

WCETR 2011

An Investigation of the Translation Problems Incurred by English-to-Persian Machine Translations: "Padideh, Pars, and Google Softwares"

Zohre Komeili^a, Jalil Allah Farughi Hendavalan^b, Ali Rahimi^{c*}

^aLinguistics Department, Islamic Azad University, Ahar Branch, Iran

^bLinguistics Department, Islamic Azad University, Ahar Branch, Iran

^cDepartment of English, Faculty of Humanities, University of Kashan, Iran

Abstract

In this paper, an effort is made to offer some approaches and solutions to remove the problems in the softwares of English to Persian translations through their comprehensive analysis. Machine translation is the automatic transfer of message from one language to another. Considering the irregularities and problems in translating texts from English to Persian caused by available softwares of machine translations, here, we are going to show that through the introduction of an appropriate framework we are able to remove the problems of the English translation weaknesses. The theoretical backgrounds and review literature are discussed in details. The linguistic aspects of machine translation, the available systems, the problems, translation approaches and evaluation of machine translation systems are also discussed. The machine translation softwares of Padideh, Pars and Google as well as bilingual dictionary of Hezareh are introduced. Data are collected and based on the obtained results we can conclude that through an analysis of lexical networks and analytic systems, we are able to reduce the current problems and enhance the quality of machine translation.

Keywords: Machine translation; Computational linguistics; Corpus studies; Translation studies; English/ Persian translation softwares.

1. Introduction

For many years, translation has been a positive and an effective factor in making different nations closer and it has made these nations share each others' experience in different areas of culture, education, technology, industry, etc. The phenomenon which we are faced as 'Global Village', has consequently been the result and the achievement of these relationships and transactions whose main body has been translation [7].

Comparing and contrasting different transactions, the readers, at times, conclude that some transactions can not convey a similar feeling when they read the original work. At these times, the acute need of an efficient tool without any personal remark and human feeling, which simultaneously regards the peripheral issues of the translation, is highly felt. In this regard computer has been chosen [3]. Computers have artificial intelligence which neither can do a job willingly nor do they have a magic power to do what they are not supposed to do. Their intelligence is as much as what human provides them with. Having an enormous memory, they are not able to choose or make decisions unless human writes programs for them. In this case they can do their job good and quick as much as the data put in their memory [1].

* Ali Rahimi. Tel.: +98-913-309-9535; fax: +98-361-555-2930

Whether the translation is done by human or computer, it will face challenging problems. The research projects related to machine translations in 1940s and 1950s became the incentive for a plenty of primary researches in computational linguistics and [10] it is still going on.

This research tries to evaluate the translations of three softwares named, Pars, Padideh, and Google (English to Persian). This paper aims to investigate the problems of machine translations of softwares from English into Persian. These machine translations are claimed to be as follows: "Recently, Pars Translator Engine is able to identify and analyze more than one million and five hundred thousand general words and technical terms in thirty eight fields of study. The treasury of words and expressions of Pars translation, is getting richer and it is promoted by our colleagues in scientific and academic centers"[15].

"Padideh Translator is unique from different aspects. This software has several information banks which encompass more than one and a half million words and phrases in general English and eighty seven technical areas. Padideh Translator Engine employs the most developed techniques of artificial intelligence and natural language processing [14]."

"Using thousands of available passages and documents in the UN, Google has gleaned more than two hundred billion words and employing comparative-statistics, it provides its users with the most versatile words and phrases."

2. Current problems in machine translation

Machine translation has some problems which are related to the employed strategy. For example, in transmission strategy, a transmission stage is needed. This transmission stage requires a bilingual dictionary by which the computer can solve many word/phrase-related problems. This part copes with the problems which might be created in machine translation and computational linguistics [7].

2.1. Lexical problems

Lexicon is list of words in which each word belongs to a syntactic category such as, noun, verb, adjective, etc. Lexicon can also be related to the sub-syntactic categories such as, transitive/intransitive verbs, proper/generic nouns, etc [8].

EX.: Paper, pencil and pen are the essential ingredients of stationery.

Pars: paper, pencil and pen are the essential tools for writing.

Padideh: paper draws with pen and pencil are necessary of stationery.

Google: paper, pencil and pen are the essential ingredients of chips.

2.2. Word conjugation and ambiguity

Lexicology and conjugation are created by recognizing and combining morphemes to make new words. Word combination is usually introduced by conjugation in which the words are combined by themselves and create new words. The meanings of these new combined words are sometimes virtually different from the word-form or they

are sometimes completely ambiguous or unpredictable. But the major problem of word combination in the scope of conjugation is that in some languages the two combined words in the combination process are immediately placed after each other and there is no hyphen or space between the two words.

EX: *Fruit flies like peach.*

Google: fruit flies such as peach.

Pars: fruit flies alike.

Padideh: fruit loves peach flies.

2.3. Syntactic problems

Syntax is putting words together or in order [9] or the principals by which words are ordered to make sentences [8].

EX: *Don't waste your money on gold.*

Pars: Waste of your money on gold.

Padideh: T-waste your money on gold.

Google: Does your money isn't on gold.

2.4. Problems at the level of production and transmission

Generally speaking, the production of the target language is much easier than analyzing the input of the source language. The production of the target language is based on the input passage analysis. One of the problems of the transmission stage is the change in the verb-tense. The order of the words in the target language sentence should be heeded so that the meaning could be conveyed appropriately.

EX. *She has been reading the novel before we arrived.*

Pars: she has been reading the novel before we arrive.

Padideh: she has been reading the novel before we arrive us.

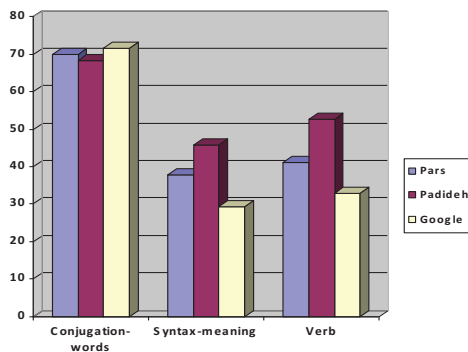
Google: she warned she refuse the proposal.

Table 1: Means of Translation Softwares

Translation Softwares	Conjugation-words	Syntax-meaning	Verb	Mean
Pars	70.3%	37.75%	41%	49.68%
Padideh	68.3%	45.75%	52.75%	55.6%
Google	71.5%	29.25%	32.75%	44.5%
<i>Mean</i>	70%	37.58%	42.16%	49.92%

According to the research hypothesis which claimed that machine translation softwares are problematic in translating English sentences into Persian, after evaluating one hundred English sentences involving simple, combined, and complicated sentences and idioms, the hypothesis is corroborated. Table 1 demonstrates that machine translators could just translate fifty percents of the English sentences into Persian successfully. These translators had the best choice in finding equivalent (70 percents) in lexicon. In this research, verbs were studied separately because the core of each sentence is its verb and the closer the verb in source language is to the translated verb in the target language, the translation would be more intelligible for the reader. This study showed that the aforementioned

softwares could just translate 43 percent of the English verbs appropriately into Persian. In this regard the verb tense was also put into consideration. In the discussion of syntax and word order, it is ascertained that the machine translators are poor (38 percent) in syntax and conveying meaning.



3. Conclusion

Although there are some problems in the translation of the machine translators, we cannot dismiss their great help and advantage of these softwares because they assist human in saving time and money. Translating in any time and any moment in any language is among the eye catching features of these translators. So, applying highly developed and specified network of lexicon and also devising an error-finding system in Persian language, the ability of these translators would be extremely ameliorated. All in all, as computer has a suitable memory for such activities, the machine-translation problems could be decreased to its smallest degree.

References

- Beikian, Ali (2005), An Evaluation of the Output Quality of Two Prevalent English Persian *Machine Translation*. *Pars Translator and Padideh Translator*. thesis in MA of Arts in Translation Studies.tehran, Allameh Tabataba'i University
- Hutchins, J. *Machine translation: a concise history*. <http://ourworld.compuserve.com/homepages/WJHutchins/>
- Hutchins, W. J. & Somers, H. L. (1992), *An introduction to machine translation*. London: Academic Press.
- Hutchins, W. John (1996), *ALPAC: the (in)famous report*. *MT News International*,14, June 1996, 9-12. Repr. (with minor revisions) in: Nirenburg, S. & Somers, H.& Wilks, Y. (Eds.) (2003), *Readings in machine translation* (pp. 131-135).Cambridge, Mass.: The MIT Press.
- Hutchins, W., John, (2000), *Warren Weaver and the launching of MT: brief biographical note*, In Hutchins, W. John (ed.) *Early years in machine translation* (pp. 17-20), Amsterdam: John Benjamins
- Padideh Translator* Home Page. <http://www.padideh.org/>
- Pars Translator* Home Page. <http://www.parstranlator.net/eng/index.htm>
- Rahimi, A. & Yazdanbakhsh, A.(2009), *Dictionary of Spoken Phrases*, Tehran, Jungle.