Applications of Language Studies
jals.hmu.ac.ir
Hazrat-e Masoumeh University

Text Readability of Machine Translation vs. Human Translation in Literary Texts: the Case for the Old Man and the Sea

Samad Mirza Suzani

Assistant Professor, Department of English, Marvdasht Branch, Islamic Azad University, Marvdasht, Iran.
smirzasuzani@yahoo.com

Abstract

Comparing human and machine abilities for translating English literary texts into Persian and paying attention to the differences in the readability levels of the items to be translated into Persian by machine and human translators were the main impetus to conduct this study. Accordingly, the present study aimed to examine the legibility of two versions of Persian translation of the book "The Old Man and the Sea" (1951) by Google translate and human translator (namely, Azima, 2006). The comparison was done by Flesch Kincaid Grade, with different sub-criteria for evaluation. The findings suggest that both machine and human translations are very different based on Flesch Kincaid Grade criteria so that human translation may even be more readable. The results of this study can affect curriculum planning, especially in the field of translation. Another benefit of this study is that students could become familiar with their strengths and weaknesses in translation and hence it can be effective in evaluating translation students as would-be translators.

Keywords: Google Translate, Human translation, Machine translation, The Old Man and the Sea, Text readability

Citation: Mirza Suzani, S. (2024). *Applications of Language Studies*, 2(2), 202-225.

Received: 2024-02-20
Revised: 2024-04-20
Accepted: 2024-06-09
Article type: Research paper

P-ISSN: 2588-6975
E-ISSN: 2588-6983
Publisher: Hazrat-e Masoumeh University Press
DOI: 10.22034/jals.2024.2023467.1016

1. Introduction

In order to achieve good translation quality, a translator must pay attention to many aspects of the source language (hereafter SL) text that affect the form of the target language (hereafter TL) text. As McDonald (2020) asserts, "the purpose of translation is not to create a new work or a new writing, but to create a bridge between the author of the source language text and the readers of the target text" (p. 36). In other words, a translator does not translate the text into a new article, but must be able to act as a facilitator between the SL text and the TL text to convey the message.

Furthermore, McDonald (2020) argues that when it comes to the issue of translation evaluation, three factors need to be considered: acceptability, accuracy, and readability. Acceptability refers to "the fluency and naturalness of the translated text in accordance with the rules, language norms of the readers of the target text" (p. 23); aaccuracy means that "the transmission of the message or information of the source language is done accurately and honestly in accordance with the author's source language (p. 30), and readability is "the degree of ease of writing and understanding its meaning" (p. 30).

Based on McDonald (2020), the term acceptability gives the idea that the text must be accepted and understood by the reader of the TL, and that "if the cultures and target languages follow adherence to the norms and cultures of the source language, translation will be very appropriate" (p. 30). On the other hand, being accurate means "accurate transmission of meaning" (p. 33), while a translated text has a high level of readability if "it is easy to read and the reader can receive the text message of the target language without matching the text of the source language" (p. 33).

According to Crosbie (2013), the reader's understanding of the written text can be called readability of the text. Thus, the readability of a text is closely related to its content (vocabulary, complexity, and syntax), how it is presented (font size, line height, character spacing, and line length) and the ability to distinguish letters or characters from each other.

This study draws its significance from four major points: first, knowing translation software and their applications as an important point in translation; second, obtaining a better

evaluation of human translations, especially in relation to translators and interpreters; third, assessing readers' understanding of the translated texts; and finally, comparing the readability of text in human and machine translation with the help of readability software. Thus, the following research questions were raised in the present study:

1. How does machine translation differ from human translation in terms of rendering the readability of literary texts from English into Persian?

2. Is there any significant difference between the levels of item readability translated in Persian by machine and human?

2. Background

Some recent and clear studies have been conducted on the importance of readability. Rezaee and Norouzi (2011) conducted a study on readability and obtained a very high correlation between readability and performance indicators of readers. In another study on the relationship between the readability criteria and the comprehensibility level of source texts and their translations, Mujiyanto (2016) carefully studied the comprehensibility and readability of English texts and their translations and concluded that the level of comprehension of texts is attributed to their readers while the ability to read more depends on the text (cited in Acar & İşisağ, 2017).

It was also revealed that the comprehensibility and readability are intertwined so that they are sometimes used interchangeably, yet there are obvious differences between them. Acar and İşisağ (2017) maintain that perception is a reader-centered approach that includes comprehension, general knowledge, readers' cognitive networks, and the like. Kolahi and Shirvani (2012) state that although readability is a prerequisite for comprehension, it cannot be concluded that all readable texts are comprehensible or vice versa.

2.1. Studies on Text Readability Checker and Readability Formulas

The concept of readability has been widely discussed and introduced decades ago, and as Mujiyanto (2016) states, this concept was used as if it was “the only important topic that could be considered as a starting point for writing a text” (p. 5). Readability is a complex concept that relates both to the properties of texts and to the people and skills of readers.

Dubay (2004) recognizes readability as what makes some texts easier to read than others. Moreover, Acar and İşisağ (2017) argue that readability has most often been sacrificed for the sake of conveying the meaning of terms.

According to Dubay (2004), principles of readability include: using short, simple, familiar words; avoiding jargon; using culture-and-gender-neutral language; using correct grammar, punctuation, and spelling; using simple sentences, active voice, and present tense; beginning instructions in the imperative mode by starting sentences with an action verb, and using simple graphic elements such as bulleted lists and numbered steps to make information visually accessible.

Chall (1958) defines four elements for which the criterion of readability is important: vocabulary load, sentence structure, idea density, and human interest. Mühlenbock and Kokkinakis (2009) consider the criteria of readability as a combination that can include different aspects of readability. Vocabulary can be measured based on the word frequency ratio or the ratio or number of long words. In this vein, Moses was used to teach the translation model (Cohen et al., 2007) and SRILM to teach the language model (Stolcke, 2002).

Regarding the text readability, Crosbie (2013) states that higher readability plays a very important role in reading speed for all readers, especially those with low comprehension. Thus, increasing the readability level from average to good in readers with moderate or poor understanding can be very decisive in the success or failure of achieving goals. Besides, Crosbie (2013) adds that readability exists in both natural and programming languages in various forms. In programming, things like programmer comments, loop structure selection, and name selection can determine how easy it is for humans to read a computer program code.

Checking the readability of texts is important for many corporate and individual users, and as Brück et al. (2008) maintain, formulas for approximate text readability have a long tradition. According to Brück et al. (2008), the text readability checker is used to identify texts that are difficult to read and it can help writers write easy-to-read texts. In addition, it

often shows a universal readability score derived from a readability formula that describes the readability of the text numerically.

According to Klare et al. (1954), the empirical evidence obtained on the trap of the readability formulas indicates that a number of experts have been asked to re-examine the reliability of these formulas as a tool to measure people's understanding of specific texts. Meanwhile, Stephens (2000), states that readability tests can only measure text surface features so that qualitative factors such as difficulties in comprehension of words, composition, sentence patterns, property, abstractness, ambiguity as well as incoherence cannot be measured mathematically.

Maria et al. (2014) argue that most traditional readability approaches examine the surface properties of a text to determine the complexity of a text. Also, she believes that these readability criteria are based on the assumption that it relates superficial features to linguistic factors influencing language. She maintains that the average number of characters or syllables per word, the average number of words in a sentence, and the percentage of words that are not among the most common words in a language are all related to the words, syntax, and semantic complexity of the text, respectively.

Mujiyanto (2016) asserts that most of the research studies have been done on the subject from perspectives such as goal setting, psychological aspects as well as socio-cultural aspects. According to DuBay (2004), in the 1980s, there were approximately 200 readability formulas and more than a thousand studies on the implementation of formulas and the validity of their statistical theories. Among the major formulas, Flesch Reading-Ease test, first introduced in the 1950s, is still known as one of the most widely used items, the purpose of which is to reveal the level of ease of reading certain texts. The formula uses the average sentence length and the number of syllables per word to determine legibility; in addition, higher scores indicate greater ease of reading while lower scores indicate that passages are more difficult (Mujiyanto, 2016).

In addition to tests, a number of tools have been used to determine the readability level of texts which include the Flesch-Kincaid Grade Level with four levels including: Gunning-Fog score, Coleman-Liau index, Simple Measurement of Gobbledygook index

(SMOG), and Automatic Readability Index (ARI). These specific criteria are based on the average number of syllables and the number of words per sentence (Mujiyanto, 2016). The Gunning-Fog score, created by Robert Gunning in 1952, estimates the years of formal education required to understand a text on first reading. However, the SMOG index- founded by McLaughlin in 1969 for the same purpose, is considered more accurate than other formulas. Depending on the factor that the characters in each word are calculated faster, ARI is basically made to produce an approximate display of US level requirements to understand a text (Mujiyanto, 2016). Still again, the Coleman-Liau index, designed by Coleman and Liau (1975) as in previous formulas, relies more on letters calculated by computer programs than syllables (Maksymski, Gutermuth, & Hansen-Schirra, 2015).

A number of studies have explored the benefits of these measures and indicators. For example, Evanciew and Jones (1996) evaluated several textbooks used in high school and advanced technology programs related to legibility including grades (baseline equivalents), human interest, and writing style. Kolahi (2012) used the Gunning-Fog index to measure the level of readability to show that Persian translations of English textbooks are less readable than English textbooks. Meanwhile, Wolfer (2015) stated that the main purpose of "readability studies was to develop formulas that could be used to directly measure the readability of a text using text surface properties such as the average length of words or sentences" (pp. 36-37).

2.2. Studies on Readability in Machine Translation

Regarding readability in machine translation, readability and simplification of the text have been extensively studied in the field of computational tools and several criteria and approaches have been proposed in the literature. According to Koehn et al. (2003), common readability criteria use universal text attributes such as ratio, type/symbol, lexical consistency, and ratio of long versus short words as indicators. Global features, such as those listed above, generally require new approaches to the SMT decryption problem. In the model implemented by Docent, SMT is based on the phrase (Koehn et al., 2003). Decryption uses a local search approach that involves the complete translation of a complete document at any time. The initial state is improved by applying the summit operation strategy to find the maximum

(local) score function (Koehn et al., 2003). The three operations used are changing the translation of phrases, changing the position of two phrases, and redistributing phrases.

There are also a number of studies which use SMT techniques for monolingual translation (e.g., Ganitkevitch et al., 2011) and sentence compression (e.g., Knight & Marcu, 2000; Specia, 2010). In addition, there is a wide range of publications that simplify monolingual sentences using other methods for compressing and writing sentences (e.g., Daelemans et al., 2004; Cohn & Lapata, 2009). Regarding the integration of extensive contextual features in machine translation such as lexical consistency, Carpuat (2009) and Carpuat & Simard (2012) investigated the effect of lexical consistency on translation.

The use of word ambiguity in SMT is another common example of including textual information in the source section (Carpuat & Wu, 2007; Chan et al., 2007). Readability has also been measured as a text summary effect by automated user criteria studies (Margarido et al., 2008) and automated benchmarks (Smith & Jönsson, 2011). These studies indicated that readability was generally better in abridged texts than in original texts.

In another study, Genzel et al. (2010) studied the translation of poetry, the composition of news text, and the maintenance of the form of poetry in the translation of poetry in which the researchers employed decoding features such as rhyme and meters and introduced restrictions on the output of the target language in order to adapt to specific task features. Also, Tiedemann (2010) suggested stored models to successfully create consistent translations on domain compliance.

Previous work in machine translation and NLP has focused on evaluating the readability and simplification of the text. Jones et al. (1997) examined the readability of MT and ASR system output humanely. As for the text simplification, Hardmeier et al. (2013) explained a document-level decoder for SMT and mentioned a case study that utilized document-wide features to improve the readability of text. Stymne et al. (2012) introduces document-level features such as type/token ratios and lexical consistency as input to the MT system. Contrary to Stymne et al. (2012), Xu et al. (2016) designed a new training objective for SMT text simplification.

In his study, Teich (2003) found that the German translation is more passive than the original German texts, which is probably due to the influence of the original English. A similar event was observed in this study for MT readability in outputs; MT tended to show variance based on ST, indicating that there was a wave of ambiguity for the same word.

Brück et al. (2007) recognized legibility and comprehensibility as two main components that compare the quality of source texts as well as texts in technical and scientific translations. In order to evaluate the quality of the formula, Ateşman Reading Ease (1997) was used in technical and scientific translation. The ease of reading formulas and indicators used in both studies showed that the desired texts and translated texts are at the level of ease of reading. Easy-to-read results showed that technical and scientific translation and literary translation are a real challenge for translators.

Koponen et al. (2012) examined human perceived effort during the translation process. It was found that sentences with less effort, as shown in higher manual scores or shorter editing time, tended to edit more in terms of word form and simply replace words from the same word group, while sentences with lower scores tended to longer edits and take longer in terms of word order, and correction of mistranslated terms (Koponen et al., 2012).

For the time being, there are different types of machine translation including FastDic, Microsoft Translation, Babylon, and Google Translate. However, it appears that the readability of text has not been considered a lot in their translations. In other words, text readability has scarcely been evaluated in machine translation. With the advancement of machine translation technology, the comparison of human functions is often generalized and exaggerated. Therefore, it is important to be able to calculate the differences correctly. Indeed, meagerness of a study to investigate readability criteria to the level of comprehensibility of source texts and their translations has triggered carrying out this research study. Thus, the main aim is to explore whether the ease of reading and the variables of the level of English texts and their reading are related to their kinds of translation, namely human and machine translation. In this vein, the present study's major interest lays in the readability analysis of translated text from English into Persian considering HT and MT. In other words, the main aim of this study is to compare the ability of HT and MT to translate

English to Persian literary texts and to explore the differences between the levels of item readability in Persian translation performed by HT and MT.

3. Methodology

3.1. Research Design

The present study was comparative-descriptive in nature. In this vein, the closest translation to the original text was evaluated according to the readability of the text. This was done using automatic readability checker comprising formulas like Flesch Kincaid Grade.

3.2. Corpus

The corpus of this study included the celebrated English book "*The Old Man and the Sea*" by Ernest Hemingway (1952) and its Persian translations by Nazi Azima (2009) and Google translate. In this study, two translations of the first twenty pages of the above-mentioned English book by Azima and Google Translate were considered. To make a comparison, after providing the translated version by Azima, Google Translate web-based translation service was consulted for rendition of the same parts. Next, both translations were compared for the level of readability by Flesch Kincaid.

3.3. Procedures

The first twenty pages of the novel "The Old Man and the Sea" in English as well as its Persian translation by Nazi Azaima were thoroughly read to be analyzed critically. Then, the same pages were fed to Google Translate web-based translation service for translation to compare the readability of both translated texts. Eventually, Flesch-Kincate Grade was used to calculate the readability of the Persian texts, as Flesch scores can indicate the readability of a text. The Flesch Reading Ease score is between 1 and 100, and the Flesch Kincaid score level represents the US education system. Also, the Flesch Reading Ease score was used to calculate the English text.

It should be noted that the ease/difficulty of reading is assessed by the reader's mental criteria. Moreover, the assessment of ease of reading is different in each language. For example, Flesch Ease Reading was used in this study for Persian; however, more than one

formula is used in English for assessing the ease of reading, (which in this study was Flesch-Kincate). Flesch Kincaid Grade Level is a broad readability formula that assesses the approximate level of text readability. Previously, the Flesch Reading Ease score had to be converted to become a reading level. A modified version was used in the 1970s for ease of use. The Navy used it for its technical guidance used in training. It is now used for a much wider variety of applications. If a text has level 8 Flesch Kincaid, it means that the reader needs level 8 or higher to understand it. Even if they are an advanced reader, they have less time to read the content.

Considering that the readers' mental criterion is one of the most important factors in assessing the ease of reading, the ease/difficulty of reading a text is different for each individual reader. In this study, the dependent samples t-test statistical procedure was used to measure different levels of readability. As it was expected that the results of the t-test procedure could indicate different levels of readability, these results were explained inferentially. The Wilcoxon test was used for this inferential explanation to test the location of samples and to compare the locations of two populations using method samples.

4. Results

4.1. Addressing the First Research Question

In order to provide response to the first research question, Flesch Reading Ease Formula was used to compare the readability of English texts with machine translation and human translation. The readability ease of English texts was measured using Flesch Reading Ease Formula and the readability of Persian texts was measured using Flesch-Kincaid Grade Level. The obtained results are described in detail in Table 1.

According to Table 2, there is almost full equivalence between the source texts and target texts as far as the reading ease is concerned (90 to 100: very easy to read). Also, it can be inferred from the table that Flesch-Kincaid Reading Ease formula can be applied to measure the ease level of the target texts in Persian.

Table 1

Comparison of Reading Degree of Ease of Levels of Source Texts (ST) (English) and Targets Texts (Persian MT versus HT)

Criterion	Text	Language	Level
Flesch Reading Ease	ST	English	17.70
Flesch-Kincaid Grade Level	MT	Persian	20.92
Flesch-Kincaid Grade Level	HT	Persian	25.67

Table 2

Comparison of Flesch Reading Ease Scores and Flesch-Kincaid Scores

Interpretation of Flesch Reading Ease Scores	Interpretation of Flesch-Kincaid Scores
100.00-90.00: Very easy to read	90 – 100: Very easy to read
90.0–80.0: Easy to read	70 – 89: Easy to read
80.0–70.0: Fairly easy to read	50 – 69: Fairly difficult to read
70.0–60.0: Easily understood	30 – 49: Difficult to read
60.0–50.0: Fairly difficult to read	1 – 29 : Very difficult to read
50.0–30.0: Difficult to read	
30.0–0.0: Very difficult to read	

A comparison of Flesch Reading Ease Formula and Flesch-Kincaid Reading Ease Formula was made using Wilcoxon Test to compare the results inferentially. The results are presented in Table 3.

Table 3

Comparison of Flesch Reading Ease Formula and Flesch-Kincaid Reading Ease Formula (Wilcoxon Test)

	Mean	Std. Deviation	Z	P
Flesch Reading Ease Formula	18.31	17.982		
				-.307[b]
Flesch-Kincaid Reading Ease Formula	19.10	14.547	0.021	

Based on the results of Wilcoxon Test in Table 3, it was confirmed that as far as they measure reading ease of the literary translation, Flesch Reading Ease Formula and Flesch-Kincaid Reading Ease Formula are compatible with each other ($p = 0.021$).

4.2. Addressing the Second Research Question

In order to address the second research question, Independent Samples t-test was used to measure the comprehensibility levels of the MT and HT for independent samples. The comparison of the readability levels of the MT and HT are presented in Table 4.

Table 4

Comparison of the Readability Levels of the MT and HT

Readability Level	N	Mean	Std. Deviation	t	P
MT	101	40.13	24.15		
HT	101	29.25	22.37	-4.571	0.001*

According to the results of Table 4.4, there is a significant difference between the readability levels of the MT and HT ($p<0.05$). Thus, it can be concluded that the comprehensibility levels of HT is significantly lower than MT.

5. Discussion and Conclusion

Regarding the first research question, the results show that despite slight differences in the readability levels of both software, they are good references to compare the differences between SL and TL. Also, the results of Wilcoxon test confirm the compatibility of these two measures with each other. Moreover, the reading ease levels of texts decline as we move from source text in English to MT and HT in Persian, respectively. This can reveal the quality of long prose that the human translators tend to apply in order to preserve the language art and transfer the artistic tone and impact on the reader, while MT might solely be confined to transferring the meaning, using shorter sentences and more simple words.

The results related to the second research question reconfirm the same process with regard to the readability in Persian MT and HT, even though HT demonstrated more perplexity in comparison with MT. Accordingly, it seems that HT tends to act more independently in translation of ST. On the other hand, MT seems to be more dependent on the ST and does not modify it based on different relevant factors such as culture, intended meaning, etc. Moreover, this result might also originate from a lack of critical attitude in MT

towards the text and its translation process. In this vein, special techniques are required for these tasks to train prospective translation machines. Hence, it seems that terminological variations are regarded more meticulously concomitant to the process of critical thinking of the translator in HT, whereas this step is highly ignored in MT.

Brück et al. (2007) compared the readability and comprehensibility as two main components that clarify the quality of source texts as well as target texts to assess the quality of Ateşman Reading Ease formula in technical and scientific translation. Although they have not followed a similar path, the results of this study can be compared with their study in that in both studies, the Reading Ease formulas proved to be effective and accurate in evaluating the equivalence between the source texts and target texts. The Reading Ease and indices formulas employed in both studies revealed that both the target texts and the translated texts are over the general reading ease levels. The reading ease results indicated that technical, scientific, and literary texts translation poses a real challenge to the translators.

In relation to the degree of editing, Koponen (2012) and Koponen et al. (2012) examined human perceived effort during the process of translation. She understood that sentences engaging less effort, as demonstrated by higher manual grades or shorter editing times, were inclined to engage more edits in relation to word forms and simple replacements of words of the same word group, while sentences with low grades or long editing times engaged more edits in relation to word order, edits where the word group was altered, and modifications of mistranslated idioms (Koponen, 2012; Koponen et al., 2012). The results of the current research study are in line with Koponen's (2012) findings in that the same problem of micro level and macro level editing was encountered in her study; hence, these factors should be taken into consideration more cautiously. However, the focus of this study was mainly on the readability of MT output and its comparison with HT, while Koponen (2012) studied the concept of human editing more deeply.

Schulz, Bernhardt-Melischnig, Kreuzthaler, Daumke, and Boeker (2013) compared three different kinds of SNOMED CT translations from English to German with the use of specialized medical translators; a free web-based machine translation service; and medical students. Average ratings of linguistic accuracy did not vary meaningfully between HT

situations. Their results are consistent with the current study. In comparing MT to HT, the linguistic accuracy differed in support of the HT and concerning content fidelity, it differed similarly in support of the HT.

In a similar vein, Yamada (2014) assessed the quality of Google Statistical Machine Translation (SMT) by examining college language students' post-editing (PE) performance. Based on his findings, learners missed about 7 errors uncorrected in their final outputs. The data also revealed only a loose correlation between the learners' general translation ability and their post-editing performance. While learners who had weak marks in a traditional translation course were indicated to be unqualified post-editors, learners who achieved good marks were not always qualified post-editors either (Yamada, 2014). In the same vein, the existence of errors in MT is highlighted in the present research study. Accordingly, the proposed notion remains the same in both studies in that human translators have an important role in the translation process that should not be taken for granted and this accuracy of HT is better to be transferred to educational systems for training specialized translator machines.

Daems et al. (2015) conducted a post-editing study for general text types from English into Dutch accomplished by masters' students of translation. They recognized six various post-editing effort indicators and understood that various types of MT errors anticipated various post-editing effort indicators. However, they investigated different factors compared to the current study, a notion is similar in both studies; the highlighted matter is remained errors as an influence of MT products. Therefore, readers and translators shall recognize all potential errors in MT outputs in order to yield more qualified products.

In another related study, Comparin and Mendes (2017) revealed the findings of a research containing an error annotation task of a body of machine translations from English into Italian. They compared error types discovered in raw MT and post-edited texts. They also recognized frequent and critical errors and perceived the errors' occurrence at different phases of the translation procedure. One relevant result was that 85% of the errors discovered in the raw MT were correctly edited through human post-editing; also, fluency errors reduced, but a moderately high number of accuracy errors were not revised. In comparison with the results of the current study, their findings denoted the influence of MT on PE

products, highlighting the unrevised errors that requires a more critical attitude of the human translators. In a similar vein, Ahrenberg (2017) highlighted the limitations of MT outputs in the context of English and Swedish languages and claimed that PE products are not even comparable with HT in respect of quality.

Acar and İşisağ (2017) conducted a study on the readability and comprehensibility levels of technical and scientific texts in English and their Turkish translations using Flesch Reading Ease, Gunning Fog, Flesch-Kincaid Grade Level, the Coleman-Liau index, the SMOG index, Automated Readability index, and Linsear Write Formula. They compared and contrasted two components (i.e., the readability and comprehensibility levels of technical and scientific texts) in the target language and the source text. They applied Atesman Reading Ease Formula to estimate the reading ease of the translated texts in Turkish and they measured the comprehensibility levels of the source texts and target texts by using a checklist, including source texts and corresponding questions that were directed to 43 English teachers. One text was translated using Google translation. Their results can be compared with the present study in using Flesh Reading Formula which was found compatible to a Turkish Reading Ease Formula. In the same vein, this formula is found to be compatible with Persian Reading Ease Formula. However, they compared the comprehensibility levels of the source texts and the target texts that were found to be higher than the readability of the texts. The comprehensibility of the target texts was also discovered to be higher than that of the source texts. Moreover, a statistical difference was met between the readability and comprehensibility levels of the texts. Google translate had the lowest comprehensibility level.

Čulo, Hansen-Schirra, and Nitzke (2017) carried out a research study in which they contrasted post-editing and human translation with regard to the aspect of term translation within the field of LSP. They used the perplexity coefficient to analyze terminological variation in term translation from English to German. Relating to the achieved results of the present study, this requires a critical approach toward the outcomes of MT. It also tackles translators' community to have a more cautious use of the updated technologies.

Based on the results of the current study, there are different levels of ease of reading in each text, which shows the lowest to highest degree of ease in the source text, MT text and

HT text, respectively. Moreover, according to the obtained results, there is almost a perfect equivalent between the source and target texts as far as ease of reading is concerned (90 to 100: very easy to read). It can be inferred that the Flesch-Kincaid Reading Ease formula can be used to measure the ease level of the target texts in Persian.

To measure different levels of readability between machine and human translations, Independent Samples t-test was run and the results showed that there is a significant difference between MT and HT readability levels ($p < 0.05$). In this regard, the level of HT perception is significantly lower than MT. Of course, in this case only one language was examined in two translation models, which could be considered a potential downside of the current study. Another potential drawback of this study is the small sample of the study. The generalizability of the evaluations can be tackled by this drawback. Also, the assessment method was mainly based on the readability methods and did not consider other modes and dimensions of translation which could be compared and contrasted. Hence, the results are highly dependent on the method through which MT products were provided solely.

Assessing the readability of the text can affect the quality of text context. This means that the use of specific words and destination grammars, especially in the form of poetry, indicate the level of knowledge of student and the quality of the educational system. So it can be concluded that this study can affect curriculum planning, especially in the field of translation. This research study can also be effective in evaluating students in the field of translation as a translator. This means that another benefit of this research is the familiarity of students with their strengths and weaknesses in translation.

Along with the results of this study, an important question will be asked in the field of empirical corpus-based translation research. The unsolved issue is whether all types of texts tend to exhibit similar MT results concerning the linguistic properties or they may differ based on the subject matter. Therefore, a staunch need appears to further shift into translation studies in order to attain a comprehensive insight of the potential potpourri between MT and HT. Some modes of MT might be more compatible with target language; while, some other modes may display a contrary outcome. However, this notion cannot be judged with this small-scale study and requires deeper investigations.

In line with the above-mentioned results, another issue that should be taken into consideration is taking a step toward the progression of online translation tools in order to produce more qualified outputs and to ease the translation process. If MT proliferates rapidly as a prevalent method in the translation world and demonstrate similar texts as ST in different procedures of translation, the produced features will be narrowed down to a constant reproduction of the same being in the long run. In other words, this event leads to a linguistic flatter. In this regard, the potential question is whether literary texts would undergo any changes considering the continuous exchange of MT used by the related readers. Due to the limitations in the generalizability of the results, the results of this study cannot be overstated and hence further generalization requires more extended and fuller in-depth research.

Conflict of interest

The author(s) certify/certifies that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in the present research paper.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

References

Acar, A., & İşisağ, K. U. (2017). Readability and comprehensibility in translation using reading ease and grade indices, *International Journal of Comparative Literature and Translation Studies*, *5*(2), 47-53. DOI: https://doi.org/10.7575/aiac.ijclts.v.5n.2p.47.

Ahrenberg, L. (2017). Comparing machine translation and human translation: A case study. Paper presented at *Proceedings of the Workshop on Human-Informed Translation and Interpreting Technology*, Varna, Bulgaria. DOI:10.26615/978-954-452-042-7_003

Ateşman, E. (1997). Türkçe'de okunabilirliğin ölçülmesi. A.Ü. *Tömer Dil Dergisi, 58*, 171-174.

Brück, T., Hartrumpf, S., & Hermann Helbig, H. (2008). A readability checker with supervised learning using deep syntactic and semantic indicators. In *11th International Multiconference*: Information Society-IS, pp. 92–97.

Carpuat, M. (2009). One translation per discourse. In *Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions (SEW-2009)*. Stroudsburg: Association for Computational Linguistics: 19-27. Retrieved February 20, 2023 from: http://aclweb.org/anthology-new/W/W09/W09-2404.pdf.

Carpuat, M., & Simrad, M. (2012). The trouble with SMT consistency. Paper presented at *Proceedings of the 7th Workshop on Statistical Machine Translation*. Stroudsburg: Association for Computational Linguistics: 442-449. Retrieved February 20, 2023 from: http://aclweb.org/anthology-new/W/W12/W12-3156.pdf.

Chall, J. S. (1958). Readability: *An appraisal of research and application*. Columbus: Bureau of Educational Research, Columbus, Ohio, USA.

Cohn, T. and Lapata, M. (2009). Sentence compression as tree transduction. *Journal of Artificial Intelligence Research, 34*, 637–674. DOI:10.1613/jair.2655

Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology, 60*(2), 283–284. https://doi.org/10.1037/h0076540

Comparin, L., & Mendes, S. (2017). *Using error annotation to evaluate machine translation and human post-editing in a business environment.* Paper presented at the Proceedings of EAMT, Prague, Czech.

Crosbie, T., French, T., & Conrad, M. (2013). Stylistic analysis using machine translation as a tool. *International Journal for Infonomics (IJI)*, *1*(1). Retrieved June 25, 3023 from http://uobrep.openrepository.com/uobrep/handle/10547/333145.

Čulo, O., Hansen-Schirra, S., & Nitzke, J. (2017). Contrasting terminological variation in post-editing and human translation of texts from the technical and medical domain. *Empirical Translation Studies: New Methodological and Theoretical Traditions*, *300*(1), 183-206. DOI:10.1515/9783110459586-007.

Daelemans, W., Höthker, A., and Sang, E. T. K. (2004). Automatic sentence simplification for subtitling in Dutch and English. In *Proceedings of the 4th International Conference on Language Resources and Evaluation* (LREC'04), pages 1045–1048, Lisbon, Portugal.

Daems, J., Vandepitte, S., Hartsuiker, R., & Macken, L. (2015). The impact of machine translation error types on post-editing effort indicators. Paper presented at *4th Workshop on Post-Editing Technology and Practice* (WPTP4), Miami, Florida. http://hdl.handle.net/1854/LU-6990271.

DuBay, W. H. (2004). *The principles of readability*. Costa Mesa, California: Impact Information. Retrieved August 12, 2022 from http://www.impact-information.com/impactinfo/readability02.pdf.

Evanciew, C. E. P., & Jones, K. H. (1996). Using readability, human interest, and writing style to evaluate technology education textbooks. *Tech Trends*, *41*(2), 37-38. http://dx.doi.org/ 10.1007/BF02818816.

Ganitkevitch, J., Callison-Burch, C., Napoles, C., & Durme, B. V. (2011). Learning sentential paraphrases from bilingual parallel corpora for text-to-text generation. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 1168–1179, Edinburgh, Scotland. https://aclanthology.org/D11-1108.

Genzel, D., Uszkoreit, J., & Och, F. (2010). "Poetic" statistical machine translation: Rhyme and meter. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pp. 158–166, Cambridge, Massachusetts, USA. https://aclanthology.org/D10-1016.

Hardmeier, C., Tiedemann, J., & Nivre. J. (2013). Latent anaphora resolution for cross-lingual pronoun prediction. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pp. 380–391, Seattle, Washington, USA. Association for Computational Linguistics.

Jones, K. H. (1997). Analysis of readability, interest level, and writing style of home economics textbooks: Implications for special need learners. *Journal of Vocational Home Economics Education, 12*(2), 13-24. Retrieved June 12, 2021 from http://www.natefacs.org/Pages/v12no2/12-2-13%20Jones.pdf.

Klare, G. R., Nichols, W. H., & Shuford, E. H. (1957). The relationship of typographic arrangement to the learning of technical training material. *Journal of Applied Psychology, 41*(1), 41–45. https://doi.org/10.1037/h0046945.

Knight, K., & Marcu, D. (2000). Statistics-based summarization — Step one: Sentence compression. In *National Conference on Artificial Intelligence* (AAAI), pp. 703–710, USA, Austin, Texas.

Koehn, P., Och, F. J., & Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Human Language Technology Conference of the NAACL*, pp. 48–54, Edmonton, Alberta, Canada. DOI:10.3115/1073445.1073462.

Kolahi, S., & Shirvani, E. (2012). A comparative study of the readability of English textbooks of translation and their Persian translations. *International Journal of Linguistics*, *4*(4), 344–361. https://doi.org/10.5296/ijl.v4i4.2737.

Koponen, M. (2012). *Comparing human perceptions of post-editing effort with post-editing operations.* Paper presented at *the 7th Workshop on Statistical Machine Translation*, Montreal, Canada.

Koponen, M., Aziz, W., Ramos, L., & Specia, L. (2012). Post-editing time as a measure of cognitive effort. Paper presented at *the Workshop on Post-editing Technology and Practice, San Diego, USA.*

Maksymski, K., Gutermuth, S., & Hansen-Schirra, S. (Eds.). (2015). *Translation and comprehensibility*. Berlin: Frank & Timme Verlagfürwissenschaftliche Literatur.

Margarido, P., Pardo, T., Antonio, G., Fuentes, V., Aluísio, S., & Fortes, R. (2008). Automatic summarization for text simplification: Evaluating text understanding by poor readers. In *Anais do VI Workshop em Tecnologia da Informação e da Linguagem Humana*, pp. 310–315, Vila Velha, Brazil. DOI:10.1145/1809980.1810057.

Maria A. C., & Dinu. L (2014). A Quantitative Insight into the Impact of Translation on Readability. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 104–113, Gothenburg, Sweden. Association for Computational Linguistics. DOI:10.3115/v1/W14-1212.

McDonald, S. V. (2020). Accuracy, readability, and acceptability in translation. *Applied Translation*, *14*(2), 21–29. https://doi.org/10.51708/apptrans.v14n2.1238.

Mühlenbock, K., & Kokkinakis, S. J. (2009). LIX 68 revisited – an extended readability. In *Proceedings of the Corpus Linguistics Conference*, Liverpool, UK.

Mujiyanto, Y. (2016). The comprehensibility of readable English texts and their back-translations. *International Journal of English Linguistics*, *6*(2), 21. https://doi.org/10.5539/ijel.v6n2p21.

Rezaee, A. A., & Norouzi, M. H. (2011). Readability formulas and cohesive markers in reading comprehension. *Theory and Practice in Language Studies*, *1*(8), 1005–1010. https://doi.org/10.4304/tpls.1.8.1005-1010.

Schulz, S., Bernhardt-Melischnig, J., Kreuzthaler, M., Daumke, P., & Boeker, M. (2013). sMachine vs. human translation of SNOMED CT terms. In *MEDINFO 2013 - Proceedings of the 14th World Congress on Medical and Health Informatics* (1-2

Aufl., S. 581-584). (Studies in Health Technology and Informatics; Band 192, Nr. 1-2). https://doi.org/10.3233/978-1-61499-289-9-581.

Smith, C., & Jönsson, A. (2011). Automatic summarization as means of simplifying texts, an evaluation for Swedish. In *Proceedings of the 18th Nordic Conference on Computational Linguistics* (NODALIDA'11), Riga, Latvia.

Specia, L. (2010). Translating from complex to simplified sentences. In *Proceedings of the 9th International Conference on Computational Processing of the Portuguese Language*, 9th International Conference (PROPOR'10), pp. 30–39, Porto Alegre, Brazil.

Stephens, C. (2000). All about Readability. Retrieved February 2, 2024 from http://www.plainlanguage.com>newreadability.

Stolcke, A. (2002). SRILM – an extensible language modeling toolkit. In Proceedings of the *7th International Conference on Spoken Language Processing*, pp. 901–904, Denver, Colorado, USA.

Stymne, S., & Smith, C. (2012). On the interplay between readability, summarization and MTranslatability. In *Proceedings of the 4th Swedish Language Technology Conference*, pp. 71–72, Lund, Sweden.

Teich, E. (2003). *Cross-linguistic variation in system and text: A methodology for the investigation of translations and comparable texts*. Berlin/New York: Mouton de Gruyter. DOI:10.1515/9783110896541.

Tiedemann, J. (2010). Context adaptation in statistical machine translation using models with exponentially decaying cache. Paper presented at *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing (DANLP)*. Stroudsburg: Association for Computational Linguistics: 8-15. Retrieved April 25, 2023 from: http://aclweb.org/anthology-new/W/W10/W10-2602.pdf.

Wolfer, S. (2015). Comprehension and comprehensibility. In K. Maksymski, S. Gutermuth, & S. Hansen-Schirra (Eds.), *Translation and comprehensibility* (pp. 33-52). Berlin: Frank & Timme Verlagfürwissenschaftliche Literatur.

Yamada, M. (2014). Can college students be post-editors? An investigation into employing language learners in machine translation plus post-editing. *Machine Translation, 29*(1), 49–67. DOI:10.1007/s10590-014-9167-7.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی