former order is translated more easily than the latter as the former matches the English word order unlike the latter. So, whenever there are VSO sentence in the source language, a problem can arise.

- قلل الرئيس الباكستاني برويز مشرف من شأن مصافحته رئيس الوزراء الاسرائيلي اريل شارون خلال وجودهما في نيويورك لحضور قمة الأمم المتحدة.

- (underestimate Pakistani president Pervez Musharraf from matter his-handshaking Israeli prime minister Ariel Sharon during their-stay in New-York to-attend summit the-united the-nations)

Reduced Pakistani President Pervez Musharraf will Msafhath Israeli Prime Minister Ariel Sharon during their stay in New York to attend the United Nations summit. (Izwaini, 2006) (Zakraoui et al, 2021:161452) (Alkhatib and Shaalan,2018:142)

B. Gender and Reference

In Arabic reference system, there are only two gender pronouns: masculine and feminine. This leads to confusion and inconsistence of AMT by translating some masculine and feminine pronouns in the singular and plural that refer to inanimate into *he*/*him*/*his* and *she*/*her* which should be translated into *it* or *its*. Therefore, AMT system needs to establish a relation between the pronoun and the noun to which it refers, and identify whether it is animate or not, in order to determine target language gender. (Izwaini, 2006) (Guessoum and Zantout, 2005:315) (Zakraoui et al, 2021:161454)

- أسعار النفط قرب مستوياتها القياسية

(prices the-oil near their-levels the-standard)

Prices of the oil near even her standard.

In translating form English to Arabic, gender disagreement between the subject, the verb, and the reference arise. Disagreement in gender and number also occurs even though there is an indication of the gender as in:

- Women are being groomed to take their place in a more diverse professional elite. But many of these women say that is not what they want.

- المراه يعده اخذ مكانه في آآثر تنوعا نخبه مهنيه .لكن الكثير من ءلاء النسوه لا نقول ما نريد.(Izwaini,2006)

C. Wrong Analysis of Input

Wrong analysis of input may be of various formats of Arabic words (like attached prepositions and pronouns or nonidentification of diacritics) which have a main role in differentiating meanings. For instance, Systran system seems to face problem in processing numbers and the nouns they modify. (Izwaini,2006) (Zakraoui et al, 2021:161454)

-200مليار درهم الاستثمارات العقارية المتوقعة خلال السنوات الخمس المقبلة

- (200 billion Dirham the-investments the-real-estate the-expected during the-years the-five the-next)
- 200 one billion Dirham the land investments expected during the years the five next.

The same happens with English translation to Arabic:

- Palestinian chief demands end to anarchy after Israeli pullout.
- نهاية طلبات الرّئيس الفلسطينيّة للفوضويّة بعد إسرائيليّ يتحرّك

In the above case, the system analyzed "demands" as a noun, and "Palestinian" as an adjective modifying the head noun as well as the noun "pullout" is recognized as a verb, leaving the modifier 'Israeli' with no noun to modify which lead to a completely wrong and incoherent translation lacking of readability and informativeness.

D. Tense and Aspect

Nominal Arabic sentences do not have verb that encode or carry the tense sense in the present tense. Therefore, the translation needs to such a verb to be inserted in the right position. The process of insertion must be obligatory in the system. Beside, past tense in Arabic is used just to refer to **probable case of condition**. (Izwaini,2006)

-إذا ذهبت إلى أثينا فلا تطلب فنجان قهوة تركي، فقد يقذف بك إلى خارج المقهى.

- (If you-go to Athens then-not ask cup coffee Turkish, may throw you to outside the-cafe)
- If I went to Athens not request a cup of coffee Turki, you jump out of the cafe.

Moreover, English and Arabic systems of tense and aspect do not match each other. Therefore, Sakhr system will translate the present perfect into the closest structure of the past tense plus the particle "قد" which in associating with the past, denotes completion of an action. Another problem is mood since imperative is not recognized in some AMT. So, "Pass the salt, please." translate into "رجاء الملح، مررت".

E. Prepositions

Translation of prepositions is one of the problematic issues in AMT. The prepositions need to be rendered to systems according to the output rather than the input or a text can be rendered with an unsuitable target language preposition which is either awkward or meaningless. Prepositions have specific verbs and nouns that always associate with them. For instance, "اختتم" often associate with "ب" which translated wrongly in AMT into is translated into **concluded by** or **in**, while it should be with. (Izwaini,2006)

F. The Definite Article

It is not necessary to translate the Arabic definite article into "*the*". Several defined words must be translated as indefinite. In addition, Generic names which are definite in Arabic, however, in English they have to be indefinite. Furthermore, titles associated with names do not take the definite article in the target language but they do so in the source language as in "الشيخ" which must be translated into Sheikh, الملك (the-king) into King, الرئيس (the-President) into President. Additionally, in Arabic, names of institution, place, countries and cities can have the article, but they needn't when transferred into English.

One of important rule is that verbs do not have definite or indefinite articles. However, AMT can violate this rule since it translates the present participle as a verb.

- the coming Saudi oil shock and the world economy
- ال يأتي سعوديّة زيت صدمة والاقتصاد عالميّ

G. Coordinators and Conjunctions

In Arabic, there is extensive use of conjunctions and coordinators. In some cases, they needn't be translated since their translation can affect negatively on cohesion of English text. Therefore, there should be a smart filter to deal with them properly such as deleting them at the beginning of Furthermore, they cause problems when they combine to other particles like"فقد" (*hence…has/have/might*) in which they are conceived by the Systran system as a verb and translate them into (*to lose*). In such case, the system should feed with information to identify whether this is a single word and a verb meaning (to lose) when it followed by a noun. But when it followed by a verb, then it is a conjunction attached to a particle.

H. Wh-words

Such words can be translated either as question words or relative pronouns. However, syntax and punctuation can guide the system to determine the correct translation.

- Who are you?
- الذي يكون أنت ؟

Discussion and Results

Arabic was considered one of the most difficult languages to process written and spoken because of its morphological, grammatical, and phonological characteristics, although aspects of Arabic were investigated in the early days of machine translation...

These aspects are specific to Arabic, which defines its linguistic characteristics and its use in a local processing approach to the language's internal system. It should focus on semantic and morphological aspects, especially the tricolor system, which differs greatly from other languages in its letters, shape, and form.

As for the general aspects that are complementary to the language aspect, opportunities may be for applications of the grammatical aspects already used on other languages, such as French and English, with or without adaptation, the latter aspects being based on the grammatical aspects of the language system in general.

It can be said that there is a need for a well-tested standard determination and evaluation methodology to be generally recognized by translators to measure the clarity and quality of its performance. Besides, efforts should be directed to surveying and classifying research work from the perspective of machine translation to fill the gap in each model.

In fact, there is an urgent need for teamwork in the Arab world in the field of linguistics, translation, engineering and computer science to improve machine translation systems, as is the need to find solutions to other problems such as the clarification of abbreviations and other expressions, definition of meaning and use, which enables the provision of contextual information and the provision of frequency for keywords. Also, software can be made to use in group linguistics that helps in identifying expressions and terms in the target language of parallel texts. This step can greatly help machine translation studies, as Whenever you enter a text for translation, the machine actually stores a large amount of readable text, and it can recognize that text in context and subject, and translate it accordingly.

The few machine translation systems available from and into Arabic, especially the Arabic-English pair, are actually improved versions of electronic dictionaries, and it should be noted that other applications available by popular companies have limited coverage of Arabic phenomena.

It is also notable that most applications of machine translation are made in the West as a purely profitable material in an environment other than the original Arabic language. There are no systems in the Arab world that process machine translation, and there is no hard work to develop this field.

Conclusion

The study concludes that there are some challenges in AMT which need for taking into account by developing the systems capable of dealing with long sentences, certain specific usages of vocabulary, word order and feeding the machine with morphological information as well as diacritics as all these were among main challenges or problems that face Arabic machine translation in the light of perspectives which are given in the study.

References:

- Alkhatib, M. and Shaalan, K., (2018), The Key Challenges for Arabic Machine Translation.
- N. Habash, Introduction to Arabic Natural Language Processing. San Rafael, CA, USA: Morgan & Claypool, 2010.
- N. Habash and F. Sadat, Arabic preprocessing schemes for statistical machine translation, in Proc. Hum. Lang. Technol. Conf. NAACL, New York, NY, USA, 2006.
- A. Almahairi, K. Cho, N. Habash, and A. Courville, First result on Arabic neural machine translation, 2016.
- Izwaini, S. (2006). Problems of Arabic machine translation: Evaluation of three systems, In Proceedings of the International Conference on the Challenge of Arabic for NLP/MT, London: The British Computer Society (BSC), 118-148.
- Elsherif, H. and Soomro, T., (2017) Perspectives of Arabic machine translation, Journal of Engineering Science and Technology Vol. 12, No. 9 (2017) 2315 – 2332 © School of Engineering, Taylor's University.

- Zantout, R.; and Guessoum, A. (2015). Obstacles facing Arabic machine translation: building a neural network-based transfer module, In Papers in Translation Studies, Cambridge, UK: Cambridge Scholars Publishing, 229-253.
- Hutchins, J. (1986). Problems, methods, and strategies, In Machine Translation: past, present, future, Chichester: Ellis Horwood.
- Zakraoui, J., Saleh, M., Al-maadeed, S., and Alja'am, J., (2021), Computer Science and Engineering Department, Qatar University, Doha, Qatar under Grant 10-0205-170346, and in part by Qatar National Library for open access publication.