Exploring Linguistic Modifications of Machine-Translated Literary Articles: The Case of Google Translate

Alireza Bonyadi

English Department, Islamic Azad University, Urmia Branch, Urmia, Iran
a.bonyadi@iaurmia.ac.ir

Abstract

Google Translate, a free multilingual machine translation service, developed by Google has attracted the attention of countless users due to its ease of use through modern means of mass communication, and has become the only translation tool in some areas. However, compared to human translation, these machine tools have not yet been able to deliver high-quality translations due to the complexity of translation process. Therefore, studying the modifications of machine translated texts is of great importance. Therefore, the current study aimed to explore the types of linguistic modifications of the texts translated from Persian into English through *Google Translate*. To this end, the abstracts of ten unpublished Persian literary articles intended to be submitted to Iranian journals were selected for the analysis. The selected abstracts were initially translated into English (target language) through *Google Translate* from Persian (source language). To identify the kinds of changes needed to make them academically acceptable, the machine translated texts were all post edited. Then, the original Google translated texts and their post edited versions were compared to figure out the types of the applied modifications. The results of this qualitative study indicated that the linguistic post edition modification of the texts included tense, literal translation, redundancy, collocations, deletion of the main verb, word-choice and proper nouns.

Keywords: Machine Translation, Literary Texts, Post-Editing, Human Translation, Linguistic Modifications

Introduction

Google Translate services were launched in 2006 with the aim of breaking language barriers and allowing for more general access to information. Since then, the number of languages supported by this software has increased from two languages into more than a hundred languages. (Thoravsky, 2006). This service makes use of a machine translation system for translation, which is a combination of linguistic modeling, statistical decision theory and matching probabilities (Ney, 1995; Brown, Cocke, Pietra, Jelinek, Lafferty, Mercer & Rossin, 1990). Simply put, the software can quickly detect dominant text patterns and provide the closest equivalent for the text by analyzing, evaluating and analyzing millions of documents and translated texts in cyberspace.

Since this machine translation is available on the internet for free and is capable of being installed on computers, tablets and smartphones, users can enjoy the services of this technology easily everywhere via an internet connection and translate various texts from one language to another. Google Translate is said to be much more popular than any other machine translation services. (Seljan, Brkić, & Kučiš, 2011).

The availability and the user-friendly nature of this technology has made it more attractive to students, teachers and researchers in educational settings. *Google Translate* service is also known as a source of independent foreign language learning. Researchers and language practitioners, of course, have different opinions about the role of machine translation in language teaching. It is believed that machine translation is usually providing the users with linguistically odd pieces of discourse which in long run might affect students' language learning. (Anderson,1995, Richmond,1994). According to Groves & Mundt, (2015) machine translation can compensate for the lack of linguistic knowledge of users in target language. These people, then, consider machine translation texts as the best possible translation. Machine translation can also be employed in translation classes making them enter into the online translation market, where they can electronically offer their translation services (Vaezian, 2010). Shei, (2002) believes that machine translation can be used as an independent source for double checking the lexical and grammatical features of non-machine texts.

Comparing the role of machine translation in teaching translation with that of a calculator in teaching mathematics, Groves & Mundt (2015) believe that the use of the calculator did not question teaching the basic concepts of the field traditionally, but accelerated it. In the same vein, machine translation can be

considered as a tool for translators, so that they can enhance the accuracy and the speed of their translation (Hutchins, 2001).

Of course, Google has itself asserted that, due to the linguistic complexity of texts, machine translation is not expected to offer the same quality and accuracy as a professional human translator might do (Huddleston, 2013). Therefore, editing and re-reviewing machine translation texts by users is essential. The term "post-translate editing" refers to the process of reviewing and editing texts already translated by a translation machine. In fact, through post editing of machine translated texts, the speed of machine translation is coupled with the precision of the professional human translators resulting in an acceptable rendering.

It is worth mentioning that the quality of machine translation also depends on the affinity between the two languages involved. In other words, the more linguistic similarities exist between the two languages, the better the machine translation quality would be. For example, translating from French to English, as well as from Italian to English, will be much more accurate than Persian to English.

The comparison of machine translation with human translation can be carried out at different linguistic levels, such as vocabulary, syntax, semantics, and pragmatics. Of course, the extent of the edition depends on the genre of the texts as well. For example, scientific texts enjoy less lexical variation than literary texts. In other words, the number of polysemous lexical items in literary texts is much greater than the other texts requiring thus a greater amount of editing.

Literature Review

The widespread use of machine translation has led some scholars to study the quality and quantity of machine translation texts from a variety of perspectives. To identify machine translation errors (from Chinese to English), Chang (2008) translated the first two paragraphs of thirty web sites into English using machine translation. Based on the results of the study, the texts with simple grammatical structures were found to enjoy a better translation quality. The most commonly encountered errors in machine translation were found to be proper nouns, abbreviations, segmentation and omission.

Vaezian (2010) investigated the students' perceptions of the use of machine translations in English language classes. She initially asked these students to edit some political texts, previously translated from English to Persian by machine translation (*Google Translate*). Based on the results of this study, 90% of the students reported that they would use machine translation in the future. However, the main editing problem indicated by the students was the grammatical structure

of Persian sentences. Of course, the study has not provided further details on the type of structures mentioned.

Al-Shabab (2013) studied the translation of legal texts from English to Arabic. To this end, he selected six English legal papers. The selected papers were translated into Arabic by his two colleagues. The same papers were then translated to Arabic this time through using *Google Translate* service. By evaluating and comparing these two translations, the study concluded that machine translation is not very successful in translating sentences bearing passive structures and auxiliary verbs as the resulting translated sentences suffered from semantic ambiguity. Of course, this study had not specified the kind of the semantic ambiguity.

To investigate the mistakes of machine translation, (Groves & Mundt, 2015) asked their students to write an article in their native language and then translate it through one of the translation software into English. By reviewing the machine translation texts, the researchers concluded that machine translation currently cannot provide acceptable translation in terms of observing grammatical rules of the language. However, they claimed that machine translation has had a profound effect on the field of English for Special Purposes (ESP).

In general, studies on machine translation indicate that this type of translation cannot yet be considered a true quality translation. Therefore, the editing and modifying the machine-translated texts is inevitable. Since the studies on machine translation of Persian language texts are very limited, the present paper examines the quality of the machine-translation of these texts. In particular, the study aims at addressing the following research question.

What kinds of translation errors are commonly encountered in the English translation of Persian literary texts using *Google Translate*?

Research Methodology

A qualitative research method was used to analyze the quality of translation of the texts. To this end, the Persian abstracts of ten papers intended to be submitted to Iranian quality journals on literary studies were selected for the analysis. Selected abstracts – sentence by sentence - have been translated into English (target language) from Persian (source language) through *Google Translate*. The reason for this sentence by sentence feeding is the fact that the degree of precision in rendering a short sentence, through *Google Translate*, is reportedly much higher than the other larger linguistic units, such as compound sentences and paragraphs (2009, Niño). Then the google translated texts were post edited by the researcher himself. Employing a bottom-up approach, commonly used in qualitative studies (Yuan, 2015, Macky& Gass, 2005), the researcher analyzed

both machine translated texts and post edited texts to code the data and identify the emerging patterns.

Results

Theme 1: Grammatical tense

Machine translation in general cannot be trusted considering grammatical rules, semantic elements, and pragmatic issues. (Li, Graesser, & Cai ,2014). Thus, it is not unexpected to come up with equivalents that are not acceptable with reference to grammatical tenses. In our data, regardless of word choices, Persian present perfect tenses have been translated as either English simple present or simple past tenses. Table 1 illustrates the point.

Table 1
An instance of rendering grammatical tense

Source text:
شاعران کودک تا جای ممکن، میان این شگردِ شعری و دنیای کودکان، نسبت ایجاد کرده ...
Target language (Machine translation):
*The poetry of the child, as much as possible, <u>created</u> a balance between this poetic device and child's world…
Target language (Post-edited):
Child poets <u>have established</u>, as far as possible, a balance between this poetic device and child's world …

Theme 2: Literal translation

Literal translation is one of the main weaknesses of online machine translation systems, which we usually encounter as an editor. The main reason for this has something to do with the *system* of the machine translation tool which has been programmed to keep the basic structure of the texts. This usually results in giving anomalous equivalents for lexical items in source language making the texts look less cohesive in target language.

Table 2

An instance of literal translation

Source text:
شعر مقاومت و پایداری

Target language (Machine translation):

*Poetry of <u>resistance or sustainability</u>

Target language (Post-edited):

<u>Resistance poetry</u>

As is evident in the table it is necessary to check machine translation texts for spotting any literal translation in target language. In fact, as Newmark (1998) has put it "if a perfectly natural SL unit produces a clumsy literal translation, … then the translation is 'wrong', however expressive the rest of the SL text" (p, 75).

Theme 3: Redundancy

Redundancy is a common linguistic phenomenon in all languages. It is mainly employed in discourse pragmatically to enhance comprehensibility, resolve ambiguity, focus on an isolated feature, compare elements, intensify a feature and create a 'poetic' effect (Wit & Gillette, 1999). Redundancy, in the same vein, has been used in Persian literary works as well. However, as the machine translation usually keeps the basic structure of sentences and phrases of the source language, in most cases, the redundant items, translated into source language look so odd.

Table 3

An instance of redundancy

Source text:
چگونگی تغییر وضعیت و جا به جایی زمانی ..
Target language (Machine translation): *<u>The way of changing the situation</u> and <u>the time span</u> …
Target language (Post-edited): <u>The mechanism of changing</u> the situation and also the <u>chronological shift</u> ...

To edit such cases, deleting the redundant item would result in natural and much more acceptable phrases in target language.

Theme 4: Collocations

Integrating collocations in a piece of discourse is one of the factors contributing to the textual coherence of a text, although their distribution and frequency is less than that of other coherence factors such as repetition, semantic and contradiction. (Hamedi & Mirshahi, 2016) Translators' awareness of the importance of collocations and their role in enhancing text coherence is of importance. However, it should be noted that there is not a one-to one correspondence between the textual devices in two languages. In other words, collocations, like any other textual devices, cannot be expected to be realized in the same form in source and target languages. Thus, as Dastjeridi & Taghizadeh (2005) commented, the editors' awareness of lack of such correspondence would help them provide appropriate equivalences for source language collocations in target language.

Table 4
An instance of collocation

Source text:	
	وصف و بیان توصیفی به عنوان عنصری زیباشناختی و شگردی موثر در شعر کودک، کاربرد دارد.
Target language (Machine translation):	
*Descriptive description and description as an aesthetic element and effective guidance in child poetry.	
Target language (Post-edited):	
Descriptive narration as an aesthetic element and as an effective device <u>is applicable</u> to child poetry.	

As evident in Table 4, the provided equivalence for Persian collocational phrases are totally unacceptable in Persian language thus demanding editors' further intervening. For Lotfipour (2003) lack of correspondence between lexical elements in the source language with their seemingly equivalent elements in the target language, in terms of collocation, poses many problems demanding translators' attention. Therefore, it is important for the editors to review the equivalents for the source language collocations provided by the machine translation.

Theme 5: Verb omission

Verb omission in target language is one of the problems noticed in some of the sentences rendered through machine translation. Obviously, the editors' familiarity with target language grammar and also the basics of translation can help him diagnose and edit such missing points.

Table 5
An instance of verb omission

Source text:
مفتون امینی
Target language (Machine translation):   *Charmed Amini
Target language (Post-edited):   Maftoon Amini

As illustrated in Table 5, the main verb of the sentence in source language has been neglected in the target language due to machine text processing mechanism. As verb is one of the core elements of a sentence that other syntactic and semantic information are loaded upon, the omission or even incorrect translation of these elements would affect the processing and understanding of the whole sentences in the target language. (Ragheebdoost & Mehrabi, 2010)

Theme 6: Word choice
Choosing a proper lexical item in the target language is one of the main challenges in the translation. According to Landauer (2002) over 80% of potential text information is carried out through the choice of the words. This suggests the importance of the right choice of the words in the crystallization of the meaning of a text. As machine translation tools are generally processing the texts literally, post-editors would most probably face with cases of contextually wrong word choices.

Table 6
An instance of word choice

Source text:
بیان روایی
Target language (Machine translation):   *narrative narrativity*
Target language (Post-edited):   narrative style

Obviously the two lexical items namely "feature" and "component" can be considered as potential equivalents for the word   in certain contexts interchangeably. However, based on the specific context, as evident in Table 6, the word "feature" is much more suitable equivalent for the word "مولفه" than its semantically related word "component".

Theme 7: Proper nouns

Since machine translation refers to word-bank and pre-translated grammatical structures for translating the linguistic items of the source language, proper names and coned words are generally left unrecognized or translated in an unusual way.

Table 7
An instance of proper nouns

Source text:
مولفه های شعر او
Target language (Machine translation):   **components of his poetry*
Target language (Post-edited):   *features of his poetry*

As evident in Table 7, machine translation cannot provide proper equivalent for the source language phrase مفتون امینی due to the novelty of the names and phrases mentioned. In fact, the given equivalent, namely *"Charmed Amini"* is obviously erroneous and thus does not make any sense in the target language as it has been treated literally.

Discussion and Conclusion

Although *Google Translate* service can translate some sentences and structures of Persian language correctly into English, it is still unable to provide proper translation; the service can be used for rendering lexical and short grammatical structures. In general, one of the main reasons why machine translation has not yet been able to provide trusted translations is in fact due to its inability to recognize the textual context of the discourse. Therefore, as long as such capability is not achieved, machine translation should not be considered as a quality translation. In most cases users opt for the software just for saving time and cost. From this perspective, post editing of Google translated texts is of great importance.

The qualitative study of the Persian literary texts translated into English by *Google Translate* service indicated that the major errors in machine translation of the texts of the present study can be classified into certain categories, namely grammatical tense (14 items), literal translation (11 items), redundancy (8 items), collocations (9 items), deletion of the main verb (3 items), word-choice (18 items), and innovative and proper nouns (8 items). These results seem to be in line with the findings of other studies such as Vaezian (2010), Groves, & Mundt. (2015), Chang-Meadows (2008), all emphasizing the post editing of the machine

translated texts.

Given the increasing growth of the use of various translation services, the increased awareness of the beginner users on possible machine translation errors can enhance the accuracy and speed of their translation. Meanwhile, the identified common errors can also be emphasized in translator training courses. Considering the limitations of the present study, namely limited set of data and focusing on translation from Persian to English, the generalizability of the results is limited as well. Thus, it is expected that future researchers will use other quantitative and qualitative methods exploring the issue of machine translation errors in a wide variety of languages.

Meanwhile, with the rapid advancement of technology, Google translate service is expected to boost its capabilities in future through increasing its databases. On the other hand, the ease of access to modern communication devices will make the users more likely to use this software in various educational and non-educational settings. It is hoped that researchers in the field would seek to present a complementary software for automatic detecting and checking of machine translation errors.
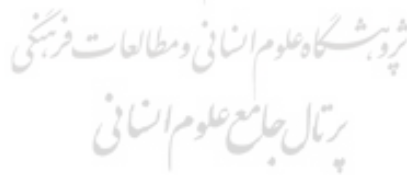
References

Al Shabab, M. (2013). The translatability of English legal sentences into Arabic by using Google translation. *International Journal of English Language and Linguistics Research*, *1*(3), 18-31.

Anderson, D. D. (1995). Machine translation as a tool in second language learning. CALICO, 13(1): 68-97.

Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., Mercer, R. L., & Rossin, P. (1990). A Statistical Approach to Machine Translation. *Computational Linguistics.* 16(2): 76-85.

Chang-Meadows, Shin. (2008). "MT errors in CH-to-EN MT systems: user feedback". In *AMTA-2008. MT at work: Proceedings of the Eighth Conference of the Association for Machine Translation in the Americas*, Waikiki, Hawai.

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, *37*, 112-121.

Hamedi Shirvan, Z., & Mirshahi, M. (2016). بررسی و کاربرد ابزارهای انسجام متن در کتاب [Investigating and applying های سطح پیشرفته از چهار مجموعه کتاب آموزش زبان فارسی. text coherence tools in advanced series of books, quartet, on Persian 516-5 ,مرکز بین المللی آموزش زبان فارسی به غیر فارسی زبانان. language teaching]. 37.

Huddleston, G. (2013, May, 21). Machine Translation (MT): Is post-editing the silver lining? [Blog post]. Retrieved from https://www.languageconnect.net/

Hutchins, J. (2001). Machine translation and human translation: in competition or in complementation. *International Journal of Translation*, *13*(1-2), 5-20.

Landauer, TK. (2002). On the computational basis of learning and cognition: arguments from LSA. In *Ross BH (Ed.) The psychology of learning and motivation*, (41). Academic Press, New York, pp. 43–84.

Li, H., Graesser, A. C., & Cai, Z. (2014, May). Comparison of Google Translation with Human Translation. In *FLAIRS Conference*.

Lotfipour Saedi, K. (2003). اصول و روش ترجمه [Principles of translation] دانشگاه پیام

نور.

Macky, A., & Gass, S. (2005). *Second language research: Methodology and design*. Lawrence Erlbaum Associates Publishers, London.

Ney, H. (1995). On the Probabilistic Interpretation of Neural Network Classifiers and Discriminative Training Criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 17(2): 107-119.

Niño, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL*, *21*(2), 241-258.

Raghibdost, Sh., & Mehrabi, M. (2010). پردازش جمله و بازنمود ذهنی فعل در زبان فارسی[Sentence processing and mental representation of verbs in Persian]. دوفصلنامه علمی–پژوهشی زبان پژوهی دانشگاه الزهرا(س), 1-24. 2.

Richmond, I.M. (1994) Doing it backwards: Using translation software to teach target language grammaticality. In CALL 7 (1): 65-78.

Seljan, S., Brkić, M., & Kučiš, V. (2011). Evaluation of Free Online Machine Translations for Croatian-English and English-Croatian Language Pairs. In *Proceedings of the 3rd International Conference on the Future of Information Sciences: INFuture2011-Information Sciences and e-Society*, 331-345. Zagreb, Croatia.

Shei, C- C (2002). *Combining Translation into the Second Language and Second Language Learning: An Integrated Computational Approach*. (Unpublished doctoral dissertation). University of Edinburgh.UK.

TAUS: Translation Automation User Society. 2006. "Putting Machine Translation to Work". https://www.taus.net/downloads/finish/56-public-reports/16-taus-putting-machinetranslation-to-work-april-2006. Accessed on: 7 January 2019.

Trotsky, B. (2016, April, 28). Ten years of google translate [Blog post]. Retrieved from https://blog.google/products/translate/ten-years-of-google-translate/

Vaezian, H. (2010). Google translator toolkit in translation classroom. (Unpublished Paper). *School of Languages, Literacies and translation.* Universiti Sains Malaysia.

Dastjerdi, H., & Taghizadeh, S. (2005). کاربرد عوامل انسجامی در ترجمه و بررسی مقابله ای [Application of cohesive devices in translation متون فارسی و ترجمه های انگلیسی آنها and comparative analysis of Persian texts and their English translations]. فصلنامه مطالعات ترجمه *12,68-71.*

Wit, E. C., & Gillette, M. (1999). *What is linguistic redundancy*? University of Chicago, USA.

Yuan, R. (2015). *Understanding language teacher educators' identities in Hong Kong*. (Unpublished doctoral thesis). The Chinese University of Hong Kong, Hong Kong.

This page is initially left blank