An Analysis of the Shortcomings and problems of Pars Translation Machine

Raheleh Karimi-Feizabadi¹, Masihollah Nemati², Heidar-Ali Amini³

Faculty of Letters and Humanities, Ferdowsi University of Mashhad, IRAN.

¹rahelehkarimi16@yahoo.com, ² ma.nemati@stu.um.ac.ir, ³ aliamini1348@stu.um.ac.ir

ABSTRACT

In the latter half of the twentieth century a new branch of applied linguistics named computational linguistics, or Engineering linguistics emerged. Computational Linguistics is an area of study between computer science and linguistics which focuses on the computational aspects of human language faculty. In this paper, we attempt to provide some evidence to show the shortcomings of a translation machine, that is Pars Translation machine. Of course we will also review its strengths.

Keywords: Machine Translation, Pars Translation Machine, Translation Criticism, Computational Linguistics

INTRODUCTION

Computational Linguistics belongs to the cognitive sciences and overlaps with the field of *artificial intelligence (AI)*, a branch of *computer science* aiming at computational models of human cognition (Uszkoreit, 2000, p. 25).

Computational linguistics might be considered as a synonym of automatic processing of natural language, since the main task of computational linguistics is just the construction of computer programs to process words and texts in the natural language. Actually, this course is slightly "more linguistic than computational (Bolshakov and Gelbukh, 2004, p. 25). Machine Translation is one of the research areas that in Computational linguistics (Uszkoreit, 2000, p. 22).

Machine translation is one of the important achievements of artificial intelligence. This science makes use of best linguistics and computer science applied theories (Valipoor, 2007, p. 35).

Work in the field of machine translation began in the late 1940s and over the past 50 years, various approaches were discussed and tested (Ramanathan, 2008, p. 44).

Due to this frequent analyses, the machine translation phenomenon has progressed; But it is clear these is a long way left to fully achieve ideal translation. Because in translation just translating the structures and vocabulary of a language to another language is not considered, but one of the most important elements of translation - and perhaps the most important one - which can be said is only limited to human is creativity in translation; and So far the translator machines are not equipped with this kind of knowledge. Moreover, it is clear that any translator machine that comes to market and is used have some defects and shortcomings and one can't remove and compensate these weaknesses without the recognition and analysis of these shortcomings.

LİTERATURE REVİEW

In this section some few studies about machine Translation in Iran are introduced and discussed.

Jahangiri and Erfani (2003, p. 89) in their paper argue that machine translation is important because of its wide use in the development of scientific knowledge; and also its relation to other fields of study has related science and technology to each other. They have analyzed the function of the linguistic components of the system, in addition to expressing the history and the strengths and weaknesses of machine translation systems; they have also analyzed the methods of this kind of translation. In this research, to be familiar with a structural analysis admission computer program, some systems have been introduced; and special attention is paid to semantic domain for its importance in deepening the source language analysis as an effective element in better functioning of the machine translation system.

Falahati and Nemati (2004, p. 84) mentioned that the purpose of their paper is to trace at least some of the problems of machine translation. The results indicate that "Padideh" translation machine has many problems that their sources can be found in three things:

- 1. Lack of experienced linguists among the group of programmers
- 2. A Syntactic based linguistic theory
- 3. Lack of contemplation and in translator machines.

They found that the first type of problems can be remedied by applying experts. The second type of problems can be solved by revising the linguistic theory and paying attention to meaning and semantics. But at least under present conditions, the machine can't be equipped with a power to reason and therefore the problems of the Third type are not solvable now. They concluded that using machines instead of human is easier and faster and more reasonable in translating internet texts and large volumes of data, particularly when the goal is to understand the general points of the text.

Taghavi Pourmehran (2004, p. 50) explains the aims of machine translation in a paper and presents some reasons about the scientific, philosophical, economic and social needs to machine translation for Iranians. He has shown that in the twenty-first century, not only the good quality of translation is important, But also there are other factors that make machine translation attractive for translators and others. The researcher briefly reviewed some kind of machine translation; the rule-based structures, the transformer kind and the linguistic knowledge. Also, the Empirical structures, the example-based and statistical kind are analyzed. He concludes that the best machine translation systems are the systems which make use of the combination of various structures. An example of these machines is the Multi-Pass Machine Translation.

Marchuk (2004, p. 121), in a paper titled "An Analysis of terminologies for use in machine translation" believes that living languages in the contemporary world are faced the phenomenon of saturation of words by technological and scientific terminologies. Formerly, when speaking about the terms of a specific knowledge for example medicine, it was expected that the existing terminologies in this scientific field, have one standard meaning. But this thought applies no longer and the terminologies in this science or other sciences are not restricted to just one meaning and hence these terminologies are of interest to linguists. He believes that by the development of machine translation and the need to computational syntactic analyses, this question was proposed that how the terminologies can be categorized and put in entries. He notes that for the analysis of morphology, his work will be easier, particularly because Russian language is a morphological language and programming for understanding the morphological status of the words is possible with required attention. But it is much more difficult to determine the meaning of words and select the right option. Because in an intermediate step, the syntactic relationships between the words should also be

specified, beforehand. In this paper, he presents some practical strategies which are the result of long years of research and work on machine projects.

Valipoor (2007, p. 55) in his research compares the performance of the best Russian and western machine translation systems in order to recognize their weaknesses and the reasons of their superiority in relation to each other. He has also tried to identify the underlying causes of machine translation systems' inefficiency and provide guidelines to improve the quality of translations.

Mirzaeeyan (2004, p. 15) believes that today many opinions are expressed about machine translation and computer capabilities in this respect. Due to the lack of understanding of the computer power in this field of and insufficient knowledge of the people, he has tried to briefly introduce machine translation and its obstacles and problems, in addition to evaluating and testing the performance of three known software in Persian language named Pars, Padideh and Google translate. He believes that machine translation is a very important and extensive matter and in the absence of economic support in this area, one can't be limited to such applications and there would be a long way left to arriving at a powerful translator.

Mousavi (2010, p. 20) believes that collocations, that is recurrent combinations of words together that their co-occurrence possibility is more than occurring randomly, are seen very extensively in natural languages. Since the bilingual dictionaries are not able to present the suitable equivalents for most of these collocations, the vast majority of machine translation systems when facing such combinations do not perform accurately and thus the quality of their output will be considerably reduced. He, in a research, examines the process of building and using a large monolingual Persian corpus. He believes that this corpus enables us to promote the problem of English collocations ambiguity when translating them into Persian by machine translation system. By using such corpus as a target language corpus and also a bilingual English-Persian dictionary, the efficiency of this corpus in finding the most relevant Persian equivalent for the English collocations is evaluated in order to improve the quality of machine translation system. The result of the Experiment which was performed on a test corpus was very hopeful and its accuracy was 83/90%.

Machine Translation

Language is an effective medium of communication. It explicitly represents the ideas and expressions of human mind. More than 5000 languages exist in the world which reflects the linguistic diversity. It is difficult for an individual to know and understand all the languages of the world. Hence, the methodology of translation was adopted to communicate the messages from one language to another. Developments in Information Communication and Technology (ICT) have brought revolution in the process of machine translation. (Tripathi and Sarkhel, 2010, p. 388)

The translation of natural languages by machine, first dreamt of in the seventeenth century, has become a reality in the late twentieth. (Hutchins, 1995, p. 431)

From 1954, the term "Machine Translation" was used gradually. The Phrases "electronic translator" and "automatic translator" were sometimes used. But the effort to build a machine translator began at the 1930s. Thitherto the machine translator was thought of as a simple tool like a calculator which easily translates from one language to another .This Oversimplification was gradually replaced by realism. One of the researchers after two decades of research on machine translation, considered it as one of the most complex activities that man has ever achieved.

Machine translation is a kind of simulation of the human brain and hence a branch of artificial intelligence (Samaei, 2005, p. 50)

The boundaries between machine-aided human translation (MAHT) and human-aided machine translation (HAMT) are often uncertain and the term computer-aided translation (CAT) can cover both, but the central core of MT itself is the automation of the full translation process. Although the ideal goal of MT systems may be to produce high-quality translation, in practice the output is usually revised (post-edited). It should be noted that in this respect MT does not differ from the output of most human translators which is normally revised by a second translator before dissemination. (Hutchins, 1995, p. 431)

Several tools, free as well as proprietary, are now available which support translation of text into one or more languages over internet, online translation is offered by Yahoo and AltaVista through Babel fish. *Bing Translator* of Microsoft and *Google Translate* from Google are tools widely used for the translation by librarians and other members of web community. Firefox uses Grease monkey application to translate the text in other languages. Google Chrome Beta offers translation if the accessed web page is in a language other than default language (mostly English) (Tripathi and Sarkhel, 2010, p. 388).

INTRODUCING PARS TRANSLATION MACHINE

This machine translation system translates English text into Persian. The first commercial version of the translator came into market with the efforts of Iranian experts in Mabna software enterprise in September 1997; its first version was presented in DOS environment; the latest improved version of Pars Translator was released in June 2003. At the entry of the Pars machine translation, the English text is typed or some contents are selected from a file. Nowadays, Pars translation engine is capable of detecting and analyzing more than one million Five hundred General Terms and Common specialized terminologies in specialized fields of study (Pars Translator website http://www.parstranslator.net/far/). Some of these fields include: Commerce and Trade, Agriculture and Fisheries, computer Science, sociology, Civil Engineering, law, banking, Air commerce, accounting, economics and management, industrial engineering, military science, knowledge of engineering, metallurgy, physical education, water engineering, Lapidary, religious history, planning and management, electrical engineering, geology, mining engineering, mechanical engineering, psychology, Food industry, political science, journalism, engineering, textile engineering, philosophy, biology, medical sciences, oil and gas.

DATA ANALYSİS

In this section some example of sentences translated by Pars Translator are provided so as to review its problems and strengths.

1. The passenger flight arrival time changes every summer.

Translation: zæma:ne vo:ru:de pærva:ze mo:sa:fer hær ta:besta:ni tæghir midæhæd (transitive verb)*.

English Equivalent: every summer's passenger flight arrival time changes

The right translation: zæma:ne vo:ru:de pærva:ze mo:sa:fer hær ta:besta:n tæghır mıko:næd

(intransitive verb).

زمان ورود پرواز مسافر هر تابستانی تغییر میدهد.

As can be seen in the translation presented by Pars Translator, it has problem in properly recognizing the phrases in the sentence; in this sentence "every summer" which is the adverb of time is added to the noun phrase "passenger flight arrival time"; also the verb must be translated into an intransitive verb in Persian, but as is seen the transitive equivalent is selected by Pars.

In the next sentence "changes" is used as a noun to see whether the machine is able to recognize and separate verbs from nouns or not.

2. The passenger flight arrival time changes will be posted.

Translation : mo:sa:fer flight tæghıra:te zæma:nevo:ru:d ersa:l ʃo:d xa:hæd bu:d[†].

English Equivalent: Passenger's flight arrival time changes is posted it will be

The right translation: tæghıra:te zæma:ne vo:ru:de mo:sa:fera:n etela: da:de xa:hæd ʃo:d.

In this example the machine properly recognized that "changes" here is a noun. But it didn't translate "flight" into Persian and it also had problem in recognizing the future passive verb "will be posted" and it was translated as two verbs "is posted it will be!!"

3. The student enters the classroom.

Translation: da:nefa:mu:z kela:se dærs va:red mifævæd[‡].

English Equivalent: the classroom student enters

The right translation: da:nefa:mu:z be kela:se dærs va:red mifævæd.

In this sentence the main problem of the machine is that it must know that the verb "enter" in Persian is used with a preposition. The machine didn't use this preposition and so the word order of the sentence is disordered and "the classroom" which is an adverb of place is combined with the noun phrase "the students".

To see whether the machine is able to successfully translate some examples of a compound noun phrases example 4 was given to it:

4. *Vertical stabilizer lower spar attachment fittings.*

Translation: ghætæa:te zæmime tiræke æsli:tæsbitko:nænde gha:eme zirin[§].

Example 1 and 4 show us that the machine is almost able to translate nominal phrases and has little or no difficulties in their translation.

It seems that the machine did not have an understanding of idioms and proverbs and their translation is quite literal and word by word and is not able at all to Translate Proverbs. To clarify the point Example 5 is given to it:

5. Take the bull by the horns.

Translation: ga:ve nær be væsile fa:xha:**.

English Equivalent: bull by the horns

The right translation: be esteghba:le xætær ræftæn.

[†]مسافر flight تغییرات زمان ورود ارسال شد خواهد بود. ‡دانشآموزکلاس درس وارد میشود. [§]قطعات ضمیمه تیرک اصلی تثبیت کننده قائم زیرین *گاو نر به وسیله شاخها This translation shows that the machine is not even able to give the reader an overall impression of the proverb and it is really defective in this important respect. The translation is too far from the right meaning that is "to risk".

6. To turn the tables^{\dagger †}.

Translation: bæra:ye gæſtæne mızha:.

English Equivalent: for turning the tables

The right translation: ɔ:za: ra: be næfe xɔ:d tæghır da:dæn.

This translation is quite literal and word by word and shows the machine's inability in translating idioms.

7. To feather my nest.

Translation: be pær a: ſi:ja:neæm^{‡‡}.

The right translation: dyibe xo:dæm ra: po:r kærdæm.

English Equivalent: These idioms are all translated word by word and show further evidence to indicate that the machine has tremendous problem in translating idioms.

So far just general sentences were given to the machine, but from No. 8 specialized sentences such as sociology, computer science and physical education are provided to it. In the 8th example a sentence from the sociology book of Hudson e is given to it:

8. Both Sapir and Whorf worked extensively on American Indian languages and made important contributions to our knowledge of those languages and also to linguistic theory (among other things).

Translation: hær do: sæpir væ vo:rf dær zæba:nha:je so:rxpu:st ka:r kærd væ sæhmha:je mo:hem be to:re mo:mtædi be æz da:nefma:ne a:n zæba:nha: væ hæm be næzæri:je zæba:ni(bejne ffizha:je digær) besa:zi: $d^{\$\$}$.

English Equivalent: Both of them Sapir and Whorf worked on American Indian languages and important shares continuously from the knowledge of those languages and also to the linguistic theory (between other things) may make.

We understand that the machine is really incapable of translating the specialized texts, with little attention to the translation of these texts by the machine; it renders us difficult and meaningless sentences and does not convey the real meaning of the sentences. But the words in brackets are properly translated and it has understood the concept of parentheses. The proper names when they are capitalized are also recognized by it and are written as proper names and the machine has no problem in recognizing them.

In another part of the Hudson's book on the sociology, the translation is as the following:

9. Having clarified some of the connections between language, culture and thought, a Mention can now be paid to the two issues which have dominated the study of language in relation to culture and thought. Translation: meghda:ri æz pejvæstegiha: ma:bejne zæba:n væ færhæŋ mo:ſæxæs ſo:de æst væ fekr beko:ni:d ha:l be do: næſr pærda:xt ſævæd ke mo:ta:leje zæba:n dær ra:bete be færhæŋ tæsælo:t pejda: ko:nid væ fekr ko:nid

The above translation suggests that the machine sometimes has problem in distinguishing a noun from a verb. For example the word "thought" in this paragraph is translated as a verb. It provides an influent and unexpected translation. It has also problem in expressing the tense of the verbs; For example, the present perfect is translated as simple past and so the meaning is far away from the original expected meaning.

In The next text which is a specialized computer text, such mistakes are also seen more or less:

10. Your reference number and PiN will be sent in two separate emails to the e-mail address you specified.

Translation: væ $\int \mathfrak{d}$: ma:re ro:dzu:eta:n dær do: dzo:da: po:ste elektero:niki be nefa:ne elektero:niki mo:fæxæsfo:dePiN xo:de fo:ma: feresta:de xa:hæd fo:d†††.

As is seen the conjunction "væ" at the start of the translation is additional. The word PiN is translated apart from its coordinate noun-phrase "reference number". The Adjective "isolated" is put before the noun as in English, while it should come after the noun as the word order of Persian requires. The active verb is also translated as a passive verb.

11. If you have several e-mail or newsgroup accounts, you can work with all of them within one window.

Translation: ægær ʃɔ:ma: pɔ:ste elekterɔ:nıkı tʃændın beda:rıd ja: gɔ:ru:he xæbærı be hesa:b bija:værıd ʃɔ:ma: mɪtæva:nɪd ba: tæma:mɪ a:nha: ra: dær mɔ:dæte jekı pændʒere ka:r kɔ:nɪd ⁺⁺⁺.

In translation, the adjective "several" which should come before the noun in Persian, is placed after the noun; In addition, the noun "accounts" is again translated as a verb and in general, the translation is so inexpressive and defective.

12. Computers are changing all our lives and also old ways of doing things with their superhuman speed.

Translation: ka:mpiju:terha: da:rænd tæma:me fjizha: ændza:m da :dæne mæsirha: hæm ghædm zendeha: jema:n vej ba: so:ræte ensa:ne xu:befa:n tæbdil miko:nænd^{§§§}.

After the translation of this sentence it seems that we can't really depend on machine translation; we should rely on human translator and the better that we are discouraged from machine translation; because it not only doesn't express the exact meaning, but also it somehow deviates us from it.

13. The average of 3, and 8 is 6.

Translation: mija:ngine 36 hæst 7 væ 8 væ^{*****}.

مقداری از پیوستگیها مابین زبان و فرهنگ روشن شده است و فکر بکنید حال به دو نشر پرداخت می شود که مطالعه زبان در رابطه به فرهنگ تسلط پیدا کنید و فکر کنید.

خود شما PiN و شماره رجوعتان در دو جدا پست الکترونیکی به نشانی الکترونیکی مشخص شده ^{۱۱۱} فرستاده نحوامد شد. ***

[:] اگر شما پست الکترونیکی چندین بدارید یا گروه خبری به حساب بیاورید، شما میتوانید ^{###} با تمامی آنها را در مدت یکی پنجره کار کنید. ۱۹۶۹

^{§§§} *کامپیوترها دارند تما*م چیزها ا*نج*ام دادن مسیرها هم قدیم زندههایمان وی با سرعت انسان خوبشان تبدیل می*ک*نند.

The only successful act of the machine in translating this phrase is recognizing the right meaning of the word "average" among its different meanings.

14. I don't like your way of life.

Translation: mæn ra:he zendegita:n du:st næda:ræm^{††††}.

The machine has problem in translating the simple sentences like the one above and it doesn't provide an impeccable translation. Here the object sign in Persian "Ra" is not mentioned and the so sentence is not natural.

In the following sentence taken from a psychology text, the machine translates as the following:

15. *A greeting is a way of being friendly to someone.*

Translation: jek tæbrik jek bu:dæne ra:he du:sta:ne be kæsi æst^{‡‡‡‡}.

The word "greeting" is translated wrongly as "congratulations and "being friendly" is translated word by word translation and so it's not eloquent and fluent like that of a human translator.

The following sentence selected from physical education specialized texts, is translated somehow successfully with some little mistakes:

16. The Olympics attract a lot of people every year.

Translation: ba:zıha:je o:læmpik besja:ri mærdo:m hær sa:le dzæzb miko:næd^{§§§§}.

This translation also lacks object sign in Persian "Ra" and also the preposition "from".

The next sentence is also translated as the following:

17. *A satellite is any natural body like the moon that orbits another object.*

Translation: jek ma:hva:re ma:h ſæbihe bædæne tæbii: æst ke dær næslha:je dīgær dær mæda:r ghæra:r mīgīræd^{*****}.

CONCLUSION

All these translations suggest that machine translations have still a long way to success and to act as a professional human translator. The Pars Translator machine has problem in translating simple English sentences, proverbs and general and specialized texts and we can't rely on it at all. It is better that we do not look at it optimistically and wait for a day to believe in machine translation's work, but it is still too early for this. Machine translation is more Problematic than helpful because we should edit its translation and it takes a lot of time. So instead it's better to spend time on translating by a human translator. In this case, the translation is more fluent and eloquent and above all we have a human translator. As seen from the translations by Pars translation machine every sentence, even the simplest ones need to be edited by a human editor; so it's not economical at all and it's time wasting. So it's better that we presently forgo Machine Translation.

میانگین 36 است 7 و 8 و ... **** §§§§ ۱۳۳۶ بازیهای المپیک بسیاری مردم هر سال جنب میکند. ***** سسی

REFERENCES

- [1]. Bolshakov, I., & Gelbukh, A. (2004). *Computational Linguistics: Models, Resources, Applications*. México DF: DirecciÓn de Publications.
- [2]. Falahati, M., & Nemati, A. (2004). "*Human Translator and Machine Translation: a Case study about the problems of machine translation of English into Persian*" Librarianship and Information science Phenomenon, Issue 7, No. 2, row 26.
- [3]. Hutchins, W. H. (1995). *Machine translation: A brief history*. Koerner, E.F.K. & Asher, R.E. (Eds.), *Concise history of the language sciences: from the Sumerians to the cognitivists* (pp. 431-445). Oxford: Pergamum Press.
- [4]. Jahangiri, N., & Erfani, M. (2003). "*The Machine Translation Process and its Restrictions*", faculty of literature and Humanities of Ferdowsi University of Mashhad's Journal, No. 4.
- [5]. Marchuk, Y. N. (2004). "Analysis of Collocations for use in Machine Translation" Foreign Language Research, No. 19: 43 – 58.
- [6]. Mirzaeeyan, H. R. (2009). "Problems of Machine Translation in Persian by Three Softwares", Translation Studies, No. 28.
- [7]. Moosavi, M.T. (2010). "*The function of large Monolingual corpus in improving the quality of machine translation*", Translation Studies, No. 29.
- [8]. Ramanathan, A. (2008). *Statistical machine translat*ion. Retrieved at July, 28, 2014, from 202. 141. 152. 9/clir/ papers/statistical_mt.pdf.
- [9]. Samaei, M. (2005). "*Numerical phrases in Machine Translation*", Science and information technology Quarterly, Issue 21, No. 1: 21 33.
- [10]. Taghvaei, P. A. (2004). "A look at Machine Translation for Persian Language", Translation Studies, No. 7 & 8.
- [11]. Tripathi, S., & Sarkhel, J. K. (2010). "Approaches to machine translation". Annals of Library and Information Studies, 57, 388-393.
- [12]. Uszkoreit, H. (2000). *What is Computational Linguistics*? Retrieved July 28, 2014, from <u>www.coli.uni-saarland.de/~hansu/what_is_cl.html</u>
- [13]. Valipoor, A. R. (2007a). "Comparing and reviewing some Russian and Western Machine Translation Systems", Foreign language Research, No. 22: 121 142.
- [14]. Valipoor, A. R. (2007b). "Analyzing the Linguistic Aspect of Machine Translation and a Brief Introduction of Related Sciences", Foreign language Research, No. 37: 103-129.

WEB REFERENC

http://www.parstranslator.net/far

APPENDİX

زمان ورود پرواز مسافر هر تابستانی تغییر می هد.* مسافر flight تغییرات زمان ورود ارسال شد خواهد بود. دانش آموز کلاس درس وارد می شود. ٹقطعات ضمیمہ تیرک اصلی تثبیت کنندہ قائم زیرین گاو نر به وسیله شاخها برای گشتن میزها جیب خودم را پر کردم و بار خودم را بستم. هر دو «سپیر» و «ورف» در زبانهای سرخپوست کارکرد و سهمهای مهم به طور ممتدی به از دانشمان آن زبانها و هم به نظریه زبانی (بین چیزهای دیگر) بسازید. مقداری از پیوستگیها مابین زبان و فرهنگ روشن شده است و فکر بکنید حال به دو نشر پرداخت میشود که مطالعه زبان در رابطه به فرهنگ تسلط پيدا كنيد و فكر كنيد. و شماره رجوعتان در دو جدا پست الکترونیکی به نشانی خود شما فرستاده خواهد شد. PiN الکترونیکی مشخص شده : اگر شما پست الکترونیکی چندین بدارید یا گروه خبری به حساب بیاورید، شما میتوانید با تمامی آنها را در مدت یکی ينجره كار كنيد. *کامپیوترها دارند تمام چیزها انجام دادن مسیرها هم قدیم زنده هایمان وی با سرعت انسان خوبشان تبدیل میکنند. میانگین 36 است 7 و 8 و ... *من راه زندگیتان دوست ندارم. یک تـبریک، یک بودن راه دوستانه به کسی است. بازیهای المپیک بسیاری مردم هر سال جذب میکند. یک ما هواره ماه شبیه بدن طبیعی است که در نسلهای دیگر در مدار قرار مےگیرد.