Research Paper

Journal of
Language and Translation
Volume 12, Number 4, 2022 (pp. 107-118)

Assessing the Performance Quality of Google Translate in Translating English and Persian Newspaper Texts Based on the MQM-DQF Model

Zahra Foradi[1*], Jalilollah Faroughi Hendevalan[2], Mohammad Reza Rezaeian Delouei[3]
[1] MA in Translation Studies, English Department, University of Birjand, Iran
[2] Associate Professor, English Department, University of Birjand, Iran
[3] Assistant Professor, English Department, University of Birjand, Iran

Received: November 10, 2021 Accepted: March 21, 2022

ABSTRACT
The use of machine translation to communicate and access information has become increasingly common. Various translation software and systems appear on the Internet to enable interlingual communication. Accelerating translation and reducing its cost are other factors in the increasing popularity of machine translation. Even if the quality of this type of translation is lower than human translation, it is still significant in many ways. The MQM-DQF model provides standards of error typology for objective and quantitative assessment of translation quality. In this model, two criteria (accuracy and fluency) are used to assess machine translation quality. The MQM-DQF model was used in this study to assess the quality of Google Translate performance in translating English and Persian newspaper texts. Five texts from Persian newspapers and five texts from English newspapers were randomly selected and translated by Google Translate both at the sentence level and the whole text. The translated texts were assessed based on the MQM-DQF model. Translation errors were identified and coded at three severity levels: critical, major, and minor errors. By counting the errors and scoring them, the percentage of accuracy and fluency criteria in each of the translated texts was calculated. The results showed that Google Translate performs better in translating texts from Persian into English; furthermore, in sentence-level translation, it performs better than the translation of the whole text. Moreover, translations of different newspaper texts (economic, cultural, sports, political, and scientific) were not of the same quality.

Keywords: Accuracy; Fluency; Machine Translation (MT); MQM-DQF Model; Translation Quality Assessment (TQA)

INTRODUCTION
In Translation Studies, the issue of translation quality has always been of great importance. In order to eliminate the effect of subjectivity on translation quality assessment (TQA), assessment should be done based on pre-defined criteria and models. "The main goal of Translation Quality Assessment (TQA) is to maintain, and deliver to the client, buyer, user, reader, etc., of translated texts" (Doherty, 2017, p. 131).

Assessing the quality of translation is very complicated due to the subtleties of natural language. One of the issues that have always been considered in machine translation is the methods and parameters for assessing translation results. "The evaluation of machine translation (MT) systems is a vital field of research, both for determining the effectiveness of existing MT systems and for optimizing the performance of MT systems" (Dorr et al, 2011, p. 745). As the translation may be done by human or machine, translation assessment also

* Corresponding Author's email:
zahraforadi@birjand.ac.ir

may be done either by human or machine; to do so, a set of predefined criteria can be used as an instruction by a human assessor or be defined in the form of a quality assessing software that performs the assessment automatically; each of them has advantages and disadvantages. Automatic assessment is low-cost and objective, but types of errors in texts are not well shown. The most valid technique for judging the quality of translation is the manual assessment by human assessors, but the inevitable negative point is that it is time-consuming and costly (see Maučec & Donaj, 2019). Furthermore, human assessment depends heavily on the skills of the assessor. Eventually, the performance quality of a machine assessor must also be assessed by a human assessor.

The structure of a translation quality assessment model includes the classification of types of errors and their weighting based on severity levels. If translation errors are considered throughout the text in general and finally a score is assigned to the whole text, the assessment is holistic, and if the errors are examined in detail, the assessment is analytic. Until recent years, all the assessment approaches have been holistic (Karimnia, 2011) and evaluation criteria have been mainly subjective (Saldanha & O'Brien, 2013).

According to Waddington (2001), the literature review on TQA shows that almost all the previous studies have been descriptive or theoretical. Waddington's model was introduced through empirical studies, which have been used in numerous research to assess the quality of translation (see, for example, Andishe Borujeni, 2020).

With the expansion of machine translation use and increasing awareness of translation consumers, the need for more detailed translation evaluation models, which could be used for machine translation evaluation, gradually increased. Around 2011, TAUS (Translation Automation User Society) sought to create a new way of measuring translation quality to satisfy its customers. O'Brien undertook a project in collaboration with TAUS to explore the potential of a dynamic quality estimation model. The first part of this study was conducted in 2011 on 11 current quality evaluation models. It was the starting point of benchmarking translation quality evaluation models to achieve a comprehensive TQA model (O'Brien, 2012).

"Translation quality assessment (TQA) has suffered from a lack of standard methods. Starting in 2012, the Multidimensional Quality Metrics (MQM) and Dynamic Quality Framework (DQF) projects independently began to address the need for such shared methods" (Moorkens et al. 2018, p. 109). According to Lommel et al. (2018, p. 23), "the Multidimensional Quality Metrics (MQM) framework for describing and defining translation quality assessment metrics was developed in the EU-funded QTLaunchPad project".

Some studies in the field of TQA have been conducted based on the MQM model, including "The evaluation of the quality of Crowdsourcing Translations of Wikipedia Articles based on the MQM model" (Vahedi Kakhki, 2018), whose results showed the relatively high quality of the translation of these texts.

"Dynamic Quality Framework (DQF) is a comprehensive suite of tools for quality evaluation of both human and machine translation developed by Translation Automation User Society (TAUS)" (Lommel et al, 2018, p. 27).

According to Moorkens et al (2018, p 109), "in 2014 these approaches were integrated, centering on a shared error typology (the 'DQF/MQM Error Typology') that brought them together."

In this way, standard benchmarks were established to assess the quality of translation. "This approach to quality evaluation provides a common vocabulary to describe and categorize translation errors and to create translation quality metrics that tie translation quality to specifications" (Moorkens et al, 2018, p. 109). From the perspective of this model, two quantified dependent variables, namely accuracy, and fluency, influence the independent variable, i.e. quality.

The MQM-DQF model was the metric used in the present research in order to evaluate the quality of Google Translate in translating newspaper texts from English into Persian and vice versa.

Numerous case studies have been conducted to evaluate the machine translation quality of various text types from English into Persian and vice versa. For example, Pajhooheshnia (2015) evaluated the quality of machine translation for technical texts from Persian into English based on two criteria of adequacy and fluency by human evaluators. The human evaluator score was 2.62 for the accuracy criterion and 2.45 for the fluency criterion.

In another study on the translation quality of proverbs translated from English into Persian by Google Translate, the results showed that "the minorities of the translated proverbs provide adequate comprehension" (Torkaman, 2013, p. iv).

According to Moradi's (2015) research, in the evaluation of machine translation quality in translating English scientific-technical articles, including four empirical sciences, i.e. biology, chemistry, mathematics, and physics, based on EuroMatrix, "GT had the best performance in translating physical subtype and the worst in mathematical, biological and chemical being second and third in rank" (Moradi, 2015, p. 89). The overall adequacy and fluency values for GT were 47.5% and 46.27% respectively.

Some comparative studies have been also performed. In a study conducted in 2017 to estimate the translation quality of two machine translators (including Google Translate) from English into Persian, most linguistic errors were found to have occurred in the areas of morphology, complex sentences, syntactic ambiguity, and semantic analysis, generation of Persian, and long sentences (Aabedi, 2017). Aabedi (2017) and Sharifiyan (2018) found that Google performed better than other translation machines.

In the field of machine translation quality assessment for literary texts, some research also has been done, including "Efficacy in Translating Verb Tense from English into Persian"; the result of this research showed that Google Translate could not translate verb tense appropriately into Persian (Hakiminejad & Alaeddini, 2016). In this study, we continued these machine translation quality assessments in the genre of newspaper text translation.

Purpose of the Study

Every day around the world, large volumes of newspapers are published in different languages. To communicate correctly and effectively with other nations and increase public awareness, the translation of these texts is necessary and is done by various news agencies around the world. Due to the high volume of production and publication of these texts that is going on all over the world every day, translating newspaper texts is necessary and, at the same time, very difficult. Using machine translation is a solution to deal with time and volume constraints in the translation of journalistic texts.

The integrated model (MQM-DQF) was used in this study to assess the performance quality of Google translate in translating newspaper texts both at the level of sentence and the whole text from English into Persian and vice versa. We were looking to answer these research questions:

1) Based on the MQM-DQF assessment model, how is the translation quality of English newspaper texts translated into Persian by Google Translate assessed?

2) Based on the MQM-DQF assessment model, how is the translation quality of Persian newspaper texts translated into English by Google Translate assessed?

3) According to the MQM-DQF model, in which case is the performance quality of Google Translate better, in translating English newspaper texts into Persian or in translating Persian newspaper texts into English?

4) According to the MQM-DQF model, at which translation level (at the level of sentence or the whole text) is the performance quality of Google Translate in translating Persian and English newspaper texts better?

METHOD

The purpose of the research was to assess the quality of Google Translate performance in translating English and Persian newspaper texts based on the MQM-DQF model. To this end, the first well-known and widely circulated

domestic and foreign newspapers were identified. Five texts (economic, cultural, sports, scientific and political) were selected randomly. A simple random sampling was also done to select these five types of texts from domestic newspapers and five from foreign newspapers.
Five selected Persian texts and five selected English texts were translated separately, both sentence to sentence and whole text, by Google Translate. Then, the translated texts were examined and analysed based on the MQM-DQF model and the two criteria (i.e. accuracy and fluency) and their subsets. Errors at three severity levels (critical, major, and minor) were identified. The error typology diagram used in the research is as follows.

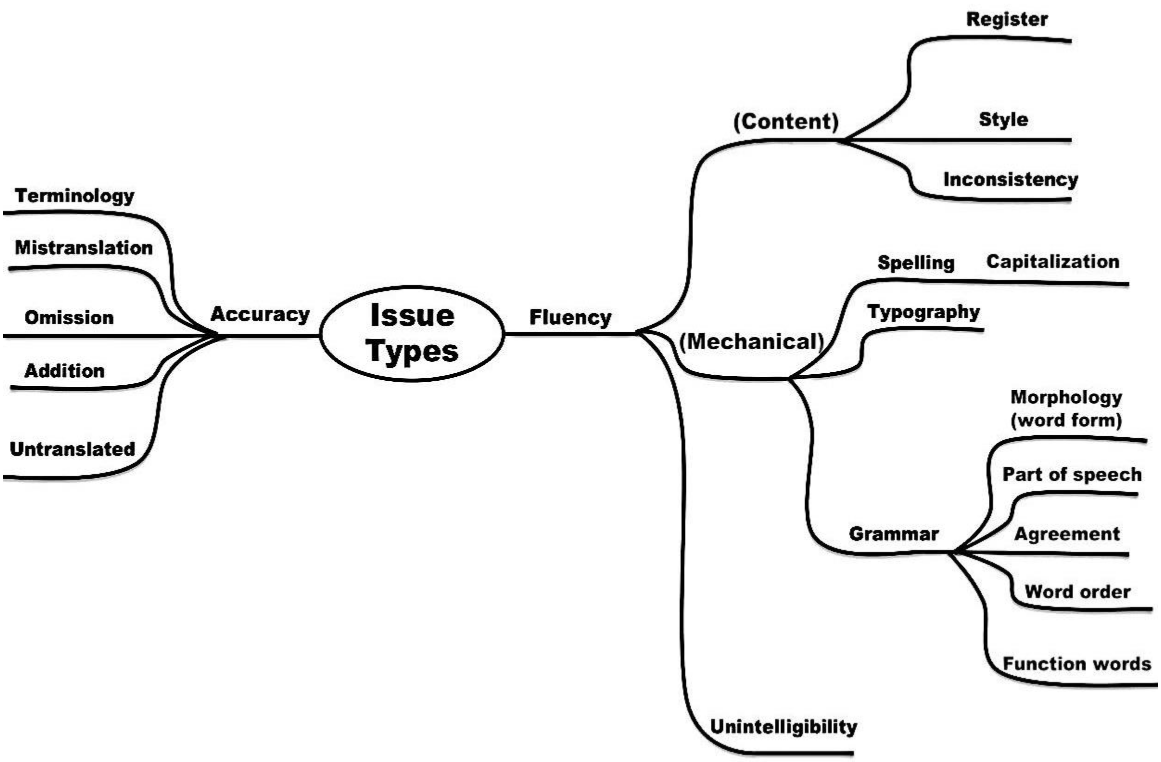

Figure 1
A subset for MT analysis (Lommel and Melby, 2018, p. 30)

Sub-categories of criteria

According to Fig. 1, there are five sub-categories to check accuracy in machine translation outputs:

Terminology: Is there a proper equivalent for each word?

Mistranslation: Has the word or phrase been mistranslated?

Omission: Has deletion happened during the translation?

Addition: Has anything been added to the translated text that did not exist in the source text?

Untranslated words or phrases: Has an untranslated word or phrase been transferred from the source text to the translated text?
There are also seven general sub-categories to assess fluency in machine-translation outputs:

Register: Have language characteristics in the source text been transferred into the target language?

Style: Has the social, cultural, and temporal position of the source language been transferred into the target language?

Inconsistency: Has a specific word or phrase repeated throughout the text been translated equally everywhere? Does the meaning of one part of the translated text not violate another part of it?

Spelling: Is the spelling of the words correct? (For example, spelling mistakes change the meaning of a word entirely or write the first letter of proper nouns in lowercase.)

Typography: Is the shape, size and spacing of the letters and paragraphing of the text appropriate?

Grammar: Are the grammatical rules in the translated text followed correctly?

Unintelligiblity: Does the translated text convey its original meaning clearly?

Error severity levels (weighting)

Evaluating a translation is not enough just to know the number of errors. "Evaluators also need to know how severe they are and how important the error type is for the task at hand" (Moorkens et al, 2018, p. 120). According to Lommel et al (2018, p. 33), "severity level is an indicator of the importance of an issue with an accompanying numerical representation and weight is a numeric representation of the importance of an issue type in a specific metric".

Critical error is an error the user does not notice. Critical errors per se make the translation unsuitable for its intended purpose. However, a critical error can prevent the translation from achieving its purpose. For example, if in the translation of an industrial text a product weighs 2 pounds (approximately 0.9 kg) and it is translated into 2 kg, the result of this translation may be to the detriment of the user while he/she does not notice it; therefore, it is a critical error. The default score for each of these errors is 100.

Major errors in translation make the meaning of the text unclear to the user. For example, if in translating an educational text about insects from Italian into English, the word "ape" that means "bee" in Italian and "monkey" in English is translated as "monkey", it has negative consequences but the user is aware of this error. However, this type of error does not harm the user because he realizes this error. The default score for this group is 10.

If there is more than one major error per thousand words in a translated text by a holistic assessment method, the assessor is required to change the assessment method to an analytic one and find the exact cause of errors.

Minor errors do not affect the usability of a translated text. In many cases, the user consciously ignores these types of errors. For example, in translating a text into English, the translator may say *to who it may concern* instead of *to whom it may concern*, which the user ignores. Minor errors do not pose a problem in conveying meaning. The default score for this group is 1.

Null errors include errors that are ignored by default before text assessment begins. For example, in assessing translation of a brochure related to the instructions of an electrical device, the style of writing may not be essential, so it will not be part of the translation criteria at all, and if such an error is seen in the text, it will be ignored. The score of these errors is always zero (Moorkens et al., 2018, pp. 116–122).

According to the definitions of the severity level of error, each previously detected type of error was labelled. Then, the percentages of both accuracy and fluency criteria were calculated using the following formula:

$$\text{Score} = 1 - \frac{\text{minor}(1) + \text{major}(10) + \text{critical}(100)}{\text{Word Count}}$$

The final score is numerically between zero and one.

The point here is to count the words in the translated texts correctly. Google does not regard zero-width non-joiner (ZWNJ) in translating from English into Persian; therefore, in translating the five texts from English into Persian, formal Persian editing was done to correct the ZWNJ to obtain the exact number of words in the translated texts.

By weighing the errors and calculating their scores, the rate of accuracy and fluency criteria can be reported quantitatively (percentage), using descriptive statistics (frequency and percentage calculation).

RESULTS

The quality of Google Translate performance in translating English newspapers (economic, cultural, sport, political, and scientific) texts into Persian and the same Persian newspaper text types into English, at both sentence-level translation and whole-text translation, was examined based on MQM-DQF model and compared by calculating percentages of two criteria or variables, i.e. accuracy and fluency. As mentioned before, the object of the current study was not to compare the translation quality of Google Translate in translating different types of newspaper texts (i.e., scientific, cultural, political, economic, sports, etc.) and such a comparison was postponed to future research. The results are as follows:

Research Findings

Economic texts: In English to Persian translation, at both levels, the maximum number of accuracy errors belonged to mistranslation subcategory and in the case of fluency criterion, the maximum errors belonged to unintelligibility subcategory.

In Persian to English translation, at both levels, the maximum number of accuracy errors was related to the terminology subcategory and in the case of fluency criterion, the maximum errors were related to the grammar subcategory.

According to the calculation of each criterion percentage in the translation of the two selected newspaper texts (Table 1 and Table 2), the percentages of accuracy and fluency criteria in the whole translation of each text are lower than the sentence level translation, and the percentages of both criteria in Persian into English translation are more than the exact percentages in English–Persian translation.

Cultural texts: In English to Persian translation, at both levels, the maximum number of accuracy errors belonged to terminology and the maximum number of fluency errors belonged to the unintelligibility subcategory.

In Persian to English translation, the maximum number of accuracy errors at the sentence level translation was related to terminology, and in the translation of the whole text, it was related to the subcategory of mistranslation. The maximum number of fluency errors in translating the whole text was related to the unintelligibility subcategory and at the sentence level translation, the number of errors in two subcategories of unintelligibility and grammar was the same.

Based on the percentages calculated for each of the criteria in two translated texts, the percentages of accuracy and fluency criteria in the complete translation of each text are less than the sentence level translation, and the percentage of both criteria in Persian–English translation is more than English–Persian translation.

Sport texts: In English to Persian translation, all accuracy errors were related to the mistranslation subcategory at both levels, and most fluency errors were related to the unintelligibility subcategory.

In Persian into English translation, most accuracy errors were related to terminology subcategory and about fluency errors; most of them were related to unintelligibility.

Based on the percentages calculated for each of the criteria in the two translated texts, percentages of accuracy and fluency criteria in the complete translation of each text were less or equal to the sentence level translation. At both translation levels, accuracy in Persian to English translation was less than English to Persian translation, but fluency in Persian to English translation was more than English to Persian translation.

Political texts: In English to Persian translation, at both levels, the highest number of errors in accuracy criterion belonged to mistranslation subcategory and about fluency criterion the most errors belonged to typography subcategory.

In Persian to English translation, at both levels, the highest number of errors in terms of accuracy was related to mistranslation subcategory and regarding fluency, the highest number of errors was related to unintelligibility.

Based on the percentages calculated for each of the criteria in the two selected texts, accuracy and fluency in the complete translation of each text were lower than translation at the sentence level. In both translation levels, accuracy in Persian–English translation was lower, but fluency was higher than English–Persian translation.

Scientific texts: In English to Persian translation, the number, type and severity of accuracy and fluency errors were similar at both levels. The highest number of accuracy errors was related to mistranslation, and the highest number of fluency errors was related to unintelligibility.

The number and type of accuracy and fluency errors in English to Persian translation were similar at both levels. Most accuracy errors were related to mistranslation. Half of the fluency errors were related to unintelligibility and the other half were related to grammar.

According to Table 1 and Table 2, accuracy and fluency criteria in English to Persian translation were equal at both levels. In Persian to English translation, accuracy criterion percentage in sentence-level translation was higher, and fluency percentage was lower than complete text translation. At both levels of translation, the percentage of both criteria was higher in Persian–English translation compared to English–Persian translation.

Table 1

Percentages of quality criteria for newspaper texts translated from English into Persian (at the level of sentence and whole text)

Texts	Economic		Cultural		Sport		Political		Scientific	
Criteria (percentage)	sentence level	whole text	sentence level	whole text	sentence level	whole text	sentence level	whole text	sentence level	whole text
Accuracy	90.4	87.6	79.7	73.8	86.9	86.9	92.2	89.7	72.2	72.2
Fluency	93.7	91.8	91.2	90.9	91.1	89.5	94.5	93	92.1	92.1

Table 2

Percentages of quality criteria for newspaper texts translated from Persian into English (at the level of sentence and whole text)

Texts	Economic		Cultural		Sport		Political		Scientific	
Criteria (percentage)	sentence level	whole text	sentence level	whole text	sentence level	whole text	sentence level	whole text	sentence level	whole text
Accuracy	93.8	87.9	94.1	89.9	80.9	76.4	90	86.9	96.6	85.2
Fluency	99.8	95.2	98.2	94.6	94.8	93	99.8	94.9	98.2	98.2

Frequent errors in translating these texts by Google Translate are summarized below:

- Google Translate could not understand the implicit meaning of many satires, proverbs, metaphors, allusions, idioms, slang terms, cultural elements, and similar items, and has made mistakes in translating them. For example, the phrase *to tick every box* is an idiom and means “to satisfy all of the apparent requirements for success” (Collins dictionary, 2021). But such a concept is not transmitted to the Persian-speaking reader, because Google Translate has translated it as برای علامت زدن هر جعبه (baray e alamat zadan e har ja'be) into Persian which is a word by word translation.
- In the UK, the Prime Minister’s building number is 10, and across the UK No. 10 implicitly means government; however, Google Translate transmitted exactly the phrase No.10 into the translation, without any additional cultural explanation.
- About some children’s activities, which have cultural aspect, Google Translate does not have the appropriate equivalent words and

phrases that can be understood in the target culture, including *rainbow trails* and *doorstep discos* or *electric milk float* that means *milk distribution car*. However, it has been translated as شناور شیر الکتریکی (shenavar e shir e electrici) into Persian. They are considered major errors because Google Translate has translated them word by word into Persian with no additional explanation.

- Google Translate may make mistakes in translating words with several common meanings (polysemy), and it may misuse different meanings. For example, the word *pilot*, which has several meanings, should have been translated as پارکینگ مسطح (pilot parking), but has been translated as خلبان (aviator).
- Google Translate may leave unfamiliar words in the target culture untranslated or translate them as the most similar word structure it knows. For example, *parklet* that means *small park*, has been translated as پارک جیبی (pocket park), which is similar to the structure of the word *booklet*.
- There were also some untransliterated words in these translations, most of which being proper nouns, and they are minor fluency errors. Google Translate has transmitted English proper nouns exactly in Latin and not in Persian alphabet, such as *CNBC* and *Covid*, which are understandable for the reader according to context.
- Google Translate made mistakes in translating proper nouns from Persian into English and translated these nouns instead of just transliterating them, such as translating *Fekri* (person’s name) into *intellectual* or *Esteghlal* and *Nassaji* (club name) into *independence* and *textile*.
- Google Translate has made mistakes in converting and transferring units of weight, distance, time, volume, etc. For example, the solar year *1398* has been transferred without conversion, so the reader realizes this error, and it is a major accuracy error. However, in another text translation, Google Translate mistranslated the same solar year into *2009*, which was considered a critical accuracy error.
- Google Translate could not recognize words with ZWNJ in Persian texts and sometimes connects parts of such words and translates the new words. Sometimes this connection creates an irrelevant word, such as the phrase کم‌کاری برخی از بازیکنان استقلال (the negligence of Esteghlal players) where the word کم‌کاری (negligence) has ZWNJ and has been changed to کمکاری (helping), and then it has been translated to *some Esteghlal players were helping*.

Table 3

Comparison of percentages averages (se is the abbreviation for sentence level, wh for whole text, A for accuracy and F for fluency)

	Accuracy average in sentence level translation	Accuracy average in whole text translation	Fluency average in sentence level translation	Fluency average in whole text translation
English to Persian	$\bar{x}_{A.se}= 84.28$	$\bar{x}_{A.wh}= 82.04$	$\bar{x}_{F.se}= 92.52$	$\bar{x}_{F.wh}= 91.46$
Persian to English	$\bar{x}_{A.se}= 91.08$	$\bar{x}_{A.wh}= 85.26$	$\bar{x}_{F.se}= 98.16$	$\bar{x}_{F.wh}= 95.18$

*A (Accuracy)
*F (Fluency)
*se (sentence level translation)
*wh (whole text translation)

The average of accuracy percentages in translation from English into Persian is about 84% in sentence-level translation and about 82% in complete text translation. The average fluency percentages are approximately 92% and 91% in sentence-level translation and complete translation, respectively (Table 3). Persian's average accuracy percentages into English are about 91% in sentence-level translation and about 85% in complete text translation. The average of fluency percentage is approximately 98% in sentence-level translation and about 95% in whole text translation (see Table 3).

As Table 3 shows, all the averages of accuracy and fluency percentages in the texts translated from Persian into English were higher than the averages of the texts translated from English into Persian. According to Table 1 and Table 2, all percentages related to sentence-level translations are equal or higher than whole text translations.

DISCUSSION

Machine translation (MT) is related to how a computer software translates texts from one language into another without human intervention. The latest approach in machine translation and the approach used in this research is the Neural Machine Translation approach (NMT). Google Neural Machine Translation (GNMT) is an NMT system that was developed by Google and was introduced in November 2016. It uses an artificial neural network to increase fluency and accuracy in Google Translate. It was also stated that with the upgrade of Google to NMT, translation of texts would be done at the sentence level (Turovsky, 2016).

In recent years, researches have been conducted on the quality of online machine translation from English into Persian and vice-versa, either by human assessor or machine assessor, in various genres. However, it should be noted that the results of these researches can no longer be reliable today due to the improvement of approaches to machine translation. Here we discuss these researches and their results.

In 2011, an automated accuracy evaluation was done using 51 languages with Google Translate based on BLEU. Due to a large number of languages in this study, human evaluation was not used. The results showed that “translations between European languages are usually good, while those involving Asian languages are often relatively poor” (Aiken & Balan, 2011, para. 1). The results were confirmed by Benjamin (2019): The corpus of stored texts in Google Translate is richer for upper languages, English above all, and poorer for bottom languages. On the other hand, in translating all pairs of non-English languages, Google acts indirectly; first, it translates each language into English and then translates it into the other language. Therefore, the corpus of English texts in Google Translate is more prosperous than other languages (Benjamin, 2019). But according to a 2016 study on Google Translate performance, the difference between the frequencies of different types of English–Persian and Persian–English errors did not reach statistical significance, the conclusion being that the direction of translation does not affect the quality of the translation of Google Translate (Ghasemi & Hashemian, 2016, pp. 15–16). As confirmed in the present study, Google’s errors in translating texts from Persian into English are more than translating from English into Persian.

In 2019, a study was conducted on fifty languages used in Aiken & Balan’s (2011) study using the same sample texts, which aimed to compare the improvement of Google Translate in terms of accuracy. In 2011, the approach of Google Translate was PBMT (Phrase-Based Machine Translation), which upgraded to NMT eight years later. In the 2011 study, the BLEU score in translation from English into Persian was 16. The 2019 study gave a 39 BLEU score for translation from English into Persian and a 66 BLEU score for translation from Persian into English (Aiken, 2019). As it was found in the current research, the quality of Google translation in translating from English into Persian is better than translating from Persian into English.

According to a study conducted in 2018 to evaluate the translation quality of three translation machines (including Google Translate) from Persian into English, out of six common error categories, the most common categories were missing words, extra words, and word orders (Sharifiyan, 2018).

In the present study, in both English into Persian and Persian into English translations, the highest number of accuracy errors was mistranslation and the highest number of fluency errors was related to unintelligibility. Among the subgroups, errors of grammar subcategory including morphology, part of speech, agreement, word order and function words, the highest number of errors in translation from English into Persian and vice versa was related to the part of speech.

As already mentioned, due to the improvement of machine translation approaches, the results of previous research in evaluation of machine translation quality have been improved and this trend continues, and because of this progress in machine translation quality, translation quality assessment models need to be improved as well.

CONCLUSION

We used MQM-DQF model, one of the most up-to-date TQA models, and its two criteria (i.e. accuracy and fluency) and their subsets to assess Google Translate quality performance in translating newspaper texts (Persian and English).

It was found that the percentages of accuracy criterion in five English newspapers texts translated into Persian were between 72.2% and 92.2%, and the percentages of fluency criterion in the texts translated were between 89.5% and 94.5% (Table 1). The percentages of accuracy criterion in five Persian newspapers texts translated into English were between 76.4% and 96.6%, and the percentages of fluency criterion in the translated texts were between 93% and 99.8% (Table 2). The accuracy and fluency criteria in sentence-level translation were equal or higher than whole-text translation. All the averages of accuracy and fluency in the newspapers translated from Persian into English were higher than those of the newspapers translated from English into Persian. Therefore, it can be concluded that Google Translate has a better performance in translating Persian newspaper texts into English than translating newspaper texts from English into Persian. Moreover, the averages of accuracy and fluency in the sentence-level translation are higher than whole text translation; therefore, it can be said that Google Translate has a better quality performance in translating newspapers texts at the level of sentence than the level of whole text.

REFERENCES

Aabedi, F. (2017). *Quality Estimation and Generated Content of Machine Translation Programs: Babylon 10 Premium Pro Review and Google Translator* (Master's thesis, Faculty of Literature and Human Science, Imam Reza International University, Mashhad, Iran). Retrieved May 8, 2021, from https://ga nj.irandoc.a c.ir/#/artic les/8 fb842d6931a485d5874ea36c95e49cb

Aiken, M., & Balan, S. (2011). An analysis of Google Translate accuracy. *Translation Journal, 16(2), April.* Retrieved July 12, 2021, from https://t ransla tionjou rnal.ne t/journal/56google.htm.

Aiken, M. (2019). An Updated Evaluation of Google Translate Accuracy. *Studies in Linguistics and Literature, 3*. p. 253. 10.22158/sll.v3n3p253.

Andishe Borujeni, F. (2020). *The quality of translation of cultural elements translated by human and machine in Natour Dasht novel based on Weddington model* (Master's thesis, Faculty of Foreign Languages, Sheikh Baha'i University). Retrieved Avril 25, 2021, from http s://gan j.irando c.ac.ir /#/articles/24d46684b88c4d7010f7c2de 74fb27ac

Benjamin, M. (2019). *When & How to Use Google Translate*, Retrieved March, 27, 2021, from https:// www.tea chyouba ckwards.com/how-to-use-google-tran slate/

Collins dictionary online. (2021). Retrieved June 9, 2021, from https:// www.coll insdictionary.com/ dictionar y/english/t ick-all-the-boxes

Doherty, S. (2017). Issues in human and automatic translation quality assessment. In D. Kenny (Ed.), *Human issues in translation technology* (pp. 131 – 148). London, UK: Routledge.

Dorr, B., Olive, J., Mccary, J., & Christianson, C. (2011). Machine Translation

Evaluation and Optimization. *Handbook of Natural Language Processing and Machine Translation*, by Olive, Joseph; Christianson, Caitlin; McCary, John, ISBN 978-1-4419-7712-0. Springer Science+Business Media, *LLC,* 2011, p. 745. -1. 745. 10.1007/978-1-4419-7713-7_5.

Ghasemi, H & Hashemian, M. (2016). A Comparative Study of Google Translate Translations: An Error Analysis of English-to-Persian and Persian-to-English Translations. *English Language Teaching*, *9*. 13. 10.5539/elt.v9n3p13.

Hakiminejad, A & Alaeddini, M. (2016). A Contrastive Analysis of Machine Translation (Google Translate) and Human Translation: Efficacy in Translating Verb Tense from English to Persian. *Mediterranean Journal of Social Sciences*, *7*. 10.5901 /m jss.2 016.v7n 4S2p40.

Karimnia, Amin. (2011). Waddington's model of translation quality assessment: a critical inquiry. *Elixir Ling. & Trans. 40*, 5219-5224.

Lommel, A and Melby, A. (2018). MQM-DQF: A Good Marriage (Translation Quality for the 21st Century), *The 13th Conference of The Association for Machine Translation in the Americas*, Retrieved December, *5*, 2020 from www.conference.amtaweb.org

Lommel et al, (2018). Harmonised Metric, Qt21 Deliverable 3.1. p.23. Retrieved March13, 2021 from http://w ww.qt2 1.eu/

Maučec, M & Donaj, G. (2019). *Machine Translation and the Evaluation of Its Quality*. DOI: 10.5772 /intechop en.89063. retrieved 28 July 2021 from: https://www.intechopen.com/chapters /6 8953.

Moorkens, J et al (2018). Translation Quality Assessment: From Principles to Practice. Berlin, Germany: Springer.

Moradi, S. (2015). *Comparison of translation of scientific-technical texts by two web-based translation machines based on Eurometrics* (Master's thesis, Faculty of Literature and Humanities, University of Shahid Bahonar, kerman, Iran). Retrieved Feb 22, 2021 from https://ganj.irandoc.ac.ir/#/articles/4182 d0197e7b33873b141664483f5fd3

O'Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *The Journal of Specialised Translation*, *17*, 55–77 Retrieved Avril 18, 2021from https://www.jostrans.org/issue17/art_ ob riNn. pNdf

Pajhooheshnia, M (2015). *Manual and automatic comparative evaluation in online machine translation systems; Case study: Persian-English translation of technical texts* (master's thesis, Sheikhbahaei University, Esfahan, Iran). Retrieved Avril 25, 2021 from https://ganj. irandoc.ac.i r/#/art icles/41b 29375cf56 G6c1ae9ff607ea8029cfe

Saldanha, Gabriela and O Brien, Sharon (2013). *Research Methodologies in Translation Studies*. London & New York: St. Jerome Publishing.

Sharifiyan, L. (2018). *Evaluating Cohesion and Comprehensibility in Persian-English Machine Translated Texts* (Master's thesis, Faculty of Persian Literature and Foreign Languages, Allameh Tabataba'i University, Tehran, Iran). Retrieved May 2, 2021 from https ://ganj.i randoc.ac .ir/#/articles/e8bb80d702dc3792ab08 e01f4c2e36ad

Torkaman, E. (2013). *A comparative Study Quality Assessment of Machine (Google translate) and Human Translation of Proverbs from English to Persian* (Master's thesis, Islamic Azad University, Central Tehran Branch – Faculty of Foreign Languages – Department of English). Retrieved July 21, 2021 from https://ga nj.irandoc .ac.ir/#/articles/51d717a038e26c50aee9 df9beb143c0f

Turovsky, B. (2016). Found in translation: More accurate, fluent sentences in Google Translate, Retrieved April 27, 2021 from https://blo g.google/pro ducts/translate/found-translation more-accurate-fluent-sentences-googletranslate/

Vahedi Kakhki, A. (2018). *Quality Evaluation of Persian Crowdsourcing Translations of Wikipedia Articles by MQM*, (Master's thesis, Faculty of Literature and Humanities- Department of Foreign Languages, University of Shahid Bahonar, Kerman, Iran). Retrieved Feb 11, 2021 from https:/ /ganj.ira ndoc.ac.ir /#/articles/1f803e5d3 3f 98f 3de 54a 959 d92bbb2d2.

Waddington, C. (2001). Different methods of evaluating student translation: The question of validity. Meta: *Translators' Journal*, *46(2)*, *311–325* Retrieved April 18, 2021 from https://www.erudit.org /fr/revues/meta/ 2001-v46-n2-meta 159 /004583ar.pdf.

Biodata

Zahra Foradi is holds an MA in Translation Studies from the University of Birjand, Iran. Her research interests include machine translation, parallel linguistic corpora, the history of translation, and bibliometric studies.
Email: *zahraforadi@birjand.ac.ir*

Dr. Jalilollah Faroughi Hendevalan is associate professor at the University of Birjand, Iran. His research interests include machine translation and critical discourse analysis.
Email: *jfaroughi@birjand.ac.ir*

Dr. Mohammad Reza Rezaeian Delouei is assistant professor at the University of Birjand, Iran. His research interests include conceptual research, interdisciplinary research, bibliometric studies, and translational language.
Email: *mrrezaeiand@birjand.ac.ir*