A Study of Computational Linguistics and Machine Translation

Hussein Shamsabadi[1], Asghar Moulavi Nafchi[2], Fereshteh Afzali[1]

[1]The Department of Arabic Language and Literature, Hakim Sabzevari University, Sabzevar, Iran
[2]The Department of English Language and Literature, Hakim Sabzevari University, Sabzevar, Iran

Email address:
drshamsabadi@gmail.com (H. Shamsabadi), a.moulavi.n@gmail.com (A. Moulavi Nafchi), fereshtehafzali1853@gmail.com (F. Afzali)

Abstract: Today the world is experiencing advanced technology on a full scale. Man is thus compelled to make use of this technology to develop machine translation as one of the means of transferring the cultural and scientific achievements of the various fields and disciplines from other languages to Arabic and vice versa. It is quite mandatory to pave the way for an efficient ground of cooperation among Arab researches, who should aim at obtaining advanced technology in the field of translation, especially during this booming period. Likewise, conveying world-wide scientific achievements in Arabic plays an essential role in establishing the slight but still ongoing endeavors that aim at the Arabization of various scientific terminologies. The present article attempts to elucidate the concept of machine translation, its objectives, its categories and types, its value in the current age, its position among other modern sciences and fields of research, and the attempts made by the authorities in the field, who look up to the growth and development of machine translation, all with an emphasis on Arabic.

Keywords: Arabic, Computer, Information, Machine Translation, Linguistics

1. Introduction

The progress and potential of machine translation have been debated much through its history. Since the 1950s, a number of scholars have questioned the possibility of achieving fully automatic machine translation of high quality (Bar-Hillel, 1964). Some critics claim that there are in-principle obstacles to automatizing the translation process. (Madsen, 2010).

The area of machine translation is still relatively unidentified throughout the Arab world, it is even unknown to the linguists, the informing specialists, and the members of the translation community. It is amply clear that this unfamiliarity is the outcome of ignoring the advanced technologies that have been used in the area, and is sometimes the result of refusing the Arab world's pressing need for using it in cultural and scientific exchange, considering all its branches and trends, from one language to another. At times, this unawareness and unfamiliarity reaches a place that causes fear and utter refusal, for man's nature is the adversary to unknown things. It is quite mandatory to pave the way for an efficient ground of cooperation among Arab researchers, who should aim at obtaining advanced technology in the field of translation, especially during this

booming period. Likewise conveying world-wide scientific achievements in Arabic plays an essential role in establishing the slight but still ongoing endeavors that aim at the Arabization of various scientific terminologies. In some Arab countries, a couple of researches on the issue have been conducted, most of them referring only to the Arabic language from computer point of view, but these researches still cannot help us to design a system of machine translation from Arabic to other languages and vice versa. A great number of machine translation programs at the market are mostly based on direct machine translation, one of the first endeavored designing systems in the 1970's, and a few of them are based on intercession method. Also, we can see that none of them follows transformation or replacement methods. At any rate, the reasons why Arabic falls short of these translation methods and systems can be listed as follows:

First, individual, partial, and personal endeavors will never lead to a scientific end, unless they unite all the capacities, complete all the processes and works, and also use the experiences of the other languages. In addition, it should not be restricted to some partial Arabic programs and tools like electronic dictionaries, orthographic programs, and grammatical, inflectional, and marking programs, for machine translation requires such things as semantic,

inflectional, and syntactic examining processes, meaning transferring processes, and substructures of Arabic sentences corresponding to their equivalents in other languages. Second, there is a considerable lack of funding and serious investing both from the government and from the private sector to support such researches or developmental plans. Additionally, the absence of correct assessment of the amount of endeavor and time required for doing such projects places detrimental effects on the experts and the researchers of the area and leaves them discouraged and unmotivated. This, as a result, will also stop them from producing machine translation systems, especially for Arabic, and then connecting them to the translation systems of other languages. One of the chief reasons why the authorities are unaware of the precious role of machine translation in development and progress is sometimes not knowing, ignoring, or fearing the technology. Third, we can see that the developments in this area, namely, machine translation systems has been revolutionized all being somehow related to the Arabic language as a source or a target language, or either way, is all Arab researchers take into serious consideration. Non-Arabs will naturally act differently as they lack similar motivation, for they do look at the issue from their own different cultural interests, the same as the Arabs themselves do. Usually, technologically well-advanced foreign countries take Arabic as a target language, because they rarely need it as a source language. So, Arab experts should take the initiative themselves, and ask for such foreign help and cooperation, which they are not practically needless of.

2. Discussion

2.1. Machine Translation

Machine translation is only one part of a great whole including areas of applied and theoretical researches that treat computer-based natural language processes, today known as computational linguistics that is a branch of artificial intelligence knowledge. This wide scientific area, computational linguistics, deals with the main mechanisms on which human thoughts and languages are based. By way of describing and reconstructing its mathematical structures, using formal structures and artificial languages and patterning them, computational linguistics attempts to discover those mechanisms and then it tries to use them in computational programs frameworks, in the same way. (Al-Humaidan, 1994)

However, on artificial intelligence, it should be mentioned that the consisting base and core of all automatic control systems and export systems are equal, and we hear of them a lot these days. This is done by such methods as computational describing, structuring, examining, and stimulating the process. Computational manufacture of automobiles and their equipment and facilities production and computational ships and planes direction can be named as the examples of automatic control systems. Instances of export systems, that are more based on the knowledge base of specific areas of knowledge, are like export systems of medicine, agriculture, and language.

These researches that are related to machine translation deal with theoretical trends and practical techniques, and use the partial steps serially through the whole translation process. This whole translation process consists of inflectional, syntactic, and semantic analysis of the source language, and also inflectional, syntactic, and semantic reconstruction of the same model in the target language. Finally, we can say that these researches can provide helpful solutions and viewpoints to overcome the obstacles existing in the way of machine translation and its partial steps. In addition, machine translation can be used on a wide scale, because it can be a tool for examining advanced theories and technologies and evaluating them in a more restricted fashion, but it should be mentioned that machine translation is restricted to the areas of computational languages and artificial intelligence.

Machine translation basic and pivotal designing steps are equally divided between information and electronics experts, linguists, and translators.

2.2. Computer and Information Sciences

The examination and study of common human languages is not the first purpose of information sciences as related to machine translation. Clearly, it is the purpose of other fields of study, which without any doubt, is the beginning of a linguist's work. Linguists are responsible for controlling, describing, and analyzing our languages, and then information experts, by employing algorithms and data structures that shows how we use language, enter such a description and analysis to the internal world of computers. Linguists study and control the language to find out its phenomena, considering all its details. They do this to become eventually capable of describing the language by using formal or rather artificial languages that match the computational programs and are favorable to the information sciences. By doing this, they want to explain the inflectional, syntactic, and semantic rules and also present an appropriate and useful description of the words and sentences substructures. Then, the design of these programs should try to improve the speed, capacity, and function of the programs. They should also enhance the way users connect to the machine translation systems. (Al-Humaidan, & Usef, 1995)

2.3. Electronics

Electronics experts and engineers' work starts when the issue is related to words and their translation. The whole chain of human language relates to two instances. The first one is when the speaker gives utterance to words and the second is when these words reach the listener's ears and he understands them. The processes of written text machine translation form the main and basic links of this chain, because it is absolutely necessary to face them when speech machine translation is being conducted. It is mandatory to analyze the translation and substitution of speech from vocal

signs to written text in the source language for it to be usable by a machine. At this step, the translated texts will be processed for creating and compounding corresponding and fitting vocal signs in the source language. Performing all these productive, substitutional, and analytical processes between speaking and writing are on the responsibilities of electronics experts.

Generally, the scientific area, dealing with this issue, is known as signal processing or form recognition, but because the mentioned signs or shapes may include pictures, writings, or both of them, it is specifically called speech processing or speech recognition. For the same reason, optical character recognition is connected to this area and follows the same process. It is necessary to say that electronics experts are not the only people who deal with this issue, but also linguists who are experts in phonetics and informing experts play an important role. Phoneticians do the controlling, analyzing, and describing of speech rules to make informing experts able to program the components of those speech patterns. (Al-Humaidan, 2009)

2.4. Linguistics

The value of linguists' role and among them the role of phoneticians, lexicologists, syntax experts, and semanticists emanates from an objective reality. This reality is that in computational translation the huge obstacles are not computational, but linguistic. These obstacles, hindering efficient machine translation systems from improving and developing are problems like lexical, inflectional, and semantic dualities; syntactic complexities, lexical pluralities in languages, figurative structures opposed to grammatical rules, and the like. We can briefly say that the extraction of the sentences and texts meanings by analyzing written signs is the problem. It means that understanding the target language letters, their religious, scientific, social, and cultural meanings, and then producing sentences and texts that are equal to the signs of another group with the same letters of the source language, having an equivalent meaning to the all mentioned aspects without changing the target language meanings, is the real obstacle.

Machine translation should pay full attention to linguistics researches improvement, and should definitely count on it, especially those branches and trends that have reached a good level of machine structures and natural language linguistics rules, including inflection, syntax, and semantics. But it is necessary to know that machine translation is not able to exercise and conduct linguistic theories simultaneously and directly. Based on this, linguists can only explain the construction and understanding of the language machines and tools, and focus on its important and fundamental features and advantages. They will not try to explain and describe everything because they know that there are some expectations in all languages and people will learn them through life experiences, but foreigners who are learning a language find these expectations, like male and female plural in Arabic language, difficult, because no rule explains them. (Shamsabadi, 2003)

Machine translation systems should deal with modern and daily texts. It should examine all the aspects of vocabularies like complicated expressions, dictations, changing and improving new words, linguistic interpretations, and semantic and stylistic dualities in figurative languages like pun, irony, imagery, etc. All of these, at first glance, are the responsibility of applied linguistics experts, then that of translators', and finally that of computational programming experts. It is necessary to know that these are not related to implicit theoretical sciences and fields of studies.

2.5. Translation Knowledge

Documentaries, scientific, and technological texts translators, not the literary texts translators, are the first and main users of machine translation systems. So, these systems, in designing components and functions, should be in line with manual translation and the way the translators of such a function. The transition process from using pen and paper, the two basic elements of manual translation, to electronic text processing programs and other helpful programs like electronic dictionaries are not the only things that machine translation is concerned with, instead, beside these, the main purpose of machine translation is aimed at how it is possible to imitate and copy the mental process of translators while translating. This mental process includes two steps, the first one is to understand the main text and the second one is to translate and reconstruct the thought structures and to write them in the target language with complete accuracy and trustworthiness in transferring meaning. (Al-Humaidan, 1998)

On the first point, all are unanimous that the advent of computer and it services in different offices and daily works has changed the way we deal with the documents and data. So, the translators should use these tools, tools like typing machines, computational dictionaries, and dictation and grammar editing programs, because these tools will facilitate and hasten their writing, revising, translating, keeping the texts, etc. And finally, unlike the past, these will improve the effectiveness and efficiency of their works, and also will help them to meet their deadlines. Furthermore, experts who are making these programs more advanced should put all their capabilities in one place in order to help the translators in using all or part of them as they want.

As to the second point, understanding the original text and translating it in the target language in written form, we have to say that the internal mental activity and work along with its detailed steps leave some reflection on the translators' outer behavior and works, and also on the way they deal with and use these tools in various steps. Being familiar with translators' method or methods and the translation theories related to their works is a fundamental element in designing machine translation systems. This will make their practical systems capable of reflecting consistency and unity to these various practical theories to an acceptable extent.

Understanding the original text as step in a process, firstly begins with knowing the locative and temporal structure of that text, the scientific area, and the readers, together with

other aspects that are important factors in picking vocabularies, style and grammatical rules. In the deeper understanding of the source language text, the translator should refer to various monolingual dictionaries to become familiar with new expressions and vocabularies of the target language, or to be able to solve the complexities coming from duality in meaning, but in translating the structure and writing the text in the target language, translators should choose an appropriate writing style from understanding the factors that have influenced the source language writing correctly. They should also consider civilizational and cultural differences between the users of source and the target languages. In the deeper stage of transferring meanings from the source language to the target language, translators should refer to general and specialized bilingual dictionaries to look up scientific and technological terms, and they should also consult monolingual dictionaries in the target language to find correct equivalents for vocabularies. Sometimes, translators find it difficult to translate new technological terms in the source language that have no equivalent in the target language. And sometimes translating some cultural concepts in the source language is difficult for them, because there is no equivalent in the target language, or they are not acceptable to the target language culture.

2.6. Machine Translation or Manual Translation

Man couldn't take serious and effective steps in this area until the beginning of the second half of the twentieth century, along with the advent of electronic computers. After the advent of electronic computers, this dream gradually came true, and today it has become a tangible reality. Now, computational programs are capable of translating various texts from one language to other languages. It is good to say that the term national languages refers to the language that people use to communicate with each other, languages like Arabic, Farsi, English, Japanese, and other common languages. This natural language is opposed to the artificial languages and formal languages, languages that are used to show the computational programs languages. These two languages were used concurrent to the arrival of mechanical and then electronic computers. They have their own vocabularies and grammars. These artificial languages use human language inflectional, syntactic, and semantic rules in the description structure of the tools that have been used in natural language, when conducing a computational or mechanical processing of these languages. (Howlett, & Rota, 1980)

Now, it is difficult to say that this reality will be well-developed one day. That day there will be no machines or tools for translating, but by pressing one or a couple of buttons man will translate a text from one language with a great quality, full automatic, and even without human intervention. This is the ideal example that researches try achiever in the future. The difficulty of reaching this state lies behind the patterning of all aspects of human mind natural processes when using these languages and the inflectional, syntactic, and semantic analyses of these processes. Historical, civilizational, social, cultural, and even critical

aspects are effective in translation process. There are lots of translation cases that have been done by one person, which other people find them unacceptable, and even sometimes the same person finds his own work dissatisfying. This is because different people have different understandings and interpretations of one thing. Based on this, we can claim that it is impossible to produce a computational translation program that pleases everyone. So, at the first place, we should notice that machine is unable to comprehend the text and transfer it to another language in a way that the people who are familiar with translation process would accept it. Thus, we should not expect from a machine what we expect from a person. Anyway, most texts that machines are used to translate are fluent and plain scientific texts that have not been written in figurative language. When there are some figurative expressions in a scientific text that causes duality in meaning, and there arise scientific writing problems.

Anyhow, whatever has been done in this area is the development and improvement of the programs that can produce the raw translation of the text that are about stated topics and are related to the definite areas and are also directed with well-defined and pre-stated methods. Then, this text undergoes machine translation edition which a translator controls and revises to improve their quality and make them useful. This will help us to achieve the average of translation productions positive capabilities, improving the quality and quantity besides decreasing time and cost. The other type is machine translation supervision that has the translated texts without revision, and they will come in handy for selecting and supervising purposes when required. (Le Systeme de traduction automatique de l'Universite de Montreal, 1973)

In some cases a machine translation work can be produced with a better quality that needs less revision and edition by a system to control and supervise the language of the text that is delivered to the machine. Such a system is activated at Montreal University. It is used throughout Canada for translating weather news from English to French and vice versa. The percentage of revision, control, and edition has been decreased to less than five percent with this system. (Chndioux, 1989)

These days, such tangible processes are traditionally called machine translation that remain vague and unknown or have been misunderstood by common people. Because of the two following reasons public understanding of machine translation is vastly wrong (Hutchins, & Somers, 1992):

1. Some people do not accept that analyzing and processing languages is a difficult task. They believe that even children can learn lots of languages readily, and also they say that a person who knows one language beside his mother tongue can translate easily. They do not believe in translation knowledge, theories, schools, and ethics. If we take a detailed look at their beliefs, theories, and characteristics, we will learn that they don't have a correct understanding of manual translation difficulties, and also they have wrong conception of mechanical examining and processing of languages difficulties to make machine translation

systems. So, they don't know the real value of machine translation achievements and translation knowledge as they should do.

2. There is another group of people that considers no role for machine translation based on computers or machines. They think this kind of translation is unable to translate figurative language, verse, Holy texts like the Quran, and religious narratives. This group is unaware of the aid and services that machine translation brings to translators, although it is not perfect and without defects. However, given the pivotal role machine translation can perform in bridging civilizations and cultures, these people don't understand the advantages of such translation in terms of facility and speed, only focusing on its temporary inadequacies. In addition, literary translation, especially from Arabic to Arabic, at least at its first steps, should be taken as scientific translation, not literary.

2.7. Using Machine Translation as a Teaching Tool

Although there have been concerns about machine translation's accuracy, Dr. Ana Nino of the University of Manchester has researched some of the advantages in utilizing machine translation in the classroom. One such pedagogical method is called using "MT as a Bad Model." MT as a Bad Model forces the language learner to identify inconsistencies or incorrect aspects of a translation; in turn, the individual will (hopefully) possess a better grasp of the language. Dr. Nino cites that this teaching tool was implemented in the late 1980s. At the end of various semesters, Dr. Nino was able to obtain survey results from students who had used MT as a Bad Model (as well as other models.) Overwhelmingly, students felt that they had observed improved comprehension, lexical retrieval, and increased confidence in their target language. (Nino, 2009)

3. Conclusion

By taking a look at this article, we can say that machine translation is not detached from scientific researches. Because it inspires and uses computational knowledge, AI, translation theories, and also thoughts, methods and technologies applied for the rapid development and improvement of efficient systems. So, machine translation is fundamentally related to practical researches, and it may take advantage of technologies and their models to use in the computational processing of natural languages in other fields.

References

[1] Al Humaidan, A. B. H. (1994). *Basis of the decision making process exercised by the translator*. College of Languages, Ein Shams University, Philology.

[2] Al Humaidan, A. B. H. (1998). *Change and development of machine translation systems* (8th Vol.). Saudi Arabia: College of Languages and Translation, Malik Saud University Publications.

[3] Al Humaidan, A. B. H. (2009). *An introduction to machine translation*. Translated by Hosein Shamsabadi and Fereshteh Afzali. Par Tehran Shahr Publications.

[4] Al Humaidan, A. B. H., & Usef, A. (1995). *Translation from theory to practice. Scientific Magazine*, no. 1, College of Languages and Translation, Malik Saud University Research Center.

[5] Bar-Hillel, Y. (1964). "A demonstration of the nonfeasibility of fully automatic high quality machine translation," in *Language and Information: Selected essays on their theory and application*. Jerusalem: Academic Press. 174-179.

[6] Chandioux. J. *Météo: 100 million words later*. In D. L. Hammond, editor, *American Translators Association Conference 1989: Coming of Age*, pages 449--453. Learned Information, Medford, NJ, 1989.

[7] Howlett, J., & Rota, G. C., Eds. (1980). *A history of computing in the twentieth century: a collection of essays*, PP. 551-561. New York: Academic Publications.

[8] Hutchins, W. J., & Somers, H. L. (1992). *An introduction to machine translation*. London: Academic Press.

[9] Le Systeme de traduction automatique de l'Universite de Montreal (The University of Montreal's System of Automated Translation. *Meta*, 18, 1-2, 277-289, Mar-Jun 73. Paper presented at the Second International Conference on Linguistics and Translation, October 4-7, 1972, Montreal, Canada.

[10] Madsen, M. (2010). The Limits of Machine Translation. Docs.google.com. Retrieved 2016-01-06.

[11] Nino, Ana. (2009). Machine Translation in Foreign Language Learning: Language Learners' and Tutors' Perceptions of Its Advantages and Disadvantages *ReCALL: the Journal of EUROCALL 21*(2), 241–258.

[12] Shamsabadi, H. (2003). *Translation theory and practical translation from Arabic to Farsi*. Tehran: Chaapar Publications.