

A Comparative Study of *Google Translate* Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations

Hadis Ghasemi¹ & Mahmood Hashemian¹

¹ Department of Foreign Languages, Isfahan (Khorasgan) Branch, Islamic Azad University, Isfahan, Iran

Correspondence: Mahmood Hashemian, Islamic Azad University, Isfahan (Khorasgan) Branch, Isfahan, Iran. Tel: 98-913-266-3269. E-mail: m72h@hotmail.com

Received: December 27, 2015 Accepted: January 30, 2016 Online Published: February 1, 2016

doi: 10.5539/elt.v9n3p13

URL: <http://dx.doi.org/10.5539/elt.v9n3p13>

Abstract

Both lack of time and the need to translate texts for numerous reasons brought about an increase in studying machine translation with a history spanning over 65 years. During the last decades, Google Translate, as a statistical machine translation (SMT), was in the center of attention for supporting 90 languages. Although there are many studies on Google Translate, few researchers have considered Persian-English translation pairs. This study used Keshavarz's (1999) model of error analysis to carry out a comparison study between the raw English-Persian translations and Persian-English translations from Google Translate. Based on the criteria presented in the model, 100 systematically selected sentences from an interpreter app called Motarjem Hamrah were translated by Google Translate and then evaluated and brought in different tables. Results of analyzing and tabulating the frequencies of the errors together with conducting a chi-square test showed no significant differences between the qualities of Google Translate from English to Persian and Persian to English. In addition, lexicosemantic and active/passive voice errors were the most and least frequent errors, respectively. Directions for future research are recognized in the paper for the improvements of the system.

Keywords: machine translation, Google Translate, evaluation of machine translation, error analysis

1. Introduction

The history of research and applications in the field of machine translation shows a variety of machine translations which they have been the subject of much research of machine translation quality assessment, such as example-based, open-source, pragmatic-based, rule-based, and statistical machine translation (e.g., Elliot, 2006; Sin-wai, 1988). Among the abovementioned machine translations, great effort has been devoted to the study of Google Translate, the most famous applicable machine translation, in recent years (Aziz, Sousa, & Specia, 2012; Karami, 2014; Komeili, Farughi, & Rahimi, 2011).

Corder (1974) was the first who studied error analysis and defined language transfer as the main process in L1/L2 language learning in the 1960s. Keshavarz (1999) defined error analysis as collecting samples, identifying errors, classifying, and evaluating them. He also categorized errors and put wrong use of prepositions, articles, plural morphemes, qualifier and intensifier, and the use of typical Persian construction in English in one group as syntactical-morphological errors, and cross association and language switch into lexical-semantic errors.

Keshavarz (1999) linguistically divided errors into four major groups as (a) orthographic errors, (b) phonological errors, (c) lexicosemantic errors, and (d) morphological-syntactic errors. In recent years, research on machine translation evaluation has become very popular and some experts have been interested in using error analysis to assess machine translations (Eftekhar & Nouraei, 2013; Koponen, 2010; Stymne, 2011).

Omidipour (2014) followed Keshavarz and Corder's models of error analysis to assess writings of adult Persian learners of English. In the paper, he showed that errors in foreign language learning can be seen as a natural phenomenon and also the crucial role of L1 is inevitable. For learners, error analysis is important as it shows the areas of difficulty in their writing.

Google Translate is a provided service to translate different written texts from one language to another and it provides translating 90 languages. It can translate not only a word, but also a phrase, a section of a text, or a Web page. To translate a text, Google Translate search different documentaries to find the best appropriate translation pattern between translated texts by human. This pattern searching is called SMT. Consequently, the quality of

Google Translate depends on the number of human translated texts searched by Google Translate (Karami, 2014).

Google Translate was first based on a rule-based machine translation. After that, it then followed an SMT utilizing statistical model to determine the translation of a word in 2006. SMT uses a bilingual text corpora which is a database of the sentences in both source language and target language. A large group of sentences translated from for example English to Persian will be provided for the machine to calculate the probability of the words. If for instance a word like X has probability 75% to be translated into Y, then it will choose Y as the translation of X.

Karami (2014) discussed different models used in Google Translate. He focused on two major engines used by Google Translate and tried to assess advantages and disadvantages of each one separately. He concluded that rule-based models are easier and efficient for machine translations translating languages which are simple in their linguistics and rules. He believes for a machine translation like Google Translate which supports 90 languages and gets advantage of statistical models the quality of translated texts is due to data provided for the machine and the pair of languages applied in translation process.

Google Translate has been evaluated by many researchers and, compared to other Persian-English machine translation systems to date, and has shown how well this system translates from Persian to English (Mohaghegh & Sarrafzadeh, 2009; Mohaghegh, Sarrafzadeh, & Moir, 2010, 2011).

Aiken and Balan (2011) did a research for the first time and assessed the translation quality of Google Translate considering 50 different languages, not just a pair of languages. At the end of the study, they pointed out that Google Translate translates a European language into another European language much better than those pairs of languages which evolve Asian languages.

Recently, another assessment to the study of Google Translate has been proposed by Bozorgian and Azadmanesh (2015). In case of subject-verb agreement, they considered both Google Translate and human translators and finally they concluded that Google Translate does not handle subject-verb agreement very well while translating English sentences into Persian compared to human translators.

Not only are the scores from automatic machine translation metrics not sufficient and clear to define machine translation quality, but also they are approximate and uncertain. Therefore, they fail in providing enough insight for error analysis (Callison-Burch, Osborne, & Koehn, 2006). To solve this issue, many researchers have proposed various methods of human assessment such as (a) adequacy and fluency scores, (b) postediting measures, (c) task-based evaluations, (d) human ranking of translations at the sentence-level, and (e) error analysis to perfect automatic metrics. There may be few studies considering Google Translate subject of English-Persian pair. This study used error analysis as human assessment to give more information on the errors and help the experts interested in improving Google Translate from the point of English-Persian pair of translations.

2. Method

2.1 Materials

In this study, a descriptive-comparative human analysis of translations by means of Keshavarz's (1999) model of error analysis was done; the material was only Google Translate which is the most popular worldwide machine translation in all around the world provided by Google. Google Translate calculates probability word distribution statistic from bilingual text corpus. If the probability of a word to be translated into a specific word in target language is about 80%, then machine translation confidently uses that translation for sure.

2.2 Procedure

Initially, following Keshavarz's (1999) model, 50 sentences in English and 50 in Persian were systematically chosen from reference sentences of Motarjem Hamrah and then two profiles of their translation by Google Translate from Persian to English (TT1) and from English to Persian (TT2) were obtained. Based on Keshavarz's model, the profiles of the translated sentences were, then, analyzed and organized in different tables as lexicosemantic errors, wrong uses of tenses, wrong word order, errors in the distribution and use of verb groups, wrong use of prepositions, wrong use of active and passive voice, and errors in the use of articles. Explanation parts were set in each table to compare and explain the difference between TT1 and TT2. The frequency of occurrences of all sources of errors was calculated. To establish inter rater reliability, a colleague was asked to study and analyze the same extracted data with the same theoretical framework.

2.3 Data Analysis

In order to analyze the collected data, first the frequencies of English to Persian translation errors and Persian to English translation errors of different types were tabulated and compared. Then, the frequencies of correct and incorrect translated tokens of the different types of translation errors (e.g., lexicosemantic errors, tense errors, wrong use of preposition, word order errors, errors in the distribution and use of verb groups, and errors related to active and passive voice) were juxtaposed in separate tables. Subsequently, the frequencies of different types of errors produced by Google Translate were counted, tabulated, and compared.

To find out whether there was any difference between English-to-Persian and Persian-to-English translation errors by Google Translate, the total frequencies of errors of each type produced by Google Translate for English-to-Persian and Persian-to-English translation errors were put in a table. Then, a chi-square test was run.

3. Results

The present study, thus, made use of a quantitative design to investigate the difference between the quality of Google Translate from Persian to English and English to Persian considering Keshavarz's (1999) error analysis framework. Accordingly, the following six types of errors as per Table 1 were identified, counted, and categorized.

Table 1. Frequencies of English-to-Persian and Persian-to-English translation errors by Google Translate

Error Types	Google Translate, English to Persian	Google Translate, Persian to English	Total
Lexicosemantic	42	26	68
Tense	17	8	25
Preposition	15	5	20
Word Order	31	5	36
Distribution and Use of Verb Group	18	9	27
Active and Passive Voice	2	2	4
Total	125	55	180

Among errors identified in Google translated sentences from English to Persian, as it could be seen in Table 1, lexicosemantic errors had the highest frequency ($f = 42$), while wrong use of active and passive voice had the lowest frequency ($f = 2$). Error types which stood in the middle were wrong use of word order ($f = 31$), wrong distribution and use of verb group ($f = 18$), errors relating to verb tense ($f = 17$), and wrong use of prepositions ($f = 15$). The total number of identified errors amounted to 125.

Focusing on the second column of the above Table specified to Persian to English translations, lexicosemantic errors had the highest frequency ($f = 26$), whereas wrong use of active and passive voice had the lowest frequency ($f = 2$). Error types which were in the middle were wrong distribution and use of verb group ($f = 9$), errors relating to verb tense ($f = 8$), word order errors ($f = 5$), and wrong use of prepositions ($f = 5$). The total number of identified errors was 55 errors.

The direction of translation might affect the quality of the translations rendered by the Google Translate since the frequencies of errors of each type were mostly different in English-to-Persian renderings from the time the translation was done from Persian to English. These differences in frequencies are displayed in the above third column.

Obviously, It could be seen that the frequencies of active and passive voice errors ($f = 2$) were the same. All other frequencies were different. To be more precise, a chi-square test was conducted to capture the possible differences between Google Translate outputs from Persian to English and English to Persian with respect to the each type of translation errors identified based on Keshavarz's (1999) error analysis framework. To figure out whether the differences between English-to-Persian and Persian-to-English translations done by Google Translate were of statistical significance or not, one should cast a look at Table 2.

Table 2. Chi-Square results for comparing English-to-Persian and Persian-to-English translation errors by Google Translate

		Value	df	Sig. (2-tailed)
Pearson	Chi-Square	7.72	5	.172
Likelihood	Ratio	8.34	5	.138
Linear-by-Linear	Association	1.72	1	.189
N of Valid Cases		180		

Because the p value under *Sig. (2-tailed)* column in Table 2 was shown to be greater than the significance level (i.e., $.172 > .05$), it could be inferred that the difference between the frequencies of different types of English-to-Persian and Persian-to-English errors did not reach statistical significance, the conclusion being that the direction of translation did not affect the quality of translation of Google Translate.

4. Discussion

This paper is a modest contribution to the ongoing discussions about the quality of Google Translate as a machine translation. We concentrated not only on English to Persian translations done by Google Translate but also on Persian to English Translations. Google Translate has been evaluated by many researchers and compared to other Persian-English machine translation systems to indicate and show how well this system translates from Persian to English or vice versa (Mohaghegh, & Sarrafzadeh, 2009; Mohaghegh, Sarrafzadeh, & Moir, 2010, 2011); however, there might not be any study done like error analysis as human assessment to provide enough insight for errors and clearly show different types of errors made by Google Translate. This might be the first study to assess the quality of Google Translate considering error analysis method presented by Keshavarz (1999). An important implication of these findings is that the direction of translation did not affect the quality of the translations rendered by the Google Translate which might have been many translators question if the direction of translation is significant in using Google Translate. As seen in Table 2, considering the direction of translations rendered by the Google Translate, the p value was shown to be greater than the significance level (i.e., $.203 > .05$) that means the difference between the frequencies of different types of English-to-Persian and Persian-to-English errors did not reach statistical significance, the conclusion being that the direction of translation did not affect the quality of translation of machine translations. The frequencies of different types of errors probably highlight to what extent translators are required to be more cautious about types of errors while using Google Translate to aid them in accelerating translation. The analysis does not enable us to determine if different errors have the same frequency in all types of texts translated by Google Translate considering the simple conversational sentences used in this study from Motarjem Hamrah.

5. Conclusion

The main concern of the paper was to compare the quality of Google Translate considering the direction of translation and providing enough insight for the errors made by Google. Summing up the results, it can be concluded that the difference between the frequencies of different types of English-to-Persian and Persian-to-English errors did not reach statistical significance; therefore, the direction of translation did not affect the quality of translation of machine translations. The single most important consideration in the quality of Google Translate was to help users decide if the Google Translate will best suit their needs and if they can trust on its translated outcomes.

From the research that has been undertaken based on Keshavarz's model (1999) of error analysis, types of errors and their frequencies were identified to accomplish automatic metrics evaluations with the purpose of improving the systems.

Machine translations, as aids to human translation besides the vast development of technology in using computers, have brought machine translation evaluation into consideration. The quality investigation of Google Translate as a machine translation system and the analysis of its weaknesses were to light a number of ideas to improve future made softwares and help users to adjust their expectations and have better understanding. The findings are of direct practical relevance.

Additionally, machine translation is an unknown field of study in Iran and needs a lot of efforts to be investigated. This study, beside other research done in Iran, may help experts to write better computer programs. The revealed errors by this study may inform the developers and project managers to perceive the strengths and weaknesses of

the Google Translate. Consequently, the use of this study as a source of possible errors may bring up a new machine translation in future.

References

- Aiken, M., & Balan, Sh. (2011). An analysis of Google Translate accuracy. *Translation Journal*, 16(2). Retrieved June 26, 2015, from <http://translationjournal.net/journal/56google.htm>
- Aziz, W., Sousa, S. C. M., & Specia, L. (2012). PET: A tool for postediting and assessing machine translation. In *The Eighth International Conference on Language Resources and Evaluation*, Istanbul, Turkey.
- Bozorgian, M., & Azadmanesh, N. (2015). A survey on the subject-verb agreement in Google machine translation. *International Journal of Research Studies in Educational Technology*, 4(1), 51-62. <http://dx.doi.org/10.5861/ijrset.2015.945>
- Callison-Burch, Ch., Osborne, M., & Koehn, Ph. (2006). Re-evaluating the role of bleu in machine translation. In *Proceedings the 11th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 249-256). Trento, Italy.
- Corder, S. P. (1974). *Error analysis: Perspectives on second language acquisition*. London: Longman.
- Eftekhari, M., & Nouraey, P. (2013). Commercial translation error analysis: A case study of Iranian products. *Linguistics and Literature Studies*, 1(2), 55-60.
- Elliot, D. (2006). *Corpus-based machine translation evaluation via automated error detection in output texts*. Unpublished doctoral dissertation, University of Leeds.
- Karami, O. (2014, January). *The brief view on Google Translate machine*. Paper presented at the meeting of the 2014 Seminar in Artificial Intelligence on Natural Language, German.
- Keshavarz, M. H. (1999). *Contrastive analysis and error analysis*. Tehran: Rahnama Publication.
- Komeili, Z., Farughi, J., & Rahimi, A. (2011). An investigation of the translation problems incurred by English-to-Persian machine translations: Padideh, Pars, and Google softwares. *Science Direct*, 28, 1079-1082. <http://dx.doi.org/10.1016/j.sbspro.2011.11.195>
- Koponen, M. (2010). *Assessing machine translation quality with error analysis*. Retrieved June 26, 2015, from https://sktl-fi.directo.fi/@Bin/40701/Koponen_MikaEL2010.pdf
- Mohaghegh, M., & Sarrafzadeh, A. (2009). An analysis of the effect of training data variation in English-Persian statistical machine translation. In *Proceedings of the Innovations in Information Technology, 2009* (pp. 105-109). Al Ain. <http://dx.doi.org/10.1109/IIT.2009.5413782>
- Mohaghegh, M., Sarrafzadeh, A., & Moir, T. (2010). Improved language modeling for English-Persian statistical machine translation. In *Proceedings of the SSST-4, Fourth Workshop on Syntax and Structure in Statistical Translation* (pp. 75-82). Coling, Beijing.
- Mohaghegh, M., Sarrafzadeh, A., & Moir, T. (2011). Improving Persian-English statistical machine translation: Experiments in domain adaptation. In *Proceedings of the 5th International Joint Conference on Natural Language Processing* (pp. 9-15), Thailand.
- Omidipour, M. (2014). An Analysis of errors in writing among adult Persian learners of English. *International Journal of Language Learning and Applied Linguistics World (IJLLALW)*, 5(3), 176-187.
- Sin-wai, Ch. (1988). *The Routledge encyclopedia of translation technology*. New York: Routledge.
- Stymne, S. (2011). Blast: A tool for error analysis of machine translation output. In *Proceedings of ACL* (pp. 56-61). Portland, Oregon, USA.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).