Archives of Advances
in Biosciences

Original Article:

Translating Medical Texts from Persian to English: Accuracy of Machine Translation

Mitra Saffari[1,2], Mostafa Pourhaji[3*], Saeed Fathi-Alishah[1], Samad Sajjadi[3], Mehrad Mohammadi[4]

1. English Language Department, Islamic Azad University, Parand Branch, Tehran, Iran
2. Iranian Biomedical Journal, Pasteur Institute of Iran, Tehran, Iran
3. English Language Department, School of Allied Medical Sciences, Shahid Beheshti University of Medical Scxiences, Tehran, Iran
4. Imam Khomeini International University, Qazvin, Iran

Cite this article as: Saffari M, Pourhaji M, Fathi-Alishah S, Sajjadi S, Mohammadi M. Translating Medical Texts from Persian to English: Accuracy of Machine Translation. Archives of Advances in Biosciences. 2023; 14:E43067. https://doi.org/10.22037/aab.v14i. 43067

https://journals.sbmu.ac.ir/aab/article/view/43067

Article info:
Received: 16 Sep 2023
Accepted: 30 Sep 2023
Published: 21 Oct 2023

*** Corresponding author:**

Mostafa Pourhaji, PhD.

Address: English Language Department, School of Allied Medical Sciences, Shahid Beheshti University of Medical Scxiences, Tehran, Iran

E-mail: m.pourhaji@sbmu.ac.ir

Abstract

Introduction: Machine translation is a new subject increasingly being used by academic writers, especially students and researchers whose native language is not English. There have been numerous studies conducted on machine translation, but few investigations have assessed the accuracy of machine translation from Persian to English at lexical, semantic, and syntactic levels.

Materials and Methods: Using Groves and Mundt's model of error taxonomy, the current study evaluated Persian-to-English translations produced by two famous online translators, Google Translate and Bing Translator. A total of 60 texts were randomly selected from the academic field of medicine. All texts were rendered by the two translation systems and then by four human translators. statistical analyses were applied using SPSS.

Results: The results indicated that Google translations were more accurate than the translations produced by the Bing Translator in the domain of medicine (lexis: 186 vs. 225; semantic: 44 vs. 48; syntactic: 148 vs. 264 errors).

Conclusion: The findings suggest that students and researchers can reasonably benefit from the systems in rendering plain texts from Persian to English, given that the translated versions are subjected to human editing.

Keywords: Bing Translator, Google Translate, Translation accuracy, Translation efficiency, Online translators

1. Introduction

Thanks to computers and technological developments, using machine translation has become increasingly prevalent among academic writers, especially students and researchers whose mother tongue is not English. At an empirical level, various studies have been carried out on the usefulness of machine translators as an auxiliary tool. Champollion [1] and Lagoudaki [2] reported that machine translation is a helpful tool for preparing a rough draft. Komeili, Farughi, and Rahimi [3] observed that automatic translation tools, to a large extent, can assist researchers when translating a text. Other investigators have suggested that machine translation systems are useful for translating different text types, including legal texts [4] and scientific publications. Despite these optimistic claims, some researchers criticized the effectiveness of machine translation. Madson [5] suggested that, due to a large number of errors, machine translation cannot produce accurate translation. Precup-Stiegelbauer [6] found

that online translators are not effective systems without human intervention.

In recent years, great efforts have been devoted to comparison between different machine translation systems. In a study, Aikan, Ghosh, Wee, and Vanjani [7] evaluated four online translation systems, including Google Translate, Yahoo SYSTRAN, Applied Language, and x10 using German-to-English and Spanish-to-English translations. They reported that Google Translate performed better than other systems. In another survey, a comparison was made among four free online translation services, including Google Translate, Stars21, InterTran, and Translation Guide. The outcome revealed that Google Translate indicated a better performance than the other three translation tools [8].

Many more investigators have assessed the efficiency of different online machine translation systems. Aiken and Balan [9] reported that the accuracy of machine translation in translating European languages was more acceptable than Asian languages. The study by Khosravizadeh and Pashmforoosh [10] on the semantic structure of Google Translate using Persian-English pairs revealed that Google Translate can be a reliable tool with the involvement of a human translator. Dhakar, Sinha, and Pandey [11] evaluated the quality of translations produced by Google Translate and Bing Translator from Hindi to English. Based on their result, Bing Translator had better performance in translating news, technology, medical, and official texts. Khalilizadeh Ganjalikhani's [12] study on the semantic structure of Google Translate demonstrated that Google Translate was unable to translate simple, compound, and complex sentences from Persian to English. A recent previous study by Groves and Mundt [13] investigated the English grammatical accuracy of Google Translate on three essays written in Malay and two essays written in Chinese by university students. Although Google Translate was able to translate grammatically correct sentences from Malay or Chinese to English, it failed to produce overall convincing results.

Based on the literature review and the examples from machine translation investigations mentioned above, there is no comprehensive research to assess the accuracy of machine translation from Persian to English at lexical, semantic, and syntactic levels. In this light, the present study was undertaken to evaluate the efficiency of Google Translate and Bing Translator in translating lexical, semantic, and syntactic features from Persian to English in the domain of medicine. To this end, in an attempt to account for the gap in Persian-to-English translation, the following research question was formulated:

Is Google Translate lexically, semantically, and syntactically more accurate than Bing Translator, when translating academically oriented phrases and sentences from Persian to English?

2. Materials and Methods

Sample selection

A total of 60 Persian texts were selected from different textbooks serving as academic books being now taught at different reputable universities in Iran. The selected texts were obtained from the field of medicine.

Evaluation and scoring procedure

Quality of machine translators was evaluated by two human raters based on Groves and Mundt's Model. Kappa statistics was performed to examine inter-rater reliability and determine consistency between the two raters (k = .89, p < .005). The reason for selecting Groves and Mundt's Model is that it allows for detailed analysis and scoring of the translated materials. The Model was slightly modified to suit the objectives of the study more effectively as follows:

1. Two categories of 'pronoun incorrect' and 'pronoun reference' were reduced to one category as 'pronoun incorrect', for the sake of simplicity.

2. The categories designated as 'verb tense', 'apostrophe', 'sentence structure', 'run on', and 'comma splice' were excluded from the study because they were not in the scope of this study, and the category 'apostrophe' was not applicable to Persian language.

3. A new category, namely 'unclear', was added up to the Model. This item was included to show the lack of clarity in the translation of phrases or sentences produced by machine translators. In fact, it referred to unclear concepts. The errors were categorized into three groups, including 1) lexis, (2) semantics, and (3) syntax. This classification was performed to achieve the objective of the present study; that is, evaluating the accuracy of machine translation lexically, semantically, and syntactically. The lexis itself included subcategories of 'wrong word', 'word choice', and 'missing word'. The semantics category was subcategorized into 'unclear' and 'word order'. The syntax included subcategories of 'word order' and 'unclear' for semantic accuracy, and the items 'verb tense', 'word form', 'article', 'plural', 'agreement', 'preposition', 'comma', 'spelling', 'pronoun incorrect', and 'fragment' for syntactical accuracy.

Procedure

All the texts under study (n = 60) were translated into

English by Google Translate and Bing Translator individually, followed by four expert translators. All translators (with Ph.D. degrees in English language and literature) were native speakers of Persian and had a full command of Translation in both Persian and English. The translations done by the four experts served as the benchmark against which machine translations could be judged. All the translated materials by humans and machines were given to the evaluators to score based on the category described above (i.e., Groves and Mund's Model). To measure the vocabulary and the structure (i.e., semantic) accuracy of the texts translated by Google Translate and Bing Translator, a five-point Likert scaling was used, with 4 indicating excellent translation, 3 very good, 2 average, 1 poor, and 0 very poor translation. All the translations within a score range of 0-2 were considered 'incomprehensible', receiving a negative point. The lexical and syntactic correctness of the machine translations were scored based on the numbers of errors found in the texts. The data were then analyzed and tabulated based on the evaluation and scoring results.

Statistical analysis and results

Chi-square (χ^2) was applied to analyse the data, using SPSS version 21.0. The two tests were used to compare (a) the machine translations with human translation and (b) Google Translate with Bing Translator. The computations were intended to find out whether there were any significant differences in the results. *P* values less than 0.05 were considered statistically significant.

As indicated in the Table 1, the highest error frequency belonged to the *sentences* rendered by Bing 147, respectively). However, the lowest error frequency can be observed in Google- and Bing-translated *very short phrases* (5 vs. 9, respectively). Error frequency in *short phrases* rendered by Bing Translator was almost close to Google Translate (*f* = 10 vs. *f* = 12, respectively); however, in lexical translation of *long phrases*, both systems indicated an equal error frequency (*f* = 22).

The highest percentage of errors, as shown in Table 2, can be observed among *sentences*. The *sentences* translated by Google Translate had 100% error frequency, whereas in Bing-translated *sentences*, there were 96.6%. The lowest percentage of errors belonged to *very short phrases* rendered by Google Translate (30%), followed by Google-translated *long phrases* (50%). In *short phrase*s rendered by the two systems and in Bing-translated *very short phrases*, an equal percentage of error (60%) was observed.

As can be seen in Table 3, the highest error frequency was related to *spelling* (*f* = 29), *word form* (*f* = 19), and comma (*f* = 16) in *sentences* rendered by Bing Translator. However, Google Translate indicated relatively very well in all of these areas. Also, the two systems demonstrated an error-free translation in some areas. Bing Translator did well in the areas of *word form* (in *very short* and *long phrases*), *plural noun* in *very short phrases*, and *article* in *long phases*. However, Google Translate performed very well in the following areas: *word form* (in all *phrases*), *agreement* (in *long phrases*), *plural noun* (in *very short phrase*), and *spelling* (in *very short* and *long phrases*).

Table 1. Frequency and percentage of Persian-to-English lexical errors in Google- and Bing-translated texts

Domain	MT	Noun phrase						Sentence	
		Very short		Short		Long			
		f	%	*f*	%	*f*	%	*f*	%
Medicine	Bing	9	45	10	40	22	21.1	184	29.7
	Google	5	25	12	48	22	21.1	147	23.7
Total of words		20		25		104		619	

MT, machine translation; f, error frequency

Table 2. Frequency and percentage of Persian-to-English semantic errors in Google- and Bing-translated texts

Domain	MT	Noun phrase						Sentence	
		Very short		Short		Long			
	Total of words	*f*	%	*f*	%	*f*	%	*f*	%
		10		10		10		30	
Medicine	Bing	6	60	6	60	7	70	29	96.6
	Google	3	30	6	60	5	50	30	100

Table 3. Frequency and percentage of Persian-to-English syntactic errors in Google- and Bing-translated medical texts

MT	Error type	Noun phrase: Very short *f*	Very short %	Short *f*	Short %	Long *f*	Long %	Sentence *f*	Sentence %	
Bing	VT	-	-	-	-	4	3.8	5	0.8	
	WF	correct	0	3	12	correct	0.0	19	3.0	
	ART	-	-	-	-	correct	0.0	3	0.4	
	PLN	correct	0	1	4	1	0.9	11	1.7	
	AGR	-	-	-	-	2	1.9	5	0.8	
	PREP	-	-	-	-	1	0.9	7	1.1	
	COM	-	-	-	-	-	-	16	2.5	
	SP	3	15	2	8	2	1.9	29	4.6	
	PRO	-	-	-	-	4	3.8	4	0.6	
	FRAG	-	-	-	-	-	-	10	0.1	
Total of errors/words		3/20		6/25		14/104		109/619		132/768
Google	VT	-	-	-	-	1	0.9	2	0.3	
	WF	correct	0	correct	0	correct	0.0	3	0.4	
	ART	-	-	-	-	-	-	3	0.4	
	PLN	correct	0	3	12	2	1.9	4	0.6	
	AGR	-	-	-	-	correct	0.0	3	0.4	
	PREP	-	-	-	-	1	0.9	10	0.1	
	COM	-	-	-	-	1	0.9	12	1.9	
	SP	correct	0	1	4	correct	0.0	4	0.6	
	PRO	-	-	-	-	6	5.7	8	1.2	
	FRAG	-	-	-	-	1	0.9	8	1.2	
Total of errors/words		0/20		4/25		12/104		57/619		75/768

MT, machine translation; *f*, error frequency; Boxes show high-frequency errors

4. Discussion

Google Translate proved to be more accurate than Bing Translator in rendering lexis in the domain of medicine. This finding indicates that the Word database used by Google Translate is richer than that of the Bing Translator. In contrast to our result, Costa-Jussá *et al.* [14] found that lexical translation of Google Translate from Catalan to Spanish in the medical domain was worse. Another study by Dhakar and co-workers [11] demonstrated a contradictory result. They reported that lexical translations of medical texts by Bing Translator from Hindi to English were more accurate than their translations by Google Translate. The reason behind these inconsistent results could be due to dissimilarity in the architecture of languages, and most likely because of technical differences.

Semantically neither Google nor Bing provided proper equivalents for the texts from the domain under study, which is at odds with Kit and Wong's [4] report according to which Google Translate semantically performed well in translating medical texts. Further, Costa-Jussá *et al.* revealed that Google Translate provide a proper translation semantically from Catalan to Spanish in the domain of medicine.

Syntactically, Google Translate had better performance than Bing Translator in translating medical texts. In a previous study, Kit and Wong found that Google Translate was able to handle the grammatical translation of European languages (Dutch, French, German, Greek, Italian, Portuguese, Russian, Spanish, and Swedish), East Asian languages (Chinese, Japanese, and Korean) and a Middle East language (Arabic) very well. In another study on four machine translators, Google Translate showed a better grammatical Catalan-to-Spanish translation in the domain of journalism. However, in translating medical texts, Google Translate was not so successful [14].

Verb tense, *comma*, and *spelling* were the areas that both translation systems could not handle very well. Both Google Translate and Bing Translator were not able to translate *verb tense* correctly, indicating that these online translation tools cannot make a proper distinction between different tenses, thereby failing to choose the right verb tense. Wrong *spelling* was

more common in the translations of Bing Translator, compared to Google Translate. Spelling errors by Bing Translator were mostly rooted in Persian words. In other words, Bing Translator failed to translate Persian words; therefore, they were left intact in the target text. This problem is probably rooted in the limitations involved in the vocabulary items of the dictionary used by Bing Translator. Although the two systems indicated some errors in using *comma*, Bing Translator outperformed Google Translate. Google Translate used comma exactly the same as the source text without any flexibility, while Bing Translator seemed to be more flexible in selecting the comma.

As for the *word form* and *article*, the two systems performed reasonably well. The hypothesis for the low error frequency of *articles* is that *article* as commonly used in English does not exist in Persian. Hence, both online translation systems are able to compensate this gap. Also, the hypothesis behind the low error frequency of *word form* could be that the two systems are capable of identifying parts of speech of words. In line with our results, Groves and Mundt's [13] finding showed no spelling error in Google-translated texts from Malay and Chinese to English.

To sum up, Persian-to-English translations made by Google Translate were lexically, semantically and syntactically more accurate than those rendered by Bing Translator in the field of medicine.

5. Conclusion

Machine translators cannot still replace human translators in rendering Persian-to-English complex sentences, although they could be very helpful in translating simple sentences and plain medical texts. As such, the usefulness of translation systems would highly depend on the way they are utilized. By this way, researchers can reasonably benefit from the systems in rendering plain texts from Persian to English, given that the translated versions are subjected to human editing.

Ethical Considerations

Compliance with ethical guidelines

We confirm that every facet of the current study has been informed by ethical guidelines. We also confirm that the methods used in this study have been in accordance with relevant guidelines and regulations of the AAB journal.

Funding

This study received no fund and/or grants from any institutions.

Author's contributions

The authors equally contributed to preparing this article.

Conflict of interest

The authors declare that there were no conflicts of interests.

Acknowledgments

We would like to thank the anonymous reviewers for their constructive comments and feedback.

References

[1] Champollion Y. Convergence in CAT: blending MT, TM, OCR & SR to boost productivity. InProceedings of Translating and the Computer 25, 2003.

[2] Lagoudaki E. The value of machine translation for the professional translator. InProceedings of the 8th Conference of the Association for Machine Translation in the Americas: Student Research Workshop, 2008. p. 262-9.

[3] Komeili Z, Rahimi A. An investigation of the translation problems incurred by English-to-Persian machine translations: Padideh, Pars, and Google Softwares. Procedia-Social and Behavioral Sciences, 2011. 28: p. 1079-82.

[4] Kit C, Wong TM. Comparative evaluation of online machine translation systems with legal texts. Law Libr. J, 2008. 100: p. 299.

[5] Madsen MW. The limits of machine translation. Center for Language Technology, Univ. of Copenhagen, Copenhagen. 2009.

[6] Precup-Stiegelbauer LR. Automatic translations versus human translations in nowadays world. Procedia-Social and Behavioral Sciences, 2013. 70: p. 1768-77.

[7] Aiken M, Ghosh K, Wee J, Vanjani M. An evaluation of the accuracy of online translation systems. Communications of the IIMA, 2009. 9(4): p. 6.

[8] Seljan S, Brkić M, Kučiš V. Evaluation of free online machine translations for Croatian-English and English-Croatian language pairs.

[9] Aiken, M., and S. Balan. 2011. An Analysis of Google Translate Accuracy. Translation Journal and the Author.

[10] K Khosravizadeh P, Pashmforoosh R. Google

translation: A semantic structure analysis. Lingoistica, 2011.

[11] Dhakar BS, Sinha SK, Pandey KK. A survey of translation quality of English to Hindi online translation systems (Google and Bing). International Journal of Scientific and Research Publications, 2013. 3(1): p. 2250-3153.

[12] Ganjalikhani MK. An investigation of Google translate machine in rendering English sentence structures into Persian. InProceeding of the Global Conference On Language Practice & Information Technology, 2014. p. 9-10.

[13] Groves M, Mundt K. Friend or foe? Google Translate in language for academic purposes. English for Specific Purposes, 201. 37: p. 112-21.

[14] Costa-Jussà MR, Farrús M, Serrano Pons J. Machine translation in medicine: A quality analysis of statistical machine translation in the medical domain.