Machine vs human translation: a new reality or a threat to professional Arabic–English translators

Muneera Muftah

Department of English, College of Languages and Translation, Najran University, Najran, Saudi Arabia and
Department of English, Faculty of Arts, Thamar University, Dhamar, Yemen

Abstract

Purpose – How closely does the translation match the meaning of the reference has always been a key aspect of any machine translation (MT) service. Therefore, the primary goal of this research is to assess and compare translation adequacy in machine vs human translation (HT) from Arabic to English. The study looks into whether the MT product is adequate and more reliable than the HT. It also seeks to determine whether MT poses a real threat to professional Arabic–English translators.

Design/methodology/approach – Six different texts were chosen and translated from Arabic to English by two nonexpert undergraduate translation students as well as MT services, including Google Translate and Babylon Translation. The first system is free, whereas the second system is a fee-based service. Additionally, two expert translators developed a reference translation (RT) against which human and machine translations were compared and analyzed. Furthermore, the Sketch Engine software was utilized to examine the translations to determine if there is a significant difference between human and machine translations against the RT.

Findings – The findings indicated that when compared to the RT, there was no statistically significant difference between human and machine translations and that MTs were adequate translations. The human–machine relationship is mutually beneficial. However, MT will never be able to completely automated; rather, it will benefit rather than endanger humans. A translator who knows how to use MT will have an opportunity over those who are unfamiliar with the most up-to-date translation technology. As MTs improve, human translators may no longer be accurate translators, but rather editors and editing materials previously translated by machines.

Practical implications – The findings of this study provide valuable and practical implications for research in the field of MTs and for anyone interested in conducting MT research.

Originality/value – In general, this study is significant as it is a serious attempt at getting a better understanding of the efficiency of MT vs HT in translating the Arabic–English texts, and it will be beneficial for translators, students, educators as well as scholars in the field of translation.

Keywords Translation adequacy, Machine translation, Human translation, Comparative study, Arabic–English language translation

Paper type Research paper

1. Introduction

The desire to make a given facet of life easier is always the driving force behind any form of improvement. As a result, it is not astonishing that efforts have been made to abolish the language barrier, which has been a source of frustration for people since the dawn of time.

Linguists and computer scientists throughout the world are working to develop cheap software that can act as universal translators, translating between many language pairs. This concept, which was once merely a dream, is one that Google, among others, wants to make a reality.

In machine translation (MT) research, there was a shift in emphasis from strictly theoretical study to practical applications, which persisted throughout the 1990s. The use of MT by large corporations has increased rapidly, particularly in the field of software localization (i.e. adapting computer programs and games to target language recipients), sales of MT software for personal computers have increased significantly and MT has been offered by an increasing number of online services, making it easily accessible to anyone with Internet access (Puchała-Ladzińska, 2016).

Once parallel data for the languages are available, these techniques allow new languages to be supported without any need for handcrafted linguistic rules (Doherty, 2016). The disadvantage is that these software systems are constrained by their lack of linguistic knowledge and their reliance on their own datasets. As a result, any new terms or formulations will be hard to accurately translate if they are not included in the systems' data.

Because MT systems are basically constructed from human translations (HTs), they help bridge the gap between human and MT. Today's systems often include millions of sentences translated by humans from which these systems gain probability patterns, while customized and freely accessible online systems can comprise even more data gathered from a huge number of translators over many years (Munkova *et al.*, 2021). These systems are constantly improving in terms of consistency and effectiveness and more high-quality translation becomes available, posing a risk to human translators. MT systems, on the other hand, have recently gained acceptance among professional translators and academics (Bowker, 2019; Vieira and Alonso, 2020; Way, 2018). Despite this, many translators are still adjusting to the changes that translation technologies have introduced to the field of translation and the translation process.

MT systems are fundamentally developed from HTs (Doherty, 2016). Furthermore, they are most effective in languages that are closely related and belong to the same family (which makes them at least a little similar) (Munkova *et al.*, 2021). This is not our case because we have focused on Arabic as the source language and English as the target language. The language families of Arabic and English are distinct. Thus, Arabic is a Semitic language that has many distinctive linguistic characteristics including writing from right to left, the dual number of nouns which is not found in English, the two genders, feminine and masculine, in addition to the root, which is the most salient feature of Semitic languages, whereas English is a Germanic language (Mendel, 2016).

The main difference between Arabic and English is in their grammatical properties. English is an analytical language (with some synthetic elements), whereas Arabic is a synthetic language (Fehri, 2012). And this is precisely what distinguishes them and causes the most difficulties in translation both machine and HTs.

Translation is such a sensitive and sophisticated task in language studies that raises some serious concerns. It also involves the transformation of a variety of distinguishing features from one language to another. As Arabic and English are of disparate origins, any translation from one script to the other can be difficult, especially in terms of vocabulary, grammar, sound, style and usage (Akan *et al.*, 2019). However, it appears that translating a text from Arabic to English is a far too challenging task for nonnative Arab EFL learners as it requires extensive bilingual knowledge (Akan *et al.*, 2019; Shahata, 2020).

In the context of Arabic translation, research work undertaken in Arabic into English MT is extremely essential. A few, if any, studies comparing Arabic–English MT versus HT have been done. In addition, due to the large number of Arabic texts to be translated in recent years, research on the adequacy, fidelity and acceptability of Arabic–English MT vs HT is required. Therefore, this study is an endeavor to

(1) compare the translation adequacy of human vs machine translation in an Arabic–English context in a comparative design.

(2) examine whether MT product is adequate and more reliable as compared to HT.

(3) find out if MT threatens professional Arabic–English translators.

Generally, this research is noteworthy because it represents a real endeavor to gain a better knowledge of the efficiency of MT vs HT in translating Arabic–English texts. It also discusses the future of MT and attempts to answer the question of whether human translators will be replaced by artificial intelligence in the near future.

2. Literature review

With the increased demand for translation, MT technology has become the main interest in the Arab world. The usage of translation technology in general, and MT, in particular, has become a requirement as the need for translation grows. Various MT systems have been developed as a result of research and are now in use in many countries. Sakhr, ATA software, Cimos and SYSTRAN are some of these systems that support the Arabic language. There are other web-based MT systems with Arabic as a source or a target language, such as Babylon, Bing Translator and Google Translate.

MT systems are currently widely utilized around the world as the demand for translation has expanded dramatically and as a result of the vast volume of content that needs to be translated in every discipline (Almutawa and Izwaini, 2015). More translation systems will be required to keep up with the global information technology revolution, and because translators will not be able to keep up with the volume of material, there is a place for MT, which can save time and energy, at least when only the gist of a text is required rather than a complete and accurate translation or when translating websites and online information. When only a quick postediting is necessary, MT can be used to make rough translations with translators postediting the output. Human translators can save time that would otherwise be spent translating simple or repetitive material in this way (Almutawa and Izwaini, 2015).

Dia et al. (2022) have also claimed that while MT has a significant impact on translation, HT work will not be replaced by MT and will continue to exist in the future. As MT improves and translation practice evolves, Şahin and Gürses (2021) concluded that MT is still a long way from being an essential part of any literary translation practice for the English–Turkish language pair and that translators' interactions with MT and negative attitudes toward it may change in a positive direction. However, according to Maghsoudi and Mirzaeian (2020), MT has progressed to the point where it can compete with the HT. This is consistent with Vasheghani Farahani (2020) study, which has found no statistically significant difference between human and machine translations when compared to the reference translation (RT) and has concluded that MTs were competent translations.

Professional translators have always been concerned with the production of an adequate and acceptable translation that delivers the content materials of the source language into the target text (Gerber, 2012), which takes too much of their time. As a result, MT, a software-assisted translation of communication from one language to another, has found its way into our lives. The quality of MT service has improved in recent years, owing to the significant growth in international communication (Li et al., 2014). To put it another way, modern MT services like Google Translate and Bing Translator have made significant progress in allowing users to read content in foreign languages (Almutawa and Izwaini, 2015).

When contrasted to MT, HT has certain distinct characteristics. To begin with, when compared to MTs, HT is often performed at a very slow speed. Additionally, HTs are edited by humans, either the translator or the editor; nevertheless, MTs require postediting, which is

done by humans. Furthermore, there is no interference of human beings in HT. It is often held that MT is assessed by comparing it to human professional translation (Papineni *et al.*, 2002). In this regard, Delpech (2014) claims that MT assessment serves two goals:

(1) First, it examines the impact of a system alteration on the quality of translations during the development of the MT system.

(2) Second, the evaluation allows us to compare the systems in question, which is normally done as part of a larger evaluation effort. Each of these goals has an equivalent evaluation technique.

The need to use MT stems from the fact that MTs is becoming increasingly popular among end users, and many rely on them for their translation needs (Li *et al.*, 2014). However, its ultimate production in the target language leaves something to be desired, as it contains flaws and inconsistencies. As a result, the concept of MT quality assessment has emerged.

Assessing MT is recognized as a critical field of research for determining the efficacy of present MTs as well as developing future MTs (Martindale and Carpuat, 2018). One type of translation quality assessment is comparing the adequacy of MT to HT.

Translation adequacy refers to the extent to which the output transmits the same meaning or information as the input. How much of the source language translation has been kept in the target language is an important feature of MT adequacy. Although varied judgments have always been made about translation adequacy, there is one common agreement among professionals that translation adequacy is directed and evaluated in comparison to the source text (Chesterman, 2016).

Despite these advancements, experts in the field of translation have generally stated that assessing the adequacy of MT services requires further research, stating that evaluating MT quality in terms of adequacy is considered an emerging topic of inquiry in academia that requires further exploration. To put it another way, evaluating MT quality has always been a fascinating and appealing topic of research but has gained less attention from academia (O'Brien, 2012). Furthermore, translation adequacy research requires greater examination of underresearched language pairs, such as Arabic–English.

In light of the aforementioned concerns and the fact that MT adequacy assessment is an emerging area of research that merits more investigation, this work was an attempt to examine translation adequacy in HT vs MT in an Arabic–English setting in a comparative design.

3. Methodology
The design of this research is comparative as it sets to compare the HT with MT in terms of adequacy and acceptability.

3.1 Reference translation
A RT that is believed to be faultless, adequate, competent and acceptable is required to determine translation adequacy. As a result, different Arabic (Modern Standard Arabic) text types were picked. The text genre was instructive, and it included written mode in a variety of subjects such as literary, legal, environmental preservation, economics, basic sciences and medical sciences, ensuring that they were dense with specialized terminologies and jargon. The number of different text genres analyzed was limited to six. They varied from the most practical to the most situation-specific. On the one hand, it is meant to select a type of text with more pragmatic information, concise and even short (where possible) sentences, and limited semantic scope. It was desired to have a highly pragmatic, stylistically and semantically rich, elaborate text. The translations were limited to circa 300 words in length. It was desired to

have a reasonable length that would provide us with a diverse range of linguistic information, including a sufficient number of terminologies and jargon.

Then, two certified professional Arabic–English translators were requested to translate the texts from Arabic into English as a RT. The translations were subsequently assessed and evaluated by two professional translators acting as the RT raters and using a holistic model established by Waddington (2001).

Waddington's model is split into four different scoring rubrics. Model C was used in this study since it was simpler and more in accordance with the research's goal. Waddington (2001) proposed a holistic paradigm of translation assessment. According to the instructions provided to the translation evaluator, any translation is assessed at five different levels and scored between 0 and 10 (see Table A1). The range of 0–10 allows the reviewer to offer higher marks to translators who produce better outcomes and lower scores to translators who produce poor results. To examine the scores of the two professional translators, the Pearson correlation test was run to ensure that the scores of the reference translators were correlated.

3.2 Machine translations vs human translations

Two translation software were selected to perform MT: Google translate software and Babylon translate software. The first program was selected since it is the predominant translation software used by translators and is freely accessible and publicly available. Moreover, Babylon translate software was chosen because it is a popular and highly recommended tool among students, businesses and linguists. It has powerful translation engines that enable it to provide comprehensive, quick and full-text translations at affordable prices.

Following the MTs, the texts were translated by two nonexpert undergraduate translation students as the HT. The participants of the study were fourth-year students of Translation Studies. They were selected through accidental sampling based on their achievement in the last two years. Students were supposed to translate the Arabic texts into English and were allowed to use dictionaries where required.

To assess their adequacy, the translations were reviewed, analyzed and compared independently to the RT by three bilinguals, professional raters who were university instructors. The set of criteria of the model of translation adequacy established by Specia *et al.* (2011) was utilized to evaluate MT.

This model was chosen because it was simple to implement and follow. The following criteria make up their translation adequacy model:

(1) the frequency of tokens in the source and target, and vice versa,

(2) the absolute difference between the number of tokens in source and target normalized by source length,

(3) the ratio of percentages of numbers, content-/functional words in the source and target the absolute difference between the number of superficial structures in the source and target: brackets, numbers and punctuation symbols,

(4) the difference in the number of PP/NP/VP/AdjP/AdvP/ConjP phrases between the source and target and

(5) difference between the number of entities, such as a person, location and organization in source and target sentences.

In addition to the above-mentioned criteria, machine-translation evaluators employed the following set of criteria (Delpech, 2014, p. 43):

(1) the number of word n-grams shared by the evaluated translation and the RT, for n between 1 and 4;

(2) the (word number) size differences between the evaluated translation and the RT.

In this study, the texts were assessed using both sets of criteria in order to provide a more accurate and reliable assessment of MT adequacy.

3.3 Sketch Engine software

Sketch Engine is a piece of window-based corpus software that is mostly used in Corpus Linguistics. Lexical Computing Ltd. (https://www.sketchengine.co.uk/) created this application. It provides researchers with a variety of options, including precise word extraction, concordance lines, context keywords and collocation patterns (McGillivray and Kilgarrif, 2013). This software was applied to extract and evaluate specific information from the texts (the RT, MT and HT).

Moreover, a chi-square test of independence was conducted to figure out whether the differences between the three translations (Google translation, Babylon translation and human translation) and the Reference translation are of statistical significance.

4. Results

To measure translation quality in practice, the RT was used as the target text and was analyzed using the criteria indicated in the methodology. Similar processes were then used in both human and MTs. It is worth mentioning that even though the study's corpus was formed by several sub-corpora, they were all treated as one unified corpus during the translation assessment phase.

4.1 Reference translation

The scores of the two professional translators were compared to ensure that they were correlated. The significant two-tailed between the two raters was 0.015, as shown in Table 1. As a result, the correlation index was satisfactory.

The findings of assessing RT are displayed in Table 2. The RT served as a reference against which the machine and human translations were measured and evaluated.

The fundamental information about the Reference translation including the number of sentences (55), words (1,447), tokens (1,608) and tags (50) are represented in Table 2. To establish translation adequacy, it is necessary to examine and distinguish between content (736) and functional words (611). In simple terms, content words are words and expressions that refer to an item, quality, situation or action and have meaning (i.e. lexical meaning) when used alone (Richard and Schmidt, 2010). Functional words, on the other hand, are terms that have little significance on their own but illustrate grammatical links in and between sentences.

The difference in the number of superficial structures between the source and target texts is another criterion for evaluating translation adequacy. Overall, there are 25 brackets, 48 numbers and 148 punctuation marks in the RT.

		Rater1PosE	Rater2PostE
Rater1PosE	Pearson Correlation	1	0.564
	Sig. (2-tailed)		0.015
	N	10	10
Rater2PostE	Pearson Correlation	0.564	
	Sig. 2-tailed)	0.015	
	N	10	10
Note(s): *Correlation is significant at the 0.05 level (2-tailed)			

Table 1.
Correlations of the reference translation

Basic info.		Type of words		No. of Superficial Structures		No. of phrases		No. of entities		No. of n-grams	
No. of Sentences	55	Content	736	Brackets	25	VP	145	Persons	3	Two Words	745
No. of Words	1447	functional	611	Numbers	48	NP	417	Locations	8	Three Words	437
No. of Tokens	1608			Punctuation marks	148	AdjP	153	Organizations	5	Four Words & more	181
No. of Tags	50					AdvP	31				
						PP	224				
						ConjP	99				
						Article	173				
						Pronoun	35				
						Auxiliary	75				

Table 2.
Results of the reference
translation

The absolute difference between the number of phrases including the Noun Phrase (NP), Verb Phrase (VP), Adjectival Phrase (AdjP), Adverbial Phrase (AdvP), Conjunctional Phrase (ConjP) and Prepositional Phrase (PP) identified in the RT as well as the machine/human translation is the next measure against which a translation is assessed. The RT contains 145 verb phrases, 417 noun phrases, 153 adjectival phrases, 31 adverbial phrases, 224 prepositional phrases, 99 conjunctions, 173 articles, 35 pronouns and 75 auxiliaries, as evidenced by the data shown in Table 2. Moreover, in the RT, there are 3 occurrences of people, 8 occurrences of places and 5 occurrences of organizations.

N-grams are also crucial criteria for evaluating MT adequacy, as per Delpech (2014). N-grams are a sequence of n components (typically words) that occur immediately one after another in a corpus, where n is two or more (McEnery and Hardie, 2012). The number of n-grams in the RT is shown in Table 2. There were 745 two-word n-grams in the Reference translation, 437 three-word n-grams and only 181 four and more than four-word n-grams.

4.2 Machine translations vs human translations

The same processes were used on the Human translation and the Google and Babylon translation software, respectively. Table 3 demonstrated the frequency distribution of basic components including sentences, words, tokens and tags in the four translation methods.

Chi-square test findings revealed no statistically significant difference between the three translations (Google translation $p = 0.47$, Babylon translation $p = 0.61$ and human translation $p = 0.53$) and the reference translation. To put it another way, in terms of the frequency of sentences, words, tokens and tags, all three translations were equivalent to the Reference translation. However, as demonstrated by the p-value results, the significant difference of the Babylon translation was less compared to Google and human translations. Furthermore, as the data indicated, there was no statistically significant difference between Google and Babylon ($p = 0.53$), Google and human ($p = 0.73$), as well as Babylon and human ($p = 0.50$).

Table 4 reported the frequency and distribution of content words (i.e. verbs, adjectives, nouns and adverbs) as compared to the functional ones (i.e. prepositions, conjunctions,

articles, pronouns and auxiliaries) in the four translation methods. Generally, content words were used more than functional words. In terms of content/functional words, Reference translation received the greatest proportion of 54.6 and 45.4% respectively.

In line with the chi-square test, the frequency analysis of content words between the Google translation and the Reference translation was proved to be significantly different ($p < 0.001$). Furthermore, the results revealed that there was no statistically significant difference between Babylon translation ($p = 0.50$) and human translation ($p = 0.05$). In other words, the difference between Babylon translation and Reference translation was slighter than the difference between Google and human translations, in keeping with the estimated p-value. Results have revealed statistically significant differences for each Babylon ($p = 0.001$) and human translation ($p = 0.005$); however, no statistically significant difference has been found between both translations ($p = 0.38$).

The frequency of superficial structures including brackets, numbers and punctuation marks was shown in Table 5. As it is obvious, punctuation marks were the most used superficial structures in the four translations with 67.0% in Reference translation, 66.1% in human translation, 57.2% in Google and 55.6% in Babylon translations, while the least often used structure is brackets.

Similarly, the chi-square test revealed that there was no statistically significant difference between Google ($p = 0.27$), Babylon ($p = 0.21$) and human ($p = 0.62$) translations from the one hand and the RT from another. That is to say, all three translations were extremely close to the RT. Nonetheless, the p-value indicated that the human translation was closer to the RT than the other translations. Additionally, no statistically significant differences were found between Google and Babylon ($p = 0.65$), Google and human ($p = 0.15$), or Babylon and human ($p = 0.12$).

Table 6 reported the distribution of the different grammatical categories including the frequency of NP, VP, AdjP, AdvP, ConjP and PP phrases in all translations. As indicated by the results, the RT contained the highest number of phrases compared to machine and human translations.

Basic info	Reference translation	Google translation	Babylon translation	Human translation
No. of sentences	55/(1.7%)	43/(1.9%)	35/(1.5%)	49/(1.9%)
No. of words	1,447/(45.8%)	994/(46.2%)	1,079/(45.4%)	1,191/(45.8%)
No. of tokens	1,608/(50.9%)	1,075/(50.1%)	1,223/(51.3%)	1,309/(50.3%)
No. of tags	50/(1.6%)	39/(1.8%)	42/(1.8%)	52/(2.0%)

Table 3. Frequency distribution of the basic information in the four translation methods

Type of words	Reference translation	Google translation	Babylon translation	Human translation
Content words	736/(54.6%)	597/(62.8%)	542/(56.3%)	601/(59.3%)
Functional words	611/(45.4%)	353/(37.2%)	420/(43.7%)	412/(40.7%)

Table 4. Frequency distribution of type of words in the four translation methods

Superficial structures	Reference translation	Google translation	Babylon translation	Human translation
Brackets	25/(11.3%)	23/(15.1%)	23/(15.0%)	25/(12.1%)
Numbers	48/(21.7%)	42/(27.7%)	45/(29.4%)	45/(21.8%)
Punctuation marks	148/(67.0%)	87/(57.2%)	85/(55.6%)	136/(66.1%)

Table 5. Frequency distribution of superficial structures in the four translation methods

The chi-square test results revealed that there was no significant difference in the frequency distribution of the grammatical constructions between RT and Google translation ($p = 0.11$) or human translation ($p = 0.43$). However, there was a slightly significant difference between the RT and the Babylon translation ($p = 0.031$). Moreover, the p-value results showed that the difference between the human translation and the RT was slight. Likewise, there was no statistically significant difference in translation between Babylon and Google translation ($p = 0.28$).

Table 7 displayed facts about the frequency distribution of person, location, and organization entities. The location entity was the most frequent in all translations. The second most devoted entity was the organization, while the person entity was the least utilized in all translations.

The chi-square test revealed that the distributive frequency of a person, location and organization with Google ($p = 0.62$), Babylon ($p = 0.71$) and human translations ($p = 0.50$) had no statistically significant difference. In other words, statistically, all three translations were close and comparable with the RT. Evidence came from p-value results between translations which has also indicated that there was no statistically significant difference between Google and Babylon ($p = 0.67$), Google and human ($p = 0.60$) and Babylon and human translations ($p = 0.75$).

The distribution of N-grams in the four translations was shown in Table 8. As can be noticed, two-word n-grams were the most common in all four translations, followed by three-word n-grams and then the four-word n-grams, which were the least common.

Phrases structures	Reference translation	Google translation	Babylon translation	Human translation
VP	145/(10.7%)	83/(9.1%)	89/(8.6%)	86/(7.8%)
NP	417/(30.8%)	305/(33.3%)	363/(35.0%)	375/(34.2%)
AdjP	153/(11.3%)	101/(11.0%)	130/(12.5%)	142/(13.0%)
AdvP	31/(2.3%)	14/(1.5%)	19/(1.8%)	21/(1.9%)
PP	224/(16.6%)	150/(16.4%)	165/(15.9%)	189/(17.2%)
ConjP	99/(7.3%)	60/(6.6%)	58/(5.6%)	91/(8.3%)
Article	173/(12.8%)	135/(14.8%)	161/(15.5%)	122/(11.1%)
Pronoun	35/(2.6%)	15/(1.6%)	10/(1.0%)	20/(1.9%)
Auxiliary	75/(5.5%)	52/(5.7%)	42/(4.1%)	50/(4.6%)

Table 6.
Frequency distribution of phrases in the four translation methods

Entities	Reference translation	Google translation	Babylon translation	Human translation
Persons	3/(18.8%)	3/(25.0%)	5/(27.8%)	5/(27.8%)
Locations	8/(50.0%)	6/(50.0%)	8/(44.4%)	7/(38.9%)
Organizations	5/(31.2%)	3/(25.0%)	5/(27.8%)	6/(33.3%)

Table 7.
Frequency distribution of entities in the four translation methods

N-grams	Reference translation	Google translation	Babylon translation	Human translation
Two words	745/(54.7%)	592/(52.9%)	778/(55.1%)	940/(64.1%)
Three words	437/(32.1%)	431/(38.5%)	444/(31.5%)	370/(25.2%)
Four words and more	181/(13.2%)	96/(8.6%)	189/(13.4%)	156/(10.7%)

Table 8.
Frequency distribution of n-grams in the four translation methods

The chi-square analysis found a statistically significant difference between Google ($p < 0.001$), Babylon ($p < 0.001$) and human translations ($p < 0.001$). In other words, when compared to the RT, all three versions exhibited a statistically significant difference. That is to say, there was a statistically significant difference between Google and Babylon translation, Google and human translation, and Babylon and Human translation.

5. Discussion and conclusion

Online text translation services are becoming increasingly popular because of their quick performance and variety. Since they do not know all languages, the majority of individuals nowadays rely heavily on MT. MT adequacy assessment is a new area of research that deserves greater attention. For this purpose, the present study used a comparative design to compare translation adequacy in machine vs human translation in an Arabic–English scenario.

Primarily, Sketch Engine software was applied to extract and assess certain textual information in all translations and to address the research objectives that appeared early in the research. The comparison of the translation adequacy of human vs MT in the Arabic–English context has shown that there was no statistically significant difference between the three translations and the RT.

In terms of the first criterion, which is the frequency of content and functional words, Google translation had the most content words, followed by human and Babylon translation. Concerning the functional words, however, it was discovered that Babylon's translation had the highest number of functional words, followed by Human and Google translations.

Pertaining to other criteria such as the distribution of superficial structures, grammatical categories as well as the number of entities, no statistically significant differences were found between Google and Babylon, Google and human, or Babylon and human. The chi-square test also revealed that there was no statistically significant difference between all three translations from one hand and the RT from another.

The distribution of n-grams in the four translations revealed that two-word n-grams were the most prevalent, followed by three-word n-grams and finally four-word n-grams, which were the least common. Statistically, significant differences between Google and Babylon translation, Google and human translation, and Babylon and human have been found.

MT quality assessment has a long history (Hutchins, 2001). This study compared and contrasted the accuracy of MT vs HT. The findings revealed that there was no statistically significant difference(s) between machine and human translations in terms of adequacy. Machine translations, in other words, could provide translations that were very equivalent to the RT and were adequate translations (Maghsoudi and Mirzaeian, 2020; Vasheghani Farahani, 2020). In the case of Arabic–English language pairs, it can be argued that translation services like Google and Babylon can provide appropriate and adequate output/translation. The final output, on the other hand, requires postediting by a human editor to adjust for MT inaccuracies that can occur.

The findings of this study diverge from those of Li *et al.* (2014), who found that MT still has to be chosen to generate a satisfactory translation into the target language. On the other hand, the findings matched those of the Abusaaleek (2016) study, which found that Google Translate could provide a satisfactory and suitable translation.

The debate over machine vs human translation continues, with the question of whether MT will eventually replace HT in an era when MT is improving all the time. MT has significantly reduced the language barrier. After all, MT outperforms humans in at least two ways when it comes to translation: they can do it much faster and for a lot less money, and these two advantages are especially appealing nowadays when saving time and money are top goals for most businesses. As a result, some translators are concerned that too much advancement in the field of MT would threaten their careers.

MTs, on the other hand, have far too many flaws to be useful in many areas of life. Google translation, Babylon translation and comparable systems' output will only be useful for a restricted purpose: determining the overall meaning of the source text message. Nonetheless, human creativity and intellect are essential parts of translation, and no software has yet been able to replicate them.

MT will be utilized, and it currently is, but there will always be a need for a person to evaluate the quality of that translation if only to ensure that everything is correct (Puchała-Ladzińska, 2016). Machines can help speed up translation, but they cannot be the entire option, and they will never be the best.

Machine translators, in the meantime, should be viewed as translation aids, with the human translator acting as a posteditor. MT can serve as a foundation for professional translators to revise, reformulate, improve the writing style and, most importantly, localize the material to suit the context and audience in the target language. This means that rather than translating a text from scratch, the translator double-checks, proofreads and revises a machine-translated text. One significant benefit of such pairing between human and machine is that it boosts the translator's productivity.

The relationship between machine and human is complementary. According to statistics and current research, new technology such as MT will never be able to completely replace humans; instead, it will aid rather than threaten them (Dia *et al.*, 2022; Şahin and Gürses, 2021). A translator who is proficient in MT will have a competitive advantage over those who are not familiar with current translation technology.

As a result, it appears that human translators' anxieties about being replaced by machines in the future are unjustified. Nonetheless, it appears that the translator's position will inevitably evolve in the future. Human translators may no longer be accurate translators, but rather editors and editing materials previously translated by machines, as MTs become more advanced.

6. Recommendations
The findings of this study can shed more light on the differences in translation adequacy between human and MT. This research is important because it paves the path for a theoretical framework for MT accuracy. In general, it is a real endeavor to gain a better understanding of the efficiency of MT vs HT in translating Arabic–English texts, and it will be useful to translators, students, educators and experts in the field.

Software developers involved in the field of MTs can benefit from the findings to improve MT adequacy. The findings may also be useful for experts in the field in conducting comparative studies in the realm of machine vs human translation.

MT research is not limited to adequacy; it can also look at other facets of the technology. Another area of investigation is MT fluency and naturalness. The focus of this study was to translate Arabic (source text) to English (target text), however; research can be extended to other language pairs. In addition, the current investigation was based on short texts; therefore, it is also recommended that other research with a longer stretch of texts should be conducted to establish generalizability. Likewise, this work is limited to only two MT systems (Google translation and Babylon translation). Other MT systems and linguistic aspects, as well as more texts and HTs, may be investigated in future research.

Technological advancements in the form of MTs have had, and continue to have, significant widespread ramifications for translators and nontranslators equally in everyday scenarios professionally and personally, where the scope of the human translator has been obscured by a growing selection of comparatively straightforward and online MT systems that do not commonly show users where their translations have come from or how good the quality is. Therefore, a more in-depth interview-based study to gain insights into translators' experiences, preferences and perspectives as to MTs' impact on the future of HT is also recommended.

References

Abusaaleek, A.O. (2016), "The adequacy and acceptability of machine translation in translating the Islamic texts", *International Journal of English Linguistics*, Vol. 6 No. 3, pp. 185-193, doi: 10.5539/ijel.v6n3p185.

Akan, M.F., Karim, M.R. and Chowdhury, A., M. K. (2019), "An analysis of Arabic-English translation: problems and prospects", *Advances in Language and Literary Studies*, Vol. 10 No. 1, pp. 58-65, doi: 10.7575/aiac.alls.v.10n.1p.58.

Almutawa, F. and Izwaini, S. (2015), "Machine translation in the Arab world: Saudi Arabia as a case study. Trans-Kom", *Journal of Translation and Technical Communication Research*, Vol. 8 No. 2, pp. 382-414.

Bowker, L. (2019), "Fit-for-purpose translation", in O'Hagan, M. (Ed.), *The Routledge Handbook of Translation Technology*, Routledge, London, pp. 453-468.

Chesterman, A. (2016), *Memes of Translation: The Spread of Ideas in Translation Theory*, J. Benjamins, Amsterdam, doi: 10.1075/btl.123.

Dai, H., Xhafa, F., Janse, B.J., LIang, H. and Ye, J. (2022), "Comparative analysis of machine translation and human translation under the background of the Internet", *International Conference on Cognitive-based Information Processing and Applications (CIPA 2021)*, Vol. 1 No. 84, pp. 877-882.

Delpech, E.M. (2014), *Comparable Corpora and Computer-Assisted Translation*, 1st ed., Wiley, London, doi: 10.1002/9781119002659.ch1.

Doherty, S. (2016), "The impact of translation technologies on the process and product of translation", *International Journal of Communication*, Vol. 10, pp. 947-969.

Fehri, A. (2012), *Key Features and Parameters in Arabic Grammar*, John Benjamins Publishing, Amsterdam, doi: 10.1075/la.182.

Gerber, L. (2012), "Machine Translation: ingredients for productive and stable MT deployments – Part 2", available at: www.translationdirectory.com.

Hutchins, W.J. (2001), "Machine translation over fifty years", *Histoire epistémologie langage*, Vol. 23 No. 1, pp. 7-31, doi: 10.3406/hel.2001.2815.

Li, H., Graesser, A.C. and Cai, Z. (2014), "Comparison of Google translation with human translation", *Paper presented at Twenty-Seventh International Florida Artificial Intelligence Research Society Conference*, USA.

Maghsoudi, M. and Mirzaeian, V. (2020), "Machine versus human translation outputs: which one results in better reading comprehension among EFL learners?", *The JALT CALL Journal*, Vol. 16 No. 2, pp. 69-84, doi: 10.29140/jaltcall.v16n2.342.

Martindale, M.J. and Carpuat, M. (2018), *Fluency over Adequacy: A Pilot Study in Measuring User Trust in Imperfect MT*, Cambridge Univ. Press, Cambridge, doi: 10.1017/CBO978051198139.

McEnery, T. and Hardie, A. (2012), *Corpus Linguistics: Method, Theory, and Practice*, 1st ed., Cambridge Univ. Press, Cambridge, doi: 10.1017/CBO9780511981395.

McGillivray, B. and Kilgarriff, A. (2013), "Tools for historical corpus research, and a corpus of Latin", in Bennett, P., Durrell, M., Scheible, S. and Whitt, R.J. (Eds), *New Methods in Historical Corpus Linguistics*, Narr, Tübingen.

Mendel, Y. (2016), "German orientalism, Arabic grammar and the Jewish education system: the origins and effect of Martin Plessner's 'theory of Arabic grammar'", *Naharaim*, Vol. 10 No. 1, pp. 57-77, doi: 10.1515/naha-2016-0004.

Munkova, D., Munk, M., Welnitzova, K. and Jakabovicova, J. (2021), "Product and process analysis of machine translation into the inflectional language", *SAGE Open*, Vol. 11 No. 4, doi: 10.1177/21582440211054501.

O'Brien, S. (2012), "Towards a dynamic quality evaluation model for translation", *The Journal of Specialized Translation*, No. 17, pp. 55-77.

Papineni, K., Roukos, S., Ward, T. and Zhu, W.J. (2002), "Bleu: a method for automatic evaluation of machine translation", *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics, Philadelphia, PA, pp. 311-318, doi: 10.3115/1073083.1073135.

Puchała-Ladzińska, K. (2016), "Machine translation: a threat or an opportunity for human translators?", *Studia Anglica Resoviensia T*, Vol. 3 No. 2016, pp. 89-98, doi: 10.15584/sar.2016.13.9.

Richards, J.C. and Schmidt, R. (2010), *Longman Dictionary of Language Teaching and Applied Linguistics*, 4th ed., Pearson Education, Edinburgh.

Şahin, M. and Gürses, S. (2021), "English-Turkish literary translation through human-machine interaction", *Revista Tradumàtica*, Vol. 19, pp. 306-310, doi: 10.5565/rev/tradumatica.284.

Shahata, L. (2020), "Sentence translation challenges among Arabic-speaking EFL students", *Open Journal of Modern Linguistics*, Vol. 10, pp. 321-339, doi: 10.41236/ojml.2020.104019.

Specia, L., Hajlaoui, N., Hallett, C. and Aziz, W. (2011), "Predicting machine translation adequacy", *Machine Translation Summit*, Vol. 13 No. 2011, pp. 19-23.

Vasheghani Farahani, M. (2020), "Adequacy in machine vs human translation: a comparative study of English and Persian languages", *Applied Linguistics Research Journal*, Vol. 4 No. 5, pp. 84-104, doi: 10.14744/alrj.2020.98700.

Vieira, L.N. and Alonso, E. (2020), "Translating perceptions and managing expectations: an analysis of management and production perspectives on machine translation", *Perspectives*, Vol. 28 No. 2, pp. 163-184, doi: 10.1080/0907676X.2019.1646776.

Waddington, C. (2001), "Different methods of evaluating student translations: the question of validity, Meta", *Translators' Journal*, Vol. 46 No. 2, pp. 311-325.

Way, A. (2018), "Quality expectations of machine translation", in Moorkens, J., Castilho, S., Gaspari, F. and Doherty, S. (Eds), *Translation Quality Assessment*, Springer, Cham, pp. 159-178, doi: 10.48550/arXiv.1803.08409.

Appendix

Level	Accuracy of transfer of ST content	Quality of expression in TL	Degree of task completion	Mark
Level 5	Complete transfer of source text information, only minor revisions to reach a professional standard	Almost all translation reads like a piece originally written in English. There may be minor lexical, grammatical, or spelling errors	Successful	9,10
Level 4	Almost complete transfer; there may be one or two insignificant inaccuracies; requires a certain amount of revision to reach a professional standard	Large sections read like a piece originally written in English. There are a number of lexical, grammatical or spelling errors	Almost completely successful	7,8
Level 3	Transfer of the general idea(s) but with a number of lapses inaccuracy; needs considerable revision to reach a professional standard	Certain parts read like a piece originally written in English, but others read like a translation. There are a considerable number of lexical, grammatical or spelling errors	Adequate	5,6
Level 2	Transfer undermined by serious inaccuracies, through revision required to reach a professional standard	Almost the entire text read like a translation; there are continual lexical, grammatical or spelling errors	Inadequate	3,4
Level 1	Totally inadequate transfer of ST content; the translation is not worth revising	The candidate reveals a total lack of ability to express himself adequately in English	Totally inadequate	1,2

Source(s): Waddington (2001)

Table A1.
Scale for holistic
model C

PRR

About the author

Dr. Muneera Muftah is an Associate Professor of Applied Linguistics and SLA at the Department of English, Faculty of Arts, Thamar University, Yemen. She is currently working in the Department of English Language at the College of Languages and Translation, Najran University, KSA. She earned Ph.D. in English Language Studies from Universiti Putra Malaysia, Malaysia, and completed a postdoctoral fellowship at the Faculty of Modern Languages and Communication, UPM. She teaches courses in linguistics, applied linguistics and translation. Her main research interests are in the areas of translation technologies, syntactic and morphological mental representation and development, vocabulary development in SLA, generative syntax and morphology, discourse studies and second language assessment. Currently, she works on information and communication technologies (ICT) in English language teaching and learning, machine translation (MT) and language learning. Muneera Muftah can be contacted at: munmef5@gmail.com