Machine versus human translation outputs: Which one results in better reading comprehension among EFL learners?

Mojtaba Maghsoudi*

Farhangian University, Tehran, Iran
maghsudi@cfu.ac.ir

Vahid Mirzaeian

Alzahra University, Tehran, Iran
Mirzaean@alzahra.ac.ir

* Corresponding author

Regular Paper

EFL university students in general and Iranian EFL university learners in particular, specifically in the commencement of their study at bachelor degree prefer to have their assigned English texts translated into mother tongue since they do not have a good command of the language. Machine translation (MT) has recently turned into a favorite tool for all students including EFL ones since it is free and readily available. The aim of this study, therefore, was to evaluate the effect of MT output on such students. To achieve this goal, two sample texts in students' mother tongue were selected and translated into English by a proficient human translator whose native language was English and by an MT system (Google Translate). Next, 167 homogenous EFL freshman students were given a proficiency test and 152 homogenous ones randomly assigned to two groups, namely, control and experimental groups. The control group was given the human translation (HT) and the experimental group was exposed to machine translation. Using an independent t-test, indicated that there was a negligible difference between mentioned groups. That is, MT has improved to such an extent that it can compete with the

HT. It can be concluded that due to its great improvements in the last few years and its wide-spread use among university students, EFL educators should accept the presence of this tool and try to implement it effectively in their teaching instead of banning students from using it.

Keywords: Text analysis; reading comprehension; machine translation; human translation

Introduction

The link between translation and Foreign Language (FL) reading comprehension has a time-honored history in English as a Foreign Language (EFL) tradition, as old as the Grammar-Translation Method and has been investigated from different aspects (Larsen-Freeman & Anderson, 2013). It has also been constructively blended with innovative methods in the field of EFL, take for example, Community Language Learning (Larsen-Freeman & Anderson, 2013), and also creatively mixed with communicative approaches to language teaching (Goodwin & Jiménez, 2016).

The long-lasting presence of translation in EFL reading classes may be attributed to both learners' positive attitude toward it as well as their beliefs regarding its effectiveness (Liao, 2006). The EFL teachers' adherence to translation in reading comprehension classes is ignited by both the research background supporting the technique as well as EFL learners' positive attitude toward it (Boshrabadi, 2014). It can also explain why it has been widely used in English for Special Purpose (ESP) classes and reported to be promising in Iranian technical universities (Avand, 2009).

However, implementation of translation to enhance reading comprehension does not seem to be fruitful among Iranian freshman EFL students since they either ask human translators to translate the passages into mother tongue for them or use readily available machine translation (MT) systems. With the wide-spread use of MT among university students in general (Chang & Yamada, 2020) and freshman EFL students in particular (Mirzaeian, 2020) in the Iranian academic context, it seems that translation has regained its popularity among university students necessitating a new look at foreign language reading comprehension from translation perspective as well as study of MT output in terms of usability and comprehensibility among EFL freshman learners as touched by in this study.

This study is not devoted to the usability of translation in reading comprehension, rather to the MT output usefulness to EFL freshman learners as long as its comprehensibility is concerned. In academic settings as witnessed today, an increasing number of students are using this free MT technology for personal and educational purposes. Theoretically speaking, MT has proven to be useful to students from both linguistic and cognitive point of view. For instance, a researcher (Chung, 2020) contended that exposing students to a translated draft of foreign language text could drastically reduce the cognitive load of the text and result in text comprehension among beginners without resort to help from peers or instructors.

MT is currently used by almost anyone inside or outside university campus; however, majority of the studies of Machine Translation Assisted Language Learning (MTALL) have been devoted to the use of MT in foreign language writing and a few studies have investigated MTALL effect on FL reading comprehension. Therefore, studies of this kind should be conducted to provide us with a detailed knowledge of what is happening among students when exposed to MT output.

Recently some researchers (Mirzaeian, 2020; Murphy, 2020 & Hsu, 2019) considered

the rising interest among university students with regard to MT; however, more studies should be conducted on MT and reading comprehension. It can be concluded that there is no evidence to prove if MT output is comparable to HT output as far as FL reading comprehension is concerned. In other words, when exposed to MT output, do students experience the same effect as exposed to HT output?

The extent of similarity between the two can inform us on how to implement this tool with regard to educational purposes in general and FL learning in particular. It has to be noted here that there have been few attempts to probe into the effectiveness of MT as far as improving reading comprehension of students with insufficient foreign language proficiency is concerned. MT was criticized for its inability to provide an acceptable output; however, with the recent advancements, the improved quality is noticeable and lack of empirical evidence has pushed language policy makers to ignore the usefulness of this tool.

This study, therefore, probes into the contribution of MT to freshman FL students who lack enough proficiency to follow up their studies in the target language. Inspired by the aforementioned rationales, this study measures EFL students' reading comprehension regarding MT output versus HT output. The following research questions have been formulated for this study:

1. How is the quality of the texts generated by MT?
2. Do the students who read the MT output and those who read the human translation, perform equally in reading comprehension?

Literature review

The majority of the studies on the use of MT for L2 learning have focused on improving learners' writing skill in the target language. For example, Amaral & Meurers (2011) concluded that MT could improve learners' writing by increasing lexical, syntactic, semantic and pragmatic knowledge, raising linguistic awareness and finally providing linguistic modeling. Correa (2014) showed that metalinguistic awareness could be improved by MT resulting in an improvement in the writing process and affecting students' final written output. In the same vein, Nino (2008) showed that to improve students' writings, paraphrasing skills should be taught to students and MT can be of great assistance in this regard. MT was also found to be of assistance to students who want to provide better written output (Garcia & Pena, 2011; Ali & Alireza, 2014). MT was shown to enhance learners' writing by scaffolding their learning (Godwin-Jones, 2015).

Despite numerous attempts employing MT for writing instruction in L2 classes, reading comprehension instruction with the aid of MT has not been researched sufficiently and there is a poor body of knowledge about using MT to teach reading comprehension to L2 learners in English for Specific (ESP) classes, in particular, where the main objective of reading instruction is to teach students how to read efficiently and quickly enough to extract meaning from the given text rapidly (Rushwan, 2017).

Reading instruction in ESP programs are of specific significance since it enables learners to read English texts for academic, and occupational purposes which consequently influences their future achievements. There is enough evidence supporting the promising role of translation in ESP reading instruction success (Mahmoud, 2006) due to its role in developing three essential qualities to language learning: accuracy, clarity, and flexibility as well (Ross, 2000). Accordingly, reading instruction in ESP classes at Iranian universities is mostly based on translation activities (Iranmehr, Atai, & Babaii, 2018). However, due to

the discrepancies in policy and practice, ESP programs, in general, and reading instruction, in particular, are docile in terms of significantly improving the learners' proficiency (Iranmehr, Atai, & Babaii, 2018).

The inefficiency of ESP programs in Iranian higher education centers and infertility of reading instruction even the translation-based ones in these programs have been echoed frequently in recent evaluations of these courses. To enumerate a few, Tavakoli and Tavakol (2018) discovered political, sociocultural and ethno-religious reasons of the dissatisfaction among the ESP education stakeholders regarding. Meanwhile, Abdelaal & Alazzawie (2020), highlighting the ESP stakeholders' perceptions about the significant role of reading comprehension as the main need of EAP students, highlighted problems raised by "students and teachers regarding the sources, materials, students' general English proficiency level, duration and timing of the classes, motivation level of both students and teachers, etc. which are discussed in the paper."

Although higher education students are in dire need of improving their reading comprehension skills to infer technical English texts (Amerian, & Marefat 2019), ESP programs at higher education centers are unable to meet their needs due to boring classes, outdated materials, non-proficient students, crowded classes, lack of technological aid (Boniadi, Ghojazadeh & Rahmatvand 2013), reading passages with inappropriate readability index (Mohebbi, Nayernia, Nemati & Mohebbi, 2017), contextual factors, improper policies, unprincipled teaching methodology, and old-fashioned spiritless textbooks (Mostafaei Alaei & Ershadi, 2017). These factors lead frustrated higher education students to abandon improving their ESP reading skills and shift to HT or MT for their immediate needs to comprehend a bulk of English texts.

In foreign/second language reading comprehension in particular, there have been attempts since as early as the 1960s; for example, Pfafflin (1965) described an experiment in which the participants were presented with MT texts, reporting that human translations were insignificantly perceived to be of higher quality by the participants. Despite the scarce studies done in this field in the fifty years since then, it has continued to interest researchers, particularly in the past two decades (Mirzaeian, 2020). For example, the practicality of using reading comprehension questions to evaluate MT has been recurrently addressed in previous studies. Jones *et al.* (2005a), for instance, used a test of language proficiency structure to assess the readability of MT texts. They found that participants were slower and less accurate at answering comprehension questions on the MT documents compared to human translated documents. In another study, Jones *et al.* (2005b) aimed to discover which level of reading comprehension participants could achieve by reading an MT document. They found that MT documents brought about an intermediate level of comprehension. They also showed that automatic assessment methods, such as BLEU are not satisfactory to distinguish between quality levels of MT documents translated by the same MT system. In a later study, Berka *et al.* (2011) used a quiz-based approach to evaluate MT using a set of yes/no questions human translated and MT texts from four different MT systems, which consisted of small paragraphs from different domains. They found that different texts from various MT systems brought about different degrees of accuracy in the participants' answers.

MT can be helpful to students from an affective perspective, too. Several studies (Nino, 2008, 2009; Kliffer, 2008; Jin, 2013; Bahri & Mahadi, 2016) have shown that MT can help provide a nonthreatening learning environment for the students. It has been argued that

using MT in post editing can improve students' participation and motivation and consequently increase confidence in foreign language learning.

In spite of advantages mentioned for MT, some disadvantages have also been reported by some studies. The MT quality highly depends on subject area, text type, size as well as language pair which imposes limitations on the quality of the output generated by MT (Godwin-Jones, 2015; Holt, Overgaard & Engel, 2020). Some studies (Shei, 2002a, 2002b) have shown that there are limitations regarding MT pre-editing at cultural, lexicosemantic and idiomatic levels. In addition, if the input is weak or ill-formed, it may result in an erroneous output full of grammatical as well as lexical errors. It has proved not to be successful to provide grammatical solutions to students' grammatical errors (Josefsson, 2011). However, it has to be noted that, MT has been glamorously promising in terms of translating expository texts, such as technical ones the higher education students have to study (O'Brien, Simard & Goulet, 2018).

Due to rapid changes in the field, MT output tends to improve gradually, so its implementation will also increase in language learning contexts. Therefore, more studies are needed to shed light on dark areas in the use of MT for language learning. However, since authorities in the field of language teaching are not so optimistic regarding the use of MT in language learning (Barr, 2013), not many researches are conducted in this field which can be called, Machine Translation Assisted Language Learning (MTALL). Since majority of the previous studies have focused on MT reliability, they have claimed MT to be an unsuitable language model. In most studies, student's learning was not measured after using MT and the analysis was limited to surveys from language learners using MT (Case, 2015).

Method

Participants of the study

In total, 167 Iranian EFL freshman university students whose average age was 21 were randomly invited to this study. They were selected from three universities of Arak, Iran in which EFL courses were offered at bachelor level. They were all given the Quick Placement Test to probe into their current level of language proficiency. They were all given Quick Placement Tests to measure their English language proficiency level. Those whose scores were one SD below and above the mean score were selected as the homogeneous ones for further study.

Materials and instruments

English proficiency test. To ascertain the homogeneity of the participants in terms of language proficiency, the Quick Placement Test (second version) was utilized. It is a standardized 50-item multiple-choice test which consists of grammar, vocabulary, and reading subsections. The entire Quick Placement Test lasted 40 minutes. The reliability (Cronbach's alpha) of the test was 0.82.

Reading comprehension test. Students who had been divided into experimental and control groups were instructed to read the allocated texts and answer reading comprehension questions followed. As mentioned before, the texts were in students' mother tongue and were translated into English by Google Translate as well as a human translator who was proficient in students' mother tongue and whose native language was English.

73

The aim of the test was to measure the level of comprehensibility of the MT and HT outputs. The original passages were meticulously selected and their level of difficulty checked to ensure their relevance to students' current level of proficiency. This was conducted in order to avoid the effect of background knowledge as well as prior familiarity with technical terms on students' performances. This test including two passages consisted of 25 multiple-choice items, assessing the participants' literal comprehension of information stated in the passage as well as higher order comprehension that required making inference and conclusions. It has to be noted that the questions and the contents were identical for both experimental and control groups. The only difference was the translator.

Semi-structured interview. Thirty students out of 167 homogenized (i.e., with similar English language proficiency) students were invited to take part in this part of the study. It is worth mentioning that the students participating in interview sessions were selected based on convenient sampling; that is, those who had frequently experienced MT and were also willing to voluntarily attend the interview sessions were invited. Several relevant and general questions were given to three expert colleagues to make sure through which the needed patterns may be extracted. The needed time for each interviewee was approximately 15 minutes.

Procedure. The researchers gained consent from the university authorities as well as participants for the study. The instruments were given to the subjects over a two-day period. The ELT proficiency test and the reading comprehension tests were administered in one day. The semi-structured interviews were conducted in another day with a one-week interval between the two. The rationale behind administering the proficiency test was that numerous studies have confirmed the effect of FL proficiency on reading comprehension. To ensure the students' English proficiency would not interfere with their reading comprehension, participants with similar degree of proficiency had to be selected for the current study. This would guarantee that the difference gained in the scores could be attributed to the comprehensibility of the outputs and not proficiency level of the students. The whole study was conducted in two phases as described below.

In the first phase, the proficiency test (Quick Placement Test, second version) was administered to all participants so that homogenous students could be identified. As a result, those who scored 1 SD below or above the mean score were selected for the study. As a result, 152 students remained after these students were excluded. An interview guide was piloted with three students who were not among the participants. The questions were created eliciting patterns based on typology. Three colleagues from the ELT department were asked to make comments regarding the interview questions.

In the second phase, semi-structured interviews with the refined questions were conducted with 30 students. Having received participants' consent to record the interviews, the researchers recorded them. The classification of errors to be generated out of interviews was based on a condensed typology by Popovic (2018). Based on his classification, errors are split into categories as: omission, syntactic error, mistranslation, addition, register, idioms, and punctuation.

Omission, as the name suggests, refers to lexical items present in the original text and absent in the translation. Syntactic errors are abundant in MT output; however, their frequency has decreased drastically in the past few years. The third category, mistranslation, is still a problem in MT since lexical items have different meanings and it is the context

that identifies the correct meaning. This is very easy to do for a human being and highly difficult for an MT system. Addition refers to the insertion of an unnecessary lexical item. Register refers to cases in which the translation is both syntactically and semantically acceptable; however, the translation does not seem to be acceptable for the given context, that is, the degree of formality versus informality. Idioms, groups of words whose meaning is different from the meaning of individual words, also are problematic for MT systems. If extra punctuation marks are either inserted or deleted in the output, this is considered a punctuation error.

Before the actual experiment was conducted, the instrument was piloted to ensure its simplicity, clarity, time allotment, and reliability which was calculated using Cronbach's alpha, achieving a score of 0.85. In addition, to ensure the instrument's efficacy to cover the content it was supposed to measure, three members of the ELT department evaluated the instrument and their comments were incorporated in the instrument.

The participants in each university were randomly divided into control and experimental groups. The students in the control group were exposed to the HT output whereas the students in the experimental group read the text generated by MT without any pre- or post-editing. Since the participants came from three different universities and the studies were conducted in their own universities, there were six groups of participants. The researchers were fortunate to be able to conduct their experiment prior to Covid-19 pandemic. To ensure the quality of the test, strict and similar procedure was implemented in all three universities.

Data analysis

The qualitative data from the interview were analyzed according to Popovic's (2018) classification of errors to answer the first research question. That is, having carefully analyzed the transcriptions, each participant's comments were coded according to each of the error types and the frequencies and percentages of their answers were tabulated on a scale ranging from strongly satisfied to strongly dissatisfied. The HT and MT reading comprehension test scores were used to answer the second research question. Having confirmed the normality of distribution, an independent-samples t-test was used to compare the observed mean scores of the two tests.

Results

Interview results

The results of the qualitative analysis of the data collected via the interview from 30 participants are presented below. The results are mainly pertinent to the first question of the study. The following table depicts the emerged pattern of the participants' answers to aspects of MT.

Table 1. Extracted patterns from MT in the interview

Extracted patterns	SS		S		NI		D		SD	
	F	P	F	P	F	P	F	P	F	P
Omission	3	10	18	60	5	16.6	4	13.4	0	0
Syntax	1	3.3	14	46.6	6	20	7	23.3	2	6.6
Mistranslation	3	10	10	33.3	2	6.6	14	46.6	1	3.3
Addition	3	10	16	53.3	5	16.6	4	13.4	2	6.6
Register	0	0	4	13.4	3	10	15	50	8	26.6
Idioms	0	0	1	3.3	1	3.3	6	20	22	73.3
Punctuation	3	10	14	46.6	3	10	7	23.3	3	10
Acceptability of MT	4	13.4	21	70	2	6.6	3	10	0	0

F: Frequency, P: Percentage, SS: Strongly Satisfied, S: Satisfied, NI: Mo Idea, D: Dissatisfied, SD Strongly Dissatisfied

According to the results reported in Table 1, the students' levels of satisfaction with omission (60%), addition (53.3%), syntax (46.6) and punctuation (46.6%) in MT implied that they believed although there were considerable instances of omission, addition, and/or inappropriate usage of language in terms syntax and punctuation in MT output, it was desirably acceptable. However, it can be argued that that most of the students (73.3%) did not welcome MT output in terms of the translated idiomatic expression. In the same line, it can be argued that existing errors of register and syntax in output were the secondary and tertiary level sources of dissatisfaction with the MT output. Despite these deficiencies, however, 83.4% of the students were generally satisfied with the output.

Reading comprehension test results

This section covers the results of the quantitative analysis of the participants' reading comprehension scores collected after administering the MT and HT tests. The results are mainly pertinent to the second research question of the study. The results are shown in the table below.

Table 2. The results of Descriptive Analysis, Test of Normality and Independent Samples T-Test for the MT and HT Reading Comprehension Scores

Groups	N	M	SD	Kolmogorov-Smirnov test			t	df	p
				Z	df	p			
MT	72	14.86	2.11	.89	152	.40	−.61	150	.52
HT	80	15.59	2.14	.67	152	.74			

As it was reported in Table 2, the normality of the distributions of the MT reading comprehension scores ($Z = .89$, $p = .40 < .05$) and HT reading comprehension scores ($Z = .67$, $p = −.74 < .05$) is confirmed considering the fact that the observed p levels are above the critical value (.05). Therefore, a parametric statistical test, independent samples t-test was used to compare the mean scores of the participants' HT and MT reading comprehension

test scores. The results ($t = -0.613$, $df = 151$, $p = 0.52 > 0.05$) indicated that there was no significant difference between the reading comprehension performances of the two groups. That is, the participants' comprehension of the HT output was comparable to that of the HT output. Accordingly, it can be argued that MT was desirably as successful as HT in terms of generating comprehensible texts.

Discussion

It is undeniable that there has been an extensive interest in improving EFL learners writing via MTALL. However, its effect on their reading comprehension ability has been scarcely researched. To fill the gap in the existing body of empirical knowledge, this study aimed at comparing the participants' reading comprehension of the MT versus HT texts. To accomplish the goal of the study, two parallel forms of a reading comprehension test, one including HT texts and the other including MT texts, with the same reading comprehension questions were distributed among the two groups participants with identical level of proficiency. In addition, the errors in MT output were also identified according to a recent typology of errors to further investigate the participants' perceptions of the extent to which each error type might lead to their (dis)satisfaction with the output in terms of their comprehension.

The analysis of the qualitative data conducted for the first research question depicted in spite of the fact that the participants were generally satisfied with the quality of the MT output, their levels of satisfaction considerably varied with regard to the different aspects, e.g. syntax, register and idioms, as measured in this study. Accordingly, it can be argued that the participants were satisfied with the micro-level errors, such as omission, addition (unrelated) words, scrambled word order (syntactic errors), and punctuation whereas they found macro-level errors such as mistranslation of idiomatic expressions, register errors and incorrect equivalent words in the target language, dissatisfying.

However, this has to be interpreted cautiously. Text difficulty is long considered to have a role in MT output quality (Pfafflin, 1965). Accordingly, it has to be born in mind that satisfaction with MT output quality in this study was based on the two texts featuring relatively identical readability level, as a measure of difficulty. Therefore, it may be concluded that the observed satisfaction with MT output quality is dependent on the audience took part in the interview and is not necessarily generalizable to all academic communities. Hence, the current MT systems have to be upgraded and enriched so that their output includes fewer micro-level and macro-level errors.

In addition, it is worth to note the fact that the participants judging the MT output quality all knew that the texts were translated by machine. In accordance with expectancy framework, diagnostic criterion categorization is influenced with such a presupposition. According to Mumma (1993), participants' evaluation may not be pure due to the information they received prior to the interview or when evaluating the output during the interview. In addition, according to the expectancy framework, their judgments may be polluted by theory-based expectations (Mumma, 1993). Consequently, the participants' judgments in this study may be to some extent infested with their expectancy ingrained in the information the researcher gave them about the source of the input before and during the interview. All in all, it is not irrational to conclude that the relatively considerable dissatisfaction of the participants with the micro-level errors in the MT output in this study, which was based on the inaccurate lexical and syntactic forms in MT text, was finally tranquilized by their awareness of the source of the text they were evaluating.

A further justification of the observed satisfaction with the MT output among the participants may be way errors are interpreted and perceived by the interviewees. The participants were initially dissatisfied mostly with lexical errors in the MT output. This aspect of dissatisfaction has been recurrently reported by several researchers (Bentivogli *et al.*, 2016; Castilho *et al.*, 2017) emphasizing MT output deficiency and its impaired language as compared with HT. However, those experts evaluating MT output quality via automatic metrics (Bahdanau *et al.*, 2014; Jean *et al.*, 2015; Bojar *et al.*, 2016; Koehn & Knowles, 2017) believe MT output has an acceptable quality. In the same line, according to the results of this study, the participants' dissatisfaction with the MT output may be due to machine's inability to recognize the proper meaning of the word according to its context and selects an irrelevant equivalent, defined as the one that is hardly ever or never selected by a human translator. Additionally, MT may be incapable of finding an equivalent jargon or register when encountering a technical term. However, MT may contain a comprehensible synonym, instead which helps reader decipher the meaning or the writer's intention. All in all, putting all the aforementioned justifications together, the MT output quality was assessed to be desirably satisfactory and it is in agreement with the findings of Castilho and Guerberof Arenas (2018) who found MT output in Chinese, English, and Spanish to be negligibly poorer than human translation output so that the readers were pleased with the output.

The results of the quantitative analysis were also in line with the findings discussed above. The independent *t*-test results indicated the observed mean scores of the two groups were similar. It is partly justified by the underlying in-built cognitive comprehension processes employed by the participants while reading the MT and HT outputs (Castilho & Guerberof Arenas, 2018). It may be argued that both groups were identical since both groups might have stimulated comparable schemata and areas of background knowledge to fulfill top-down cognitive processes they needed to accomplish the task. In addition, in accordance with Karnal and Pereira (2015) they may have employed the same strategies to infer MT output as the ones used to infer HT output. It is noteworthy that the lower observed MT reading comprehension test mean score may indicate the difficulties imposed by erroneous sentences, idiomatic expressions, and stylistic distractions, as discussed above (Roturier, 2006; Scarton & Specia, 2016).

The results of the quantitative analysis are also confirmed by the findings of the previous research on measuring usability of MT output via reading comprehension, such as Scarton and Specia (2016), Tomita *et al.* (1993). As indicated in previous research, reading comprehension tests is a valid means of evaluating MT output, especially, considering output informativeness, comprehensiveness and fluency measured by the number of correct answers provided for the comprehension questions (Fuji, 1999). The findings also support Fuji *et al.* (2001) and Jones *et al.*'s (2005b) in that judging usefulness of an MT system based on mere automatic metrics, unlike actual comprehension tests or human evaluation, is awry since they cannot measure informativeness, comprehensiveness, or fluency appropriately (Roturier, 2006).

Conclusion and implications

The findings of this study confirm the suitability or usefulness of MT system has been measured based on informativeness, comprehensiveness, and fluency of its output via a reading comprehension test. Although at the first glance the finding of this study seemed to be inconsistent with that some previous research, for example Mitkov (2004), the way

MT output was investigated in the present study was unlike the previous studies devising artificial-intelligence-oriented approaches (e.g. Bahdanau *et al.*, 2014; Bojar *et al.*, 2016; Jean *et al.*, 2015; Koehn & Knowles, 2017); instead of using mathematical and automated methods to evaluate MT output, the researcher focused on MT output quality defined in terms of its usefulness.

The findings discussed so far open a new vista to evaluation of MT output from a cognitive perspective relying on readability and usefulness. The results are broadening our view to MT output evaluation via noting the significance of source text type, reader expectancy, and comprehensibility of MT output as well as its informativeness as the significant domains to be investigated. In short, since MT output is both lexically and syntactically different from human generated text norms, these aspects has to be included in cognitively oriented evaluation of MT output usefulness. However, some variables highlighted in previous studies, such as readers' levels of literacy (Berka *et al.*, 2011; Jones *et al.*, 2005a), was neglected in this study. Furthermore, it is not unlikely that a number of unstudied variables such as question types and texts features (Castilho *et al.*, 2014), having been neglected in this study, may have influenced the results.

Considering the contribution of text difficulty to MT output quality, further research has to be done to explore MT output qualities across various texts differ in terms of their readability, styles and academic fields as well. Further investigation is required to pinpoint English-Persian MT output quality and the interchange between MT output characteristics and Persian readers' comprehension of the product. Moreover, considering unique generic features of different text types such as different sections of research papers (Thompson & Hunston, 2019) further research is needed to evaluate MT output including this variable. Additionally, considering the role of the reader's expectancy, it is required to replicate this research taking the role of readers' expectancy into account. Also, further empirical studies are looked-for to discover the possible effects of using MT on improving L2 learners' reading comprehension ability in English for academic purpose programs.

With regard to the pedagogical implications, the findings may valuable for higher education L2 teachers and learners. Provided that HT and MT output are similarly comprehensible, as indicated in this study, higher education students are expected to gradually be more interested in using MT systems rather that HT services to meet their needs in terms of covering their assignments requiring reading technical English texts. In addition, being seen as one of the limitations of this study, it has to be noted that inclusion of texts featuring topics other than those of the participants' fields of studies can diminish the intervening effect of their background knowledge on HT or MT output comprehension. Generally, the findings of this study eradicate the long-held perception that MT systems simply substitute source text words with target ones language and are unable to produce comprehensible texts. Confirming the limitations and deficiencies of MT systems, the present research depicted in spite of the proven limitations of MT systems, they are, to a great extent, useful in terms of rendering the source texts into comprehensible ones in target languages. Accordingly, L2 learners are recommended to enjoy the assistances MT offers as they struggle with comprehending English texts.

References

Abdelaal, N., & Alazzawie, A. (2020). Machine Translation: The case of Arabic-English translation of news texts. *Theory and Practice in Language Studies, 10*(4), 408–418. http://doi.org/10.17507/tpls.1004.09

Ali, K., & Alireza, F. (2014). The effect of computer-assisted translation on L2 leaners' mastery of writing. *International Journal of Research Studies in Language Learning, 3*(5), 29–44. http://doi.org/10.5861/ijrsll.2013.396

Amaral, L., & Meurers, D. (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL, 23*(6), 4–24. http://doi.org/10.1017/S0958344010000261

Amerian, M., & Marefat, F. (2019). A triangulated study of professional english needs of university graduates in business and economics in today's Iranian business sectors. *Applied Research on English Language, 8*(2), 227–260.

Amores, M. (1997). A new perspective on peer-editing. *Foreign Language Annals, 30*(4), 513–522. http://doi.org/10.1111/j.1944-9720.1997.tb00858.x

Avand, A. Q. (2009). Using translation and reading comprehension of ESP learners. *The Asian ESP Journal, 5*(1), 44–60.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. Retrieved online from *arXiv preprint arXiv:1409.0473* on February 2, 2020.

Bahri, H., & Mahadi, T. (2016). Google Translate as a supplementary tool for learning Malay: A case study at Universiti Sains Malaysia. *Advances in Language and Literary Studies, 7*(3), 161–167. http://doi.org/10.5755/j01.sal.0.21.1469

Barr, D. (2013). Embedding technology in translation teaching: Evaluative considerations for courseware integration. *Computer Assisted Language Learning, 26*(4), 295–310. http://doi.org/10.1080/09588221.2012.658406

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: a case study. *Retrieved online from arXiv preprint arXiv:1608.04631* on February 1, 2020.

Berka, J., Černý, M., & Bojar, O. (2011). Quiz-based evaluation of machine translation. *The Prague Bulletin of Mathematical Linguistics, 95*(5), 77–86.

Bernardini, S. (2016). Discovery learning in the language for translation classroom. *Cadernos de Traducao, 36*(3), 14–35. http://doi.org/10.5007/21757968.2016v36nesp1p14

Briggs, N. (2018). Neural machine translation tools in the language learning classroom: Students' use perceptions, and analyses. *The Jalt Call Journal, 14*(1), 3–24.

Bojar, O., Chatterjee, R., Federmann, C., Graham, Y., Haddow, B., Huck, M., & Negri, M. (2016). Findings of the 2016 conference on machine translation. In *Proceedings of the First Conference on Machine Translation: Volume 2, Shared Task Papers* (pp. 131–198).

Boniadi, A., Ghojazadeh, M., & Rahmatvand, N. (2013). Problems of an English for Specific Purpose course for medical students in Iran. *Khazar Journal of Humanities and Social Sciences, 3*(2), 48–88.

Boshrabadi, A. M. (2014). Pedagogical utility of translation in teaching reading comprehension to Iranian EFL learners. *International Journal of Language Learning and Applied Linguistics World, 5*(2), 381–395.

Case, M. (2015). Machine translation and the disruption of foreign language learning activities. *E-Learning Papers, 45*(8), 4–16. http://doi.org/10.1017/S0958344009000172

Castilho, S., & Guerberof Arenas, A. (2018). Reading comprehension of machine translation output: what makes for a better read? In J. Pérez-Ortiz, F. Sánchez-Martínez, M. Esplà-Gomis, M. Popović, C. Rico, A. Martins, J. Van den Bogaert, & M. L. Forcada (Eds.), *Proceedings of the 21st Annual Conference of the European Association for Machine Translation* (pp. 79–88). Alacant, Spain: University of Alacant.

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, P., & Gialama, M. (2017). *A comparative quality evaluation of PBSMT and NMT using professional translators*. MT Summit: Nagoya, Japan.

Castilho, S., O'Brien, S., Alves, F., & O'Brien, M. (2014). Does post-editing increase usability? A study with Brazilian Portuguese as target language. In *Proceedings of the Seventeenth Annual Conference of the European Association for Machine Translation* (pp. 16–18). Dubrovnik: University of Zagreb, Croatia.

Chang, C. & Yamada, M. (2020). Translation tasks for learning collocations: Effects of machine translation plus post-editing and sight translation. *English Teaching and Learning, 3*(5), 87–101. http://doi.org/ 10.1007/s42321-020-00059-x

Chen, I. J., & Yen, J. C. (2013). Hypertext annotation: Effects of presentation formats and learner proficiency on reading comprehension and vocabulary learning in foreign languages. *Computers & Education, 63*, 416–423.

Chang, C.-K., & Ching-Kun Hsu, C.-K. (2011). A mobile-assisted synchronously collaborative translation–annotation system for English as a foreign language (EFL) reading comprehension. *Computer Assisted Language Learning, 24* (2), 155–180, http://doi.org/ 10.1080/09588221.2010.536952

Chung, S. (2020). The Effect of L2 proficiency on post-editing machine translated texts. *The Journal of Asia TEFL, 17*(1), 182–193. http://doi.org/ 10.18823/asiatefl.2020.17.1.11.182

Correa, M. (2014). Leaving the "peer" out of peer-editing: Online translators as a pedagogical tool in the Spanish as a second language classroom. *Latin American Journal of Content and Language Integrated Learning, 7*(1), 1–20. http://doi.org/10.5294/laclil.2014.7.1.1

Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first- and second-language learners. *Reading research quarterly, 38*(1), 78–103.

Fuji, M. (1999). Evaluation experiment for reading comprehension of machine translation outputs. In *Proceedings of MT Summit VII* (pp. 285–289). Tokyo: AAMT.

Fuji, M., Hatanaka, N., Ito, E., Kamei, S., Kumai, H., Sukehiro, T., & Isahara, H. (2001). Evaluation method for determining groups of users who find MT useful. In *MT Summit VIII: Machine Translation in the Information Age* (pp. 103–108). Santiago de Compostela, Spain.

Garcia, I., & Pena, M. (2011). Machine translation-assisted language learning: Writing for beginners. *Computer Assisted Language Learning, 24*(5), 471–487. http://doi.org/10.1080/ 09588221.2011.582687

Godwin-Jones, R. (2015). Contributing, creating, and curating: Digital literacies for language learners. *Language Learning & Technology, 19*(3), 8–20. http://doi.org/10125/44427

Golkar, M., & Yamini, M. (2007). Vocabulary, proficiency and reading comprehension. *The Reading Matrix, 7 (3)*, 88–112.

Goodwin, A. P., & Jiménez, R. (2016). TRANSLATE: New strategic approaches for English learners. *The Reading Teacher, 69*(6), 621–625.

Groves, M., & Mundt, K. (2015). Friend or foe? Google Translate in language for academic purposes. *English for Specific Purposes*, *37*(8), 112–121. http://doi.org/10.1016/j.esp.2014.09.001

Holt, Overgaard, & Engel, L. (2020). Health literacy, digital literacy and eHealth literacy in Danish nursing students at entry and graduate level: A cross sectional study. *BMC Nurse*, *19*(22), 33–50. https://doi.org/10.1186/s12912-020-00418-w

Hsu, H.-C. (2019). Wiki-mediated collaboration and its association with L2 writing development: An exploratory study. *Computer Assisted Language Learning*, *4*(2), 23–41 http://doi.org/10.1080/09588221.2018.1542407

Iranmehr, A., Atai, M. R., & Babaii, E. (2018). Evaluation of EAP Programs in Iran: Document Analysis and Expert Perspectives. *Applied Research on English Language*, *7*(2), 171–194.

Jean, S., Firat, O., Cho, K., Memisevic, R., & Bengio, Y. (2015). Montreal neural machine translation systems for WMT'15. In *Proceedings of the Tenth Workshop on Statistical Machine Translation* (pp. 134–140). Stroudsburg, PA: Association for Computational Linguistics.

Jin, L. (2013). Foreign language learners' use and perception of online dictionaries: A survey study. *MERLOT Journal of Online Learning and Teaching*, *9*(4), 513–533. http://doi.org/10.1076/call.14.5.367.5775

Jones, D., Gibson, E., Shen, W., Granoien, N., Herzog, M., Reynolds, D., & Weinstein, C. (2005a). Measuring human readability of machine generated text: three case studies in speech recognition and machine translation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.* (Vol. 5) (pp. v-1009). Piscataway, NJ: IEEE.

Jones, D., Shen, W., Granoien, N., Herzog, M., & Weinstein, C. (2005b). Measuring human readability of machine generated text: Studies in speech recognition and machine translation. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA: IEEE.

Josefsson, E. (2011). Contemporary approaches to translation in the classroom: A study of students' attitudes and strategies. *Retrieved from* http://du.diva-portal.org/smash/get/diva2:519125/FULLTEXT01.pdf.

Karnal, A. R., & Pereira, V. V. (2015). Reading strategies in a L2: A study on machine translation. *Reading Matrix: An International Online Journal*, *15*(2), 69–80.

Kliffer, M. (2005). An experiment in MT post-editing by a class of intermediate/advanced French majors. *In Proceedings EAMT 10th Annual Conference* (pp. 160–165). Pázmány Péter Catholic University & MorphoLogic, Budapest, Hungary.

Kliffer, M. (2008). Post-editing machine translation as an FSL exercise. *Porta Linguarum*, *9*(5), 53–67.

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. *Retrieved online from arXiv preprint arXiv:1706.03872* on February 3, 2020.

Larsen-Freeman, D., & Anderson, M. (2013). *Techniques and principles in language teaching* (3rd ed.). Oxford: Oxford University Press.

Lee, S. (2019). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, *8*(6), 102–124. http://doi.org/10.1080/09588221.2018.1553186

Liao, P. (2006). EFL learners' beliefs about and strategy use of translation in English learning. *RELC Journal*, *37*(2), 191–215.

Mahmoud, A. (2006). Translation and foreign language reading comprehension: A neglected didactic procedure. *English Teaching Forum 44* (4), 28–39.

Mirzaeian, V. (2020). Machine Translation Output Assessment and its Impact on Reading Comprehension. *Technology of Education Journal (TEJ), 14*(2), 697–710. https://doi.org/10.22061/JTE.2019.5473.2227

Mitkov, R. (Ed.). (2004). *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press.

Mohebbi, H., Nayernia, A., Nemati, M., & Mohebbi, B. (2017). Readability of ESP textbooks in Iran: A neglected issue or a taken-for-granted one? *Journal of Teaching English for Specific and Academic Purposes, 4*(3), 641–654.

Mostafaei Alaei, M., & Ershadi, A. R. (2017). ESP program in Iran: A stakeholder-based evaluation of the program's goal, methodology, and textbook. *Issues in Language Teaching, 5*(2), 306–279.

Mumma, G. H. (1993). Categorization and rule induction in clinical diagnosis and assessment. *The Psychology of learning and motivation, 29*, 283–326.

Murphy, D. (2020). Supporting pre-service English teachers' academic reading and writing with online machine translation. *STEM, 21*(2), 25–32. http://doi.org/10.16875/stem.2020.21.2.123

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nino, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning, 21*(1), 29–49. http://doi.org/10.1080/ 09588220701865482

Nino, A. (2009). Machine translation in foreign language learning: Language learners' and tutors' perceptions of its advantages and disadvantages. *ReCALL, 21*(2), 241–258. http://doi.org/10.1017/S0958344009000172

O'Brien, S., Simard, M., & Goulet, M. J. (2018). Machine translation and self-post-editing for academic writing support: Quality explorations. In *Translation Quality Assessment* (pp. 237–262). Cham, Switzerland: Springer.

Papineni, K., Roukos, S., Ward, T., & Zhu, W. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of The 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311–318), Philadelphia, PA: Association for Computational Linguistics.

Pfafflin, S. M. (1965). Evaluation of machine translations by reading comprehension tests and subjective judgments. *RCT, 1*(32), 28–40.

Popovic, M. (2018). Error classification and analysis for machine translation. *Quality Assessment, 8*(6), 52–63. http://doi.org/10.1007/978-3-319-91241-7_7.

Ross, N. J. (2000). Interference and intervention: Using translation in the EFL classroom. *Modern English Teacher, 9*(3), 61–66.

Roturier, J. (2006). *An investigation into the impact of controlled English rules on the comprehensibility, usefulness and acceptability of machine-translated technical documentation for French and German users.* [Unpublished doctoral dissertation]. Dublin City University.

Rushwan, I. M. H. (2017). The role of translation in developing ESP learners' reading comprehension skills-A case study of medical students at Najran University-KSA. *International Journal of Applied Linguistics and English Literature, 6*(3), 243–253.

Scarton, C., & Specia, L. (2016). A reading comprehension corpus for machine translation evaluation. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (pp. 3652–3658). Philadelphia, PA: Association for Computational Linguistics.

Shei, C. (2002a). *Combining translation into the second language and second language learning: An integrated computational approach.* [Unpublished doctoral dissertation]. University of Edinburgh, Edinburgh.

Shei, C. (2002b). *Teaching MT through pre-editing: Three case studies*. In Proceedings of 6th EAMT Workshop Teaching Machine Translation (pp. 89–98). Philadelphia, PA: Association for Computational Linguistics.

Tavakoli, M., & Tavakol, M. (2018). Problematizing EAP education in Iran: A critical ethnographic study of educational, political, and sociocultural roots. *Journal of English for Academic Purposes*, *3*(1), 28–43.

Thompson, P., & Hunston, S. (2019). *Interdisciplinary research discourse: Corpus investigations into environment journals.* London: Routledge.

Tomita, M., Shirai, M., Tsutsumi, J., Matsumura, M., & Yoshikawa, Y. (1993). Evaluation of MT Systems by TOEFL. In *Proceedings of the Theoretical and Methodological Implications of Machine Translation (TMI-93)*. Allschwil, Switzerland: European Association of Machine Translation.