A Contrastive Analysis of Machine Translation (Google Translate) and Human Translation: Efficacy in Translating Verb Tense from English to Persian

Hakiminejad Afshin*  
Dr. Mohammad Ali Alaeddini 2

1 Instructor, Department of Foreign Languages, Payame Noor University, Iran  
2 Assistant Professor, Department of Foreign Languages, Payame Noor University, Iran

Doi:10.5901/mjss.2016.v7n4s2p40

Abstract

This study aims to examine the efficiency of Google translation in translating verb tense from English to Persian, the acceptable translation of Google translation, and the areas of possible errors in the Google translation. To this end, Google translated four passages of Oliver Twist. The Google translations were compared with the translation of Gharib based on the selected codes for the domain and extent of errors. The results were analyzed and evaluated in terms of accuracy and appropriateness. The findings of the study showed that Google Translate is not able to translate verb tense from English to Persian and the translated passages are not acceptable. Also, most of the grammatical errors occurred in translation of aspects, passives, and compounds, respectively.

Keywords: domain of error, extent of error, machine translation, human translation

1. Introduction

The origin of translation goes back to human being’s demand for communication. However, ever-expanding information exchange has changed the kind of human being’s demand for translation. Now, no one expects when s/he is reading an internet page and does not comprehend some components or whole of it; a translator assists her/him and renders the text [1]. This major change in everyone’s viewpoint and rapid speed of information exchange in the modern era has resulted in thinking of developing a system, which is capable of translating the given text automatically and independently. This system has been called machine translation [1]. Automatic machine translation was one of the first natural language processing operations analyzed in computer science and has captivated many researchers across the world for over 50 years[16]. Within this period, many paradigms have been acquainted and expanded: Dictionary-based machine translation, Example-based machine translation and Rule-based machine translation [16]. Moreover, paradigms for evaluation of machine translation emerged. Hutchins [9] states that despite the general agreement about the main features of machine translation evaluation there are no generally accepted and certain methods and scales for it. Meanwhile, he (ibid) suggests three types of evaluation: Adequacy, Diagnostic, and Performance. In this study, the translations of verb tense produced by Google Translate were compared with the translation of Gharib [5] using Hutchins’s [9] second category, diagnostic one for evaluation of Google Translate and Lennon’s(as cited in Brown [2]) category for description of errors.

2. Literature Review

2.1 The Concept of Translation

House [7] puts forward two definitions related to translation: ‘overt translation’ and ‘covert translation’. In an overt translation, the receivers of the rendition text are completely "overtly" not being clearly addressed. A covert translation possesses the situation of a genuine source text in the target culture. Newmark [13] suggests two terms for translation: ‘semantic’ and ‘communicative’ translation. Communicative translation creates on its readers the same effect that acquired on the readers of the original. Semantic translation heeds, as closely as to both the semantic and syntactic structures of the second language [13].

According to Hutchins [8], translation is the interpretation of a linguistic meaning of a text in the source language
and production of a new equivalent text in the target language. Traditionally, translation has been a human activity although attempts have been made to automate and computerized the translation of texts. The goal of translation is establishing a relation of content between the source text and the target text, and considering a number of constraints including context, rules of grammar of language, writing conventions, idioms, etc. [8].

Tymoczko (as cited in Calvo [17]) demonstrates that 'translation' is a bunch sense used in various cultural, social, and technological conditions because of its boundless nature.

2.2 Errors in Translation

Lennon (as cited in Brown [2]) states that there are several categories for the description of errors on the language of learner:

1. The most overall classification can be recognized by errors of addition, omission, substitution, and ordering, ensuing standard mathematical categories.
2. Within each category, levels of language can be taken into account as phonology, lexicon, grammar, and discourse. Often, of course, it is difficult to differentiate various levels of errors.
3. According to Burt and Kiparsky’s (as cited in Brown [2]) model, errors can be seen as global or local. Global errors impeded communication; they avert the hearer from grasping some aspect of the message. Local errors do not impede the message from being heard usually because there is just a slight violation of one segment of a sentence, permitting the hearer / reader to make a precise guess about the expected meaning.
4. Domain and extent are the two associated features of error in any error analysis. Domain is the rank of linguistic unit from phoneme to discourse that must be considered as context in order for the error to become patent, and extent is the rank of linguistic unit that would have to be deleted, replaced, supplied, or reordered in order to repair the sentence.

2.3 The History of Machine Translation

Chéragui [3] summarizes the history of machine translation in five periods as follows:

2.3.1 First Period (1948-1960)

- 1949: Warren Weaver in his note of 1949 tendered the first opinion on the application of computers in translation, by excerpting the word computer translation.
- 1954: The emergence of the first automatic translator by some researchers from Georgetown University in collaboration with IBM, translating more than 60 Russian sentences into English.
- 1954: Victor Yngve issued the first journal on MT, called « Mechanical translation devoted to the translation of languages by the aid of machines »

2.3.2 Second Period (1960-1966)

- Early 1960s the parsing is propounded for research to develop the MT. Thus, there are previously numerous parsers developed from various sorts of grammars, such as grammar and dependency grammar Tesnière Stratificationnelle Lamb.
- 1961: Computational linguistics is born in this year, regarding lectures set up by David G. Hays at the Rand Corporation in Los Angeles.
- 1964: the establishment of ALPAC(Automatic Language Processing Advisory Committee) with American government to investigate the viewpoints and the opportunities of MT.
- 1966: ALPAC issued his well-known statement emphasizing that its endeavors on MT is only squandering of time and money.

2.3.3 Third Period (1966-1980)

- 1970: Russian researchers commenced the project REVERSO.
2.3.4 Fourth Period (1980-1990)

- 1983, NEC advances the system of translation entitled PIVOT. 1986: Promotion of system PENSEE by OKI, which is a Japanese-English translator.
- 1986: Hitachi improved his own translation system based on an approach obtained by transfer (Japanese-English).

2.3.5 Fifth Period (since 1990)

- 1993: The project C-STAR (Consortium for Speech Translation Advanced Research). The theme of project is in the area of tourism.
- 1998: Trading REVERSO by the firm Softissimo.
- 2000: The advancement of system ALPH (Japanese-English and Chinese – English). This translator is based upon example-based approach.

2.4 Evaluation of Machine Translation

2.4.1 Human Evaluation

Hutchins [9] states that because humans are the best standard for using language, human evaluation is the necessary for the evaluation of machine translation. However, human evaluation is time consuming and thus not always an option. Additionally, maintaining objectivity is important in human evaluation. Now, most machine evaluation metrics are based on human references. In the case of human evaluation, the evaluation is often involved in the research, but one can take same approaches with MT and just use a couple of evaluation [9]. Hutchins (ibid) states that there are no generally accepted and certain methods and scales for the evaluation of machine translation, and evaluation methodology has been the subject of much discussion in recent years. Meanwhile, he suggests three types of evaluation:

2.4.1.1 Adequacy Evaluation

This evaluation determines the suitability of MT systems within a specified operational context. It is usually performed by potential users and/or purchasers of systems, individuals, companies, or agencies. Adequacy evaluations by potential purchasers typically involve the testing of systems with a series of typical documents. Nevertheless, these are necessarily limited to specific areas and generally applicable and objective test suites are required for diagnostic and performance evaluation. These test studies are now under development.

2.4.1.2 Diagnostic Evaluation

This evaluation characterizes imitations, errors and deficiencies, which may be corrected or modified by the research team or by the developers; researchers and developers consider it.

2.4.1.3 Performance Evaluation

This kind of evaluation is concerned with the stages of system development or various technical usages; it may be done by researchers, developers, or potential users. Features involved in MT evaluations are not used in evaluations of other systems. These features comprise the quality of the unedited translations, intelligibility, accuracy, fidelity, appropriateness of style or register; capability of facilities for making and updating dictionaries, post-editing texts, managing input language, customization of documents; the expansion to new language pairs or new subject areas; and cost benefit comparisons with human translation performance.

- 1970: Peter Toma improved SYSTRAN1 (Russian-English)
- 1976: Formation of system WEATHER in the project TAUM by Alai Colmerauer for MT weather prediction.
- 1978: Generation of system ATLAS2 by the Japanese company FUJITSU, this translator can translate from Korean to Japanese and vice versa.
2.5 Persian Language

Persian is an Indo-Iranian part of the Indo-European languages using a revised Arabic script and is spoken in Iran and in the south of the Persian Gulf, and some other countries. Generally, 134 million people across the world use it as first or second language [16]. Persian is an inflective language in which different word-forms are made by the combination with affixes. In Persian, words are very incongruous, they can openly transfer along the sentence, and personal pronouns are optional [16]. In Persian, there are four written forms for a character relying on its location in a word. Many Persian words are written in various forms through deletion or addition of spaces within words or by using different forms of characters. Persian texts are kept using various scripts in the computers [16]. Megerdoomian (as cited in Pilevar & Faili [16]) asserts that Persian has high inflectional morphology: Nouns are either singular or plural. They may also come along with the ezafe, the indefinite marker, the enclitic particle, the possessive clitic pronoun and the copula.

2.6 The History of English-Persian Machine Translation

There are few English-Persian MT systems produced that some of them are given as follows:

Younesi Far (as cited in Ahangar et al [1]) planned and made a MT system based on Augmented Transition Networks (ATN). This system could generate a word for word translation of English sentences into Persian. Pars translator and Padideh translator are two common English-Persian MT programs. Pars translator is a MT system that translates English texts into Persian sentences. The first commercial version of software issued to public in June 1997 and the last updated version delivered in April 2004. Recently, it is able to recognize and scrutinize more than 1.5 million general words and technical terms in 38 fields of study. The depository of words and expressions of this software is developing optionally [16]. In Persian, words are very incongruous, they can openly transfer along the sentence, and personal pronouns are optional [16]. In Persian, there are four written forms for a character relying on its location in a word. Many Persian words are written in various forms through deletion or addition of spaces within words or by using different forms of characters. Persian texts are kept using various scripts in the computers [16]. Megerdoomian (as cited in Pilevar & Faili [16]) asserts that Persian has high inflectional morphology: Nouns are either singular or plural. They may also come along with the ezafe, the indefinite marker, the enclitic particle, the possessive clitic pronoun and the copula.

Following the literature, it is evident that no cohesive studies have been conducted on English-Persian MT and most of them are scattered in which translating verb tense and aspect have not been taken into consideration. On the other hand, with the rapid movement towards computer sciences and IT, unfortunately, any outstanding developments in English-Persian MT systems are not observed in Iran and the existing systems have some weak points, so they cannot satisfy our requirements. To this end, a comprehensive, cohesive, and systematic research is needed. Therefore, in the current research I made a comparison between Google translation and human translation in order to evaluate the efficacy of Google translate in translating verb tense from English to Persian and identified extent and domain of errors produced by Google translate. To do so, the following research questions were raised:

1. Is the Google Translate efficient in translating verb tense from English to Persian?
2. Is the translation produced by Google Translate acceptable apart from possible errors in translation?
3. What are the areas of possible errors in the translation produced by Google Translate?

Moreover, three research hypotheses were developed:

1. Google Translate is not efficient in translating tense from English to Persian.
2. The translation produced by Google Translate is not acceptable.

3. Passives and aspects are not regarded as areas of possible errors.

3. Method

This research has been designed to investigate the efficiency of Google translation in translating verb tense from English to Persian, the acceptable translation of Google translation, and the areas of possible errors in the Google translation. To this end, Google translated four passages of *Oliver Twist* [4]. The Google translations were compared with the translation of Gharib [5] based on the selected codes for the domain and extent of errors. To determine errors produced by Google Translate and to analyze the results, Lennon’s (as cited in Brown, [2]) category and Hutchins’s [9] category were used respectively. A detailed description of these categories is presented in the Procedure section.

3.1 Participants

The selected passages from *Oliver Twist* [4] as the sample, have been translated by Gharib [5] Therefore, there are not any subjects in this research.

3.2 Instruments

The instruments of this study are Google Translate, the selected passages from *Oliver Twist* [4] and the translation of *Oliver Twist* by Gharib [5].

3.3 Procedure

The procedure of this study can be summarized as follows:

1. Four passages of *Oliver Twist* [4] were selected.
2. The verbs in these samples were identified and underlined to make the comparison easier.
3. Samples were translated by Google Translate.
4. Translated samples were collected; the verbs were identified and underlined to be compared with the human translation of Gharib [5].
5. Translations produced by Google Translate and the translation of Gharib [5] were compared according to the selected codes for the domain and extent of errors, and were recorded in separate tables and bar graphs.

Lennon (as cited in Brown, [2]) states that there are several categories for the description of errors on the language of learner:

- The most overall classification can be recognized by errors of addition, omission, substitution, and ordering, ensuing standard mathematical categories.
- Within each category, levels of language can be taken into account as phonology, lexicon, grammar, and discourse. Often, of course, it is difficult to differentiate various levels of errors.
- According to Burt and Kiparsky’s (as cited in Brown [2]) model, errors can be seen as global or local. Global errors impeded communication; they avert the hearer from grasping some aspect of the message. Local errors do not impede the message from being heard usually because there is just a slight violation of one segment of a sentence, permitting the hearer / reader to make a precise guess about the expected meaning.
- Domain and extent are the two associated features of error in any error analysis. Domain is the rank of linguistic unit from phoneme to discourse that must be considered as context in order for the error to become patent, and extent is the rank of linguistic unit that would have to be deleted, replaced, supplied, or reordered in order to repair the sentence.

Therefore, in the evaluation of the translations produced by Google Translate, the extent of errors was considered as follows:

- Whether the translation produced by Google Translate is wrong.
- Whether Google Translate does not translate the word.
- Whether Google Translate makes shift in translation, which may include addition, deletion, substitution, or ordering.

According to Burt & Kiparsky (as cited in Brown, [2]) errors can be seen as global or local. Global errors
impeded communication; they avert the hearer from grasping some aspect of the message. Local errors do not impede the message from being heard usually because there is just a slight violation of one segment of a sentence, permitting the hearer / reader to make a precise guess about the expected meaning. Based on this definition, global errors are regarded as unacceptable errors and local errors are categorized as acceptable errors in this research. So, in the case of any errors or in the case of existence of local errors:

- Whether the translation produced by Google Translate exactly equals the human translation.
- Whether the translation produced by Google Translate equals the human translation with some shift.

6. Domain of errors in the translations produced by Google Translate was determined as follows:

- Lexicon
- Discourse
- Grammar (which may be)
- Passive Verbs
- Causative Verbs
- Compound Verbs
- Aspect of the Verbs

7. The results were analyzed and evaluated in terms of accuracy and appropriateness based on Hutchins's [9] category. Hutchins [9] categorizes evaluation in three groups:

**Adequacy Evaluation:** This type of evaluation determines the fitness of MT systems within a specified operational context; potential users or purchasers of systems, individuals, companies, or agencies typically perform it. Adequacy evaluation by potential purchasers usually includes the testing of systems with sets of typical documents. However, these are necessarily restricted to specific domains, and for diagnostic and performance evaluation there is a need for more, generally applicable and objective test suites; these are now under development.

**Diagnostic Evaluation:** This type of evaluation identifies limitations, errors and deficiencies, which may be corrected or improved by the research team or by the developers; it is the concern of researchers and developers.

**Performance Evaluation:** This type of evaluation examines stages of system development or different technical implementations. It may be undertaken by researchers, developers, or potential users. In the case of production systems there are also assessment of marketability undertaken by or for MT system vendors.

In this study, the second type of evaluation, diagnostic one, was used concentrating on the efficacy of Google Translate.

8. The results were appeared in descriptive tables and bar graphs.

### 3.4 Data Collection Procedure

Translations produced by Google Translate, and the translation of Gharib [5] were collected and along with the selected codes for the domain and extent of errors, recorded in separate tables and bar graphs. Domain of errors in translations of Google Translate was determined as follows:

- Lexicon: LE
- Discourse: DI
- Grammar: (which may be)
- Passive Verbs: GRPA
- Causative Verbs: GRCA
- Compound Verbs: GRCO
- Aspect of the Verbs: GRAS

The extent of errors was determined as follows:

- Whether the translation produced by Google Translate is wrong: WR
- Whether Google Translate does not translate the word: DON
- Whether Google Translate makes shift in translation, which may include addition, deletion, substitution, or ordering: SH/A, SH/D, SH/S, SH/O

In the case of absence of any errors or in the case of existence of local errors:

- Whether the translation produced by Google Translate exactly equals the human translation: EQ
- Whether the translation produced by Google Translate equals the human translation with some shift: ESH
3.5 Data Analysis Procedure

The results of this research were analyzed and evaluated in terms of accuracy and appropriateness based on Hutchins's [9] category. According to Hutchins [9], there are three types of evaluation for MT systems: Adequacy evaluation, diagnostic evaluation, and performance evaluation. In this study, the second type of evaluation, diagnostic one, was used concentrating on the efficacy of Google Translate.

4. Results

Based on the analysis of the translation of Google Translate, the translation of Gharib [5] and the extent and domain of errors according to pre-determined factors obtained in the main phase of this study the results revealed that most of the grammatical errors occurred in translation of aspects, passives, compounds, respectively, and Google Translate could not translate verb tense appropriately into Persian. Moreover, Google rarely made ordering and addition errors. Most of the errors of shift (e.g. deletion and substitution) occurred when it deleted something true and substituted something false to translation. On the other hand, lexical and discourse errors were the other errors made by Google translate. Lexical errors occurred when Google translated verbs wrongly or did not translate them and discourse errors occurred when Google could not recognize the meaning of the verb in its context. Table 1 presents the percentage of extent and domain of errors occurred by Google translate.

Table 1. The percentage of extent and domain of errors

<table>
<thead>
<tr>
<th>Extent of Errors</th>
<th>Percent</th>
<th>Domain of Errors</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wrong</td>
<td>35.9</td>
<td>Discourse</td>
<td>5.9</td>
</tr>
<tr>
<td>Don’t</td>
<td>10.4</td>
<td>Lexical</td>
<td>35</td>
</tr>
<tr>
<td>Shift</td>
<td>53.7</td>
<td>Grammar</td>
<td>59.1</td>
</tr>
</tbody>
</table>

5. Discussion and Conclusions

5.1 Discussion

The finding provides evidence that Google Translate is not able to translate verb tense from English to Persian. In addition, most of the errors of shift are related to deletion and substitution respectively. On the other hand, lexical errors are the other errors made by Google Translate, that is, the treasury of its literary terms and other areas are poor. Altogether, the findings of this study is in agreement with Komeili et al. [11] research findings, which was conducted on the evaluation of the translations of three softwares named, Pars, Padideh, and Google. The only difference is that the present study examined verb tense in a literary text, specified errors according to Lennon's category and used the second type of Hutchins's [9] category for the evaluation of machine translation, but Komeili et al. [11] investigated conjunction words, syntax-meaning and verb tense in simple, combined, and complicated sentences and idioms.

Thanks to the advancement of the Internet over recent years, there is now a mutation towards using Google Translate just to get an idea of the content of pages written in another language. In short, a text that can be translated by a computer must be written in a way that the computer can understand: there must be no ambiguity, and it must contain only terms contained in the computer's dictionary and which always have the same meaning. More important than that, the resulted translation should not be used in academic purposes, it should be just used for personal knowledge.

It is believed that MT can be useful for particular types of technical documentation. However, the efficiency of MT, be it Google Translate or any other expensive program, is determined by the quality and the volume of the specialized dictionaries that the program comprises. Consequently, MT is a problem far from being solved. Experts in the field agree that computers do not translate like people. However, on some texts, particularly highly technical texts treating a very narrow topic in a rather dry and monotonous style, computers sometimes do quite well. Nevertheless, with other texts, that are more general and more interesting to humans, computers are very likely to produce atrocious results. Professional human translators, on the other hand, can produce good translations of many kinds of texts. People can handle a range of text types that computers cannot. Unfortunately, the experts did not find an answer to the question of why computers are so limited in their ability to translate.
5.2 Conclusions

The purpose of the current study was to determine the efficacy of Google Translate in translating verb tense in a literary text from English to Persian. In this respect, Lennon’s (as cited in Brown, [2]) category for description of errors and second type of Hutchins’s [9] category for evaluation of machine translation was implemented. Important conclusions drawn from this work include: 1- Google Translate is not efficient in translating verb tense from English to Persian. 2- Most of the grammatical errors belong to aspects, passive, compounds, respectively. 3- Ordering and addition errors are the least errors made by GoogleTranslate while deletion and substitution errors are the most ones. 4 - Google Translate makes lexical errors too. Therefore, it can be said that utilizing highly developed and specified meshwork of lexicon in general terms and technical areas would decrease lexical errors in Google Translate. On the other hand, discourse errors occurred also by Google Translate. This is the case that any researches have‘nt taken it into consideration so far. In this regard, Critical Discourse Analysis (CDA) would probably solve this problem and improve the quality of MTs in future.

Basically, MTs in many cases, are helpful to some extent, but they are not to be trusted without the intervention of the human mind. All in all, one can only get a grasp of what the genuine says, as it is obvious that translations have been made mechanically and they need much improvement in order to be completely understandable.

However, MT technology is advancing all the time. Many programs are running across the world now and it establishes an exciting area of translation research, especially when unified with the human touch. It is likely that, over time, this research will gradually broaden the boundaries within which the MT can operate or within its evolution.

As I see, machine translators are time conserving, but they are unprofitable without the human intervention. They can’t custody the milieu of the source text, because they can’t think, but they distribute information that has been put into them by humans.

In addition, I would say that using the facilities that the computer and the internet are supplying is an art too, requiring patience and mind-openness in order to acquire the best solution the computer can provide one with.

5.3 Recommendations for Further Work

It is recommended that further research be undertaken in the following areas: Further investigations are needed to examinin the efficiency of Google translate in translation of relative clauses and subordinate clauses; word order; conjunctions; collocation; complex sentences, and lexical cohesion. And last but not least, taking into account Critical Discourse Analysis (CDA) would probably improve the quality of Google translate too.

It is hoped that this study opens MT systems developers’ perspectives on MT translation in Iran. It should be mentioned that the present research is definitely open to further discussions and there could still be some problems to be solved out of the author’s restricted knowledge and materials. Thus, this study can be considered as an impetus for further research on the area of MT. I sincerely hope this research will evoke more attention and study on MT in Iran.

References

