A Cognitive Approach to Machine Translation Post-editing of English-Arabic Literary Texts

Sama Dawood Salman
(Associate Professor of Translation and Interpreting)
(English Department – Faculty of Al-Alsun & Mass Communication)
Misr International University

Abstract

Machine Translation (MT) of literary texts is a rich area of research that has not been investigated extensively. The conventional wisdom is that literary translation by nature is challenging even for the most competent translators as it requires preserving not only the meaning, but also the style of the author. This makes the MT output of any literary work viewed as inadequate no matter how accurate it is in conveying the meaning. However, while the standards of literary translation are still high for machines, the recent advances in MT seem promising and could open the door to an adequate post-edited MT for literary texts. This is a descriptive study that looks into the limitations and potential of MT in facilitating the task of literary translators. It analyzes the language errors in the Arabic MT of selected short English literary texts. The MT is generated from three systems: IBM Watson, Bing Microsoft, and Google Translate. The language errors in the raw MT output are detected by two professional post-editors, and classified by the researcher according to Temnikova's (2010) cognitive approach of MT post-editing. The findings of the study show that the highest number of errors fall under categories that are cognitively easy to correct by post-editors. It is hoped that this contribution can be useful to build MT tools with more adequate output for English-Arabic translation in general and literary translation, in particular.

Keywords: cognitive approach; error ranking; literary texts; machine translation, post-editing

نهج ذهني للتحرير اللاحق للنصوص الأدبية المترجمة آليًا من الإنجليزية إلى العربية

المستخلص

تعتبر الترجمة الآلية للنصوص الأدبية مجالًا ثريًا للبحث إلا أن ما كتب فيه مازال محدودًا. ثمة اتفاق شائع على أن الترجمة الأدبية بطبيعتها تمثل تحديًا حتى بالنسبة للمترجمين الأكثر كفاءة لأنها لا تقتصر على نقل المعنى، وإنما تسعى كذلك إلى الحفاظ على أسلوب المؤلف. وعليه فإن مخرجات الترجمة الآلية لأي عمل أدبي يُنظر إليها على أنها ترجمة لا تتميز بالجودة المطلوبة بغض النظر عن مدى دقتها في نقل المعنى. غير أنه وبالرغم من أن مستوى الترجمة الآلية مازال بعيدًا عن معايير جودة الترجمة الأدبية، فإن التطورات الحديثة في الترجمة الآلية تبدو واعدة ويمكن أن تفتح المجال أمام الترجمة الآلية الملائمة للنصوص الأدبية. تهدف هذه الدراسة الوصفية إلى البحث في إمكانات الترجمة الآلية في تسهيل مهمة المترجمين الأدبيين من خلال تحليل الأخطاء اللغوية في الترجمة الآلية العربية لنصوص أدبية قصيرة مختارة باللغة الإنجليزية. تقوم الدراسة على تحليل الأخطاء اللغوية التي تولدها ثلاثة أنظمة في الترجمة الآلية والمقارنة بينها من خلال الاستعانة بمحررين محترفين وفقًا لتصنيفات النهج الذهني للتحرير اللاحق الذي تقترحه إيرينا تمنيكوفا (2010). تشير نتائج التحليل إلى أن أكبر عدد من الأخطاء اللغوية الناتجة عن الترجمة الآلية للنصوص المختارة يمكن تصنيفه ضمن تلك التي يسهل على المحررين اللاحقين تصحيحها. تكمن أهمية هذه الدراسة في تسليط الضوء على أخطاء الترجمة الآلية بهدف الوصول إلى مخرجات أكثر دقة في مجال الترجمة عمومًا والترجمة الأدبية على وجه الخصوص بين الإنجليزية والعربية.

A Cognitive Approach to Machine Translation Post-editing of English-Arabic Literary Texts

Sama Dawood Salman
(Associate Professor of Translation and Interpreting)
(English Department – Faculty of Al-Alsun & Mass Communication)
Misr International University

1. Introduction

Despite the great advances in Machine Translation (MT), the adequacy of machine output is still questionable. This is especially the case when evaluating the MT between languages that belong to different families such as English and Arabic with their distinct linguistic and cultural systems. In certain types of texts such as those of legal and scientific nature, MT has proved to be a practical solution as it saves time and money. In literary texts, however, the accuracy of the machine output is still inferior to human translation. This has discouraged studies on MT quality when it comes to literary texts.

Building on the limited research conducted in this area, the present study investigates the type of MT literary translation language errors generated by three MT systems. It aims to conduct a qualitative analysis of the errors from a post-editing perspective. This research fills a gap in the study of literary MT between English and Arabic. Most previous studies in this field focused either on comparing between MT and human translation for pedagogical purposes (see Yamin 2014; and Omar & Gomaa 2020), or on conducting automated or manual assessment of the quality of MT (see Almahasees 2017; Al-khresheh & Almaaytah 2018). Analyzing MT errors as a means to a further end in order to direct the post-editing phase is still an under-researched topic. The argument advocated by the study is that although MT may not reflect the stylistic features of literary texts, it should not be excluded altogether. A post-editing phase that is based on full understanding of the nature of errors produced by the machine can make literary translators benefit from the advantages of MT and guide developers of MT systems towards more efficient tools.

2. Literature review

2.1 Machine Translation of Literary Texts

Initial attempts to study MT date back to the end of 1940s (Poibeau, 2017, p. 50). However, it is only during the past decade that the usability of MT in translating literature has started to attract the attention of scholars. In what follows a survey of the most significant contributions in this area, such as those of Voigt and Jurafsky (2012), Almahasees

(2017), Al-khresheh and Almaaytah (2018), Omar and Gomaa (2020), and Guerberof-Arenas and Toral (2022). It is to be noted that there are some relevant studies on the MT of language pairs other than English and Arabic included in this review. While the findings of such studies may not necessarily reflect the type of challenges involved in English-Arabic MT, they clearly indicate a gap in the number of studies in the language combinations which this study focuses on.

To begin with, Voigt and Jurafsky (2012) examine the effectiveness of MT compared to human translation of literary texts in terms of referential cohesion. The findings show that human translators are more able to get the referential cohesion of the text than the MT system. They view their study as a significant step that can spearhead research in the field of literature MT, and encourage more interested scholars (2012, p.18). The years that followed Voigt and Jurafsky's (2012) study have witnessed an increasing interest in MT of literature, showing contradictory results.

Yamin (2014) compare between machine translation and human translation of English literary texts translated into Arabic, taking the sentence to be the unit of analysis. The author maintains that the results of the comparison can help novice translators better understand the nature of literary translation. He recommends incorporating grammar lessons into translation classes (2014, p.159).

Besacier and Schwartz (2015) attempt translating a short story from English to French using a phrase-based statistical MT system. They conclude that the quality of the output is acceptable with some post-editing effort though it lacks creativity (2015, p.120).

Comparing the output of two MT systems is the main aim of Almahasees's (2017) study. He evaluates the adequacy of Google Translate and Microsoft Bing in translating Khalil Gibran's *The Prophet* from Arabic into English. The evaluation, which is based on BLEU automatic evaluation metric, shows that the MT output is not accurate and sometimes incomprehensible due to failure to handle metaphors and cultural peculiarities. The author recommends future studies in the correlation between automatic and manual evaluation approaches to have deeper insights into the applicability of MT systems to literary translation. Automatic evaluation alone is proved to be unreliable because it compares the MT output to a human translation (2017, p.158).

Al-khresheh and Almaaytah (2018) investigate the quality of MT of English proverbs into Arabic. The aim of their study is to shed light on the lexical and syntactic obstacles encountered by Google translate when

dealing with proverbs. To overcome these obstacles, the researchers suggest designing a special MT system that is fed by an interpretation of both English and Arabic proverbs. They recommend conducting comparative studies on different MT systems (2018, p.163).

Furthermore, Toral and Way (2018) assess the quality of neural machine translation (NMT) of novels through building a special NMT system that is fed with over 100 million words. They compare the quality of the NMT output to that of phrase-based statistical machine translation (PBSMT) according to BLEU metric. The comparison show that there is an 11% improvement in the quality of NMT over PBSMT (2018, p.263). Based on the findings of the study, they recommend further effort in this area to make MT part of professional literary translation workflow (p.286).

Approaching MT of literary texts from a different perspective, Taivalkoski-Shilov (2019) underline some ethical issues that need to be taken into consideration in this regard. She maintains that poor quality MT of such texts can harm the reputation of the original authors by overlooking their stylistic features, and allow linguistic abnormalities into the target language (2019, pp. 691-92).

Fonteyne, Tezcan, and Macken (2020) evaluate the raw MT of a whole novel by Agatha Christie from English to Dutch using Google system. The results show that 44% of the translated sentences were error-free. The results also reveal that mistranslated sentences and lack of coherence are the most frequent problems of the output (p. 3784).

Back to the ethical concerns of MT of literary texts, Kenny and Winters (2020) are specifically concerned about the voice of the translator. The authors carry out an experiment in which they ask a professional translator to produce two translations for a literary work from English into German: one is totally human, and the other is a post-edited MT. The results show the translator's voice is almost invisible in the post-edited translation compared to its clear presence in his own translation. They recommend further studies in literary MT that put the focus on human translators (2020, p.147).

The study of Omar and Gomaa (2020) is perhaps the most significant contribution on the problems of MT of literary texts in an English-Arabic context though their main concern is pedagogical. They use two MT applications, namely Google Translate and QTranslate. Two literary works are translated by the machines and then the resultant translations are compared to human translations. The errors in the machine translation are on the lexical, structural, semantic, and pragmatic

levels. The authors conclude that "literary translation is not a job for which MT systems have been designed" (2020, p.232).

Guerberof-Arenas and Toral (2022) compare the translation quality of a short story from English to Catalan and Dutch using machine translation, post-edited machine translation, and human translation. The three translations are evaluated by experts, and the highest creativity score is given to the human translation, while they view both MT and post-edited MT as poor and not fit for publication.

This study is yet another attempt to explore the reliability of MT systems to translate literary texts. It is hypothesized that understanding the nature of errors in MT output is a pivotal step towards achieving publishable post-edited MT of literature.

2.2. Assessment of cognitive load of post-editing

Cognitive load (CL) is defined as "a variable that attempts to quantify the extent of demands placed by a task on the mental resources" (Chen et al. 2016). Within the context of MT, a number of methods have been proposed to measure the kind of efforts that translators need to exert to post-edit a machine-translated text. They could be classified into two main approaches: physiological sensor-based and translation accuracy-based approaches. Moorkens (2018), Herbig, Pal, Vela, Krüger, and van Genabith (2019) and (2021), as well as Almanna, Jamoussi, (2022) are among the scholars who extensively investigate the issue of cognitive load during MT post-editing.

In order to record and measure eye-tracking indicators to two machine-translated texts, Moorkens (2018) uses a special online editing tool and an eye-tracker to record pupil dilation, fixation duration and fixation count. The importance of the study lies mainly in the final section where the researcher draws attention to the limitations of this approach to measuring the cognitive load. The small number of participants willing to take part in the experiment, and timetabling eye-tracking sessions are listed among the main difficulties involved in conducting such studies (p. 61).

Rather than focusing only on eye fixation/movement during the post-eding process, Herbiga, Pala, Krügera, and van Genabitha (2019) develop an innovative model to measure the cognitive load in terms of an array of physiological and behavioral indicators such as those related to eye, skin, and heart. Their sensors include a keylogger to detect and save keyboard strokes, a high-quality eye-tracker to count the amount of gazing, blinking, and fixation, a galvanic skin response sensor, and a heart belt to measure heart rates during post-editing. Ten translators have participated in the study and the data gathered indicate that this multi-

modal measurement approach is better at assessing the level of cognitive load than other single-modal approaches (p. 98).

Moving to a translation accuracy-based approach, Almanna, and Jamoussi (2022) adopts cognitive classification that includes point of emphasis, plexity, scope of intention and extent of causation, pace and time lapse, state of dividedness, state of boundedness, and degree of extension. The findings of their study reveal that measuring the cognitive load in terms of morphological, structural, and contextual aspects is more useful in the case of translation between English and Arabic (p. 325).

This study intends to measure the cognitive load during post-editing machine-translated literary texts in terms of an error-classification approach. This is the focus of the next section.

3.1 Theoretical framework: A cognitive approach to MT post-editing

In order to explore the process of MT post-editing, Temnikova (2010) sketch out a cognitive model that measures the quality of MT product in terms of the efforts that need to be exerted by post-editors. Temnikova draws heavily on MT error classifications suggested earlier by Vilar, D'Haro, and Ney (2006), and develops it through providing an explanation of the errors from a cognitive perspective, taking into account the effort that post-editors should exert in order to improve the MT output (2010, p. 3487).

According to Vilar et al (2006, p. 699), MT errors fall into four main categories: 1) missing word; 2) word order; 3) incorrect words; and 4) punctuation errors. All these categories are further sub-classified into minor categories as shown in Table 1.

Table 1: MT error classification by Vilar et al (2006, p. 699) with explanation provided by Temnikova (2010, p. 3488)

Type	#	Sub-type	Explanation
Missing word	1.1	Missing content word	Error correction requires adding the missing word
	1.2	Missing filler word	
Word order	2.1	Word level	Error correction requires moving single words
	2.2	Phrase level	Error correction requires moving whole phrases
Incorrect words	3.1	Wrong word	Error correction requires replacing the wrong word with a completely different one
	3.2	Correct word with an incorrect ending	Error correction requires replacing with the correct ending
	3.3	An incorrect word	Error correction requires using a synonym
	3.4	Extra word	Error correction requires deleting the extra word
	3.5	Error due to incorrectly recognized idiomatic expressions	Error correction requires replacing with the correct translation of the idiomatic expression

| Punctuation errors | 4.1 | Missing punctuation sign | Error correction requires adding the missing punctuation sign(s) |
| | 4.2 | Incorrect punctuation sign | Error correction requires replacing the incorrect punctuation sign(s) with the correct one(s) |

Temnikova upgrades Vilar et al's error classification by linking it to the cognitive effort of post-editors to detect and correct those errors. She suggests an additional classification that involves error ranking, arranging MT error correction from the easiest (1) to the hardest (10) as shown in Table (2).

Table 2: Cognitive MT error ranking (Temnikova, 2010, p. 3488)

Morphological level	1. Correct word, incorrect form
Lexical level	2. Incorrect style synonym
	3. Incorrect word
	4. Extra word
	5. Missing word
	6. Idiomatic expression
Syntactic level	7. Wrong punctuation
	8. Missing punctuation
	9. Word order at word level
	10. Word order at phrase level

Temnikova (2010, p. 3488) points out that this ranking is based on the findings of studies in comprehension and memory due to the similar nature between these two areas and the task of post-editing. The errors that require only memory activation of already stored lexical representations (i.e. morphological and lexical errors) are "less cognitively costly" than those which involve processing of the whole sentence (i.e. syntactic level). It is to be noted that Vilar et al's (2006) classification as well as that of Temnikova's (2010) exclude MT stylistic errors since they do not result in change of meaning. Therefore, stylistic errors will also be ruled out in this study.

3. Methodology

The Arabic MT of excerpts from three English literary texts were analyzed to answer the following questions:

1- What are the types of language errors in the MT of literary texts?
2- Are there similarities between the MT systems concerning the types of errors generated?
3- What are the cognitive efforts required to post-edit MT of literary texts?

Short texts of about 180 words each excerpted from three English novels were selected to be the source language texts: *Animal Farm* by George Orwell, *A Tale of Two Cities* by Charles Dickens, and *Eat, Pray, Love* by Elizabeth Gilbert. The three texts were chosen because they belong to different time periods and each one of the authors has a unique style of writing. This is to make sure that the source language texts although are all literary, contain various difficulties and their MT would yield different types of errors. The selected excerpts from the three works contain some challenging aspects for translators such as the use of archaic words, figurative language, words with no equivalents in the target language, and complicated sentence structure.

The target language texts were the Arabic MT of these excerpts generated by three distinct MT systems: IBM Watson, Bing Microsoft, and Google Translate. The MT output errors were first identified by two professional English-Arabic post-editors, and then classified by the researcher according to Temnikova's MT error ranking (2010).

4. Analysis of data
Table 3: MT errors of excerpt (1) - *Animal Farm*

English Text	MT System	Error	Rank	Explanation
… of the Manor Farm	IBM Watson	من مزرعة مانور	3	The preposition من *min* (literally means from) does not convey the relation of belonging expressed by the preposition "of". Error correction requires using a different word.
	Bing Microsoft	من مزرعة (مانور)	3	
	Google Translate	من مزرعة (مانور)	3	
Mr. Jones … locked the hen-houses	IBM Watson	حبس بيوت الدجاجة	2 + 1	The first error is the translation of the verb "locked". Although one of its meaning in Arabic is حبس *ḥabasa* (literally means prison) it does not seem the right option here because it does not collocate with بيوت الدجاج. A better option would be أغلق. The other error is a morphological one. The plural form of الدجاجة *al-dajaja* (hen) should be used. Thus, error corrections require using a synonym in case of حبس and a word with a different ending in the case of دجاجة.

English Text	MT System	Error	Rank	Explanation
	Bing Microsoft	أغلق بيوت الدجاجة	1	The two systems again translated "Hen" as دجاجة because it is used in its singular form in English. This is a morphological error because in Arabic it should be used in the plural form. Error correction requires using the same word with a different ending.
	Google Translate	أغلق بيوت الدجاجة	1	
popholes	IBM Watson	الفتحات	-	No error
	Bing Microsoft	الثقوب	2	This is not the adequate translation that matches the context. فتحات *fataḥat* seems a better option to describe a chicken house. Error correction requires using a synonym.
	Google Translate	Not translated	5	A missing word error as the word is not translated.
With the ring of light from his lantern dancing from side to side, he lurched across the yard	IBM Watson	مع خاتم الضوء من فانوسه الراقص من جانب إلى آخر، وقال انه طارد في جميع أنحاء الساحة	3 + 4 + 10	"Ring" is wrongly translated as خاتم *ḫatam* قال is an extra word that is not in the English text, "lurched" means staggered and is not طارد *ṭarada* (literally means chased), and the whole phrase needs to be re-structured according to Arabic rules starting with the main clause then the subordinate or inserting a verb to the subordinate clause. Error corrections require using a different word instead of خاتم, delete the extra word قال, and change the word order of the sentence.

English Text	MT System	Error	Rank	Explanation
	Bing Microsoft	مع حلقة الضوء من فانوس له الرقص من جانب إلى آخر، وقال انه ترنح عبر الفناء	1 + 10	فانوس له *fanus lahu* is not the correct way in Arabic to say "his lantern". The possessive relation is expressed in Arabic by using the joined pronoun ه as in فانوسه. The adjective "dancing" is translated wrongly as the noun الرقص and قال is an extra word that is not in the SL. The word order is awkward and does not follow the Arabic rules. Error corrections require using the same word with a different ending, using a different form of the word الرقص and change the word order.
	Google Translate	مع حلقة الضوء من فانوسه يرقص من جانب إلى آخر، كان يتجول في الفناء	10	The main problem here is the word order. Starting with the main clause is more preferable in Arabic. Error correction requires changing the word order.
kicked off his boots at the back door	IBM Watson	ورفس حذائه في الباب الخلفي	3	The preposition "at" is wrongly translated as في *fī*. Error correction requires using a different word.
	Bing Microsoft	بدأ حذائه عند الباب الخلفي	2	The verb "kicked off" in this context means that he took off his shoes by kicking the door but not بدأ *bada'* (literally means started). Correction requires using a synonym.
	Google Translate	وركل حذائه من الباب الخلفي	3	The preposition "at" is wrongly translated as من *min*. Error correction requires using a different word.
drew himself a last glass of beer from the barrel in the scullery	IBM Watson	ورسم لنفسه آخر كأس من البيرة من برميل في نحت	3 + 2	The verb "drew" here means to take out a glass from the barrel and it is wrongly translated as رسم *rasama* which is one of the meanings of "draw" but not the one that matches this context. Error correction requires using a synonym. Scullery is like a mini kitchen and it is wrongly translated as نحت *naḥata* (literally means sculpture). Error corrections require using a different word.

English Text	MT System	Error	Rank	Explanation
	Bing Microsoft	ووجه نفسه كوب آخر من البيرة من برميل فيscullery	3 + 2	The verb "drew" is wrongly translated as وجه *wajaha* which is one of the meanings of "draw" but not the one that matches this context. Error correction requires using a synonym. No Arabic equivalent was given to "scullery". Error correction requires using a different word in the case of "drew", and "scullery" which was not translated.
	Google Translate	ووجه لنفسه آخر كوب من البيرة من البرميل في المنجد	3	Scullery is wrongly translated as المنجدal-*munjd* (literally means upholstered or saver). Error correction requires using a different word.
made his way up to bed	IBM Watson	وقطع طريقه إلى السرير	3	The translation of this phrase does not match the context. Error correction requires using a different word.
	Bing Microsoft	وجعل طريقه حتى السرير	3	جعل is not the right collocation with طريقto mean to head to somewhere. Error correction requires using a different word.
	Google Translate	شق طريقه إلى السرير	-	No error
As soon as the light in the bedroom went out there was a stirring and a fluttering	IBM Watson	بمجرد أن ضوء في غرفة النوم ذهب إلى هناك كان هناك ركاب وتناثر	3 + 10	"Went out" is wrongly translated as ذهب *ḍahaba* which does not collocate with ضوء. Stirring and fluttering are both mis-translated without considering the context. The whole structure is awkward and does not follow the Arabic word order. Error corrections require using different words and changing the word order.
	Bing Microsoft	بمجرد أن ذهب الضوء في غرفة النوم كان هناك إثارة ورفرفة	3 + 2	Went out is wrongly translated as ذهب *ḍahaba* which does not collocate with ضوء. Stirring is wrongly translated without considering the context. Error corrections require using a different word in the first error, and a synonym in the second one.

English Text	MT System	Error	Rank	Explanation
	Google Translate	بمجرد أن ينطفئ الضوء في غرفة النوم ، كان هناك ضجة ورفرفة	1	The tense of "went out" was wrongly translated as أن ينطفئ. Error correction requires using a different form of the word.
Word had gone round	IBM Watson	وكانت الكلمة قد ذهبت	6	An idiomatic phrase that should not be translated literally. Error correction requires replacing with the correct translation of the idiomatic expression
	Bing Microsoft	كان كلمة قد ذهب مستديرة	6	An idiomatic phrase that should not be translated literally. Error correction requires replacing with the correct translation of the idiomatic expression.
	Google Translate	كانت الكلمة قد دارت	6	An idiomatic phrase that should not be translated literally. Error correction requires replacing with the correct translation of the idiomatic expression.
old Major	IBM Watson	الرائد العجوز	3	Major is a proper noun that should not be translated. Error correction requires using a different word.
	Bing Microsoft	قديمة رائد	3 + 2 + 9	Major is a proper noun that should not be translated. Old is wrongly translated as قديمة with the wrong ending that does not match the gender of the noun. The word order follows the English one. Error corrections require using a different word for major, a synonym for old, and moving single words.
	Google Translate	الرائد القديم	3 + 2	Major is a proper noun that should not be translated. Old is wrongly translated as قديم *qadim*. Error corrections require using a different word for major, and a synonym for old.
wished to communicate it	IBM Watson	وكان يرغب في إيصاله	2	إيصاله is not the right collocation with a dream. Error correction requires using a synonym for communicate.
	Bing Microsoft	ورغب أن يبلغ هو	2 + 1	يبلغ *Yubaliġ* is not the right collocation with a dream. هوshould be replaced with the joined pronoun ه. Error corrections require using a synonym for communicate and a word with a different ending for يبلغ
	Google Translate	وتمنى توصيله	2 + 3	توصيله is not the right collocation The combination of the verb and preposition does not collocate with dream. Error correction requires using a synonym for communicate and different word (preposition).

English Text	MT System	Error	Rank	Explanation
as soon as Mr. Jones was out of the way	IBM Watson	حالما يخرج السيد جونز عن الطريق	6	To be out of the way is wrongly translated as يخرج عن الطريق. Error correction requires replacing with the correct translation of the idiomatic expression
	Bing Microsoft	بمجرد أن يكون السيد جونز للخروج من الطريق	6	To be out of the way is wrongly translated as يخرج عن الطريق. Error correction requires replacing with the correct translation of the idiomatic expression.
	Google Translate	بمجرد خروج السيد جونز من الطريق	6	To be out of the way is wrongly translated as يخرج عن الطريق. Error correction requires replacing with the correct translation of the idiomatic expression

Table 4: MT Errors of excerpt (2) – *A Tale of Two Cities*

English Text	MT System	Error	Rank	Explanation
it was the age of foolishness	IBM Watson	كان عصر البهر	3	"Foolishness" is mis-translated as البهر *al-buhr* (literally means dazzle). Error correction requires using another word.
	Bing Microsoft	كان عصر الحماقة	-	No error
	Google Translate	كان عصر الحماقة	-	No error
it was the epoch of incredulity	IBM Watson	Not translated	5	Error correction requires adding a missing word
	Bing Microsoft	كان عصر التشكك	-	No error
	Google Translate	كان حقبة الغموض	-	No error
it was the season of darkness	IBM Watson	كان موسم الظلام	-	No error
	Bing Microsoft	كان كان موسم الظلام	4	Delete extra wordكان
	Google Translate	كان موسم الظلام	-	No error
we had everything before us, we had	IBM Watson	كان لدينا كل شيء قبلنا، لم	2	"Before" is mis-translated as قبلنا . Error correction requires using a synonym. لدينا seems a more accurate suggestion.

English Text	MT System	Error	Rank	Explanation
nothing before us		يكن لدينا أي شيء قبلنا		
	Bing Microsoft	كان لدينا كل شيء قبلنا، لم يكن لدينا شيء قبلنا	2	"Before" is mis-translated as قبلنا (literally means in front of us). Error correction requires using a synonym.لدينا seems a more accurate suggestion.
	Google Translate	كان لدينا كل شيء أمامنا، لم يكن لدينا شيء أمامنا	-	No error
we were all going direct to Heaven	IBM Watson	كنا جميعا نسير مباشرة إلى السماء	2	Translating "heaven" as الجنة *al-janna* (literally means paradise) fits the context more. Error correction requires using a different word.
	Bing Microsoft	كنا جميعا ذاهبين مباشرة إلى السماء	2	Translating "heaven" as الجنة *al-janna* (literally means paradise) fits the context more. Error correction requires using a different word.
	Google Translate	كنا جميعا نذهب مباشرة إلى الجنة	-	No error
the period was so far like the present period	IBM Watson	كانت الفترة حتى الآن مثل الفترة الحالية	3	"So far" is mis-translated. Error correction requires using another word.
	Bing Microsoft	كانت الفترة حتى الآن مثل الفترة الحالية	3	"So far" is mis-translated. Error correction requires using another word.
	Google Translate	كانت الفترة تشبه إلى حد بعيد الفترة الحالية	-	No error
some of its noisiest authorities	IBM Watson	بعض من أكثرها ضوضاء. أصرت السلطات	9 + 2	The whole phrase is incomprehensible due to wrong word order. Noisiest is closest to صاخب *ṣaḵab* in this context rather than ضوضاء. Correction requires using a synonym.
	Bing Microsoft	بعض من سلطاتها صاخبة أصر	9 + 1	Wrong word order, the verb is placed at the end and the definite article ال is not attached to صاخبة. Errors correction require changing the word order and using a word

English Text	MT System	Error	Rank	Explanation
				in the correct form.
	Google Translate	أصرت بعض سلطاتها المزعجة	-	No error
… a queen with a plain face	IBM Watson	ملكة مع وجه عادي	3	عادي is an incorrect word that needs to be replaced by another. Plain is closest to ugly in this context. Correction requires using a different word. مع is the wrong preposition here and needs to be replaced by another word.
	Bing Microsoft	ملكة ذات وجه عادي	2	عادي is an incorrect word that needs to be replaced by another. Plain is closest to ugly in this context. Correction requires using a different word.
	Google Translate	ملكة ذات وجه عادي	2	عادي is an incorrect word that needs to be replaced by another. Plain is closest to ugly in this context. Correction requires using a different word.
a queen with a fair face	IBM Watson	وملكة مع وجه عادل	3 2	عادل is an incorrect word that needs to be replaced by another. "Fair" is closest to beautiful in this context. Correction requires using a different word. مع is the wrong preposition here and needs to be replaced by another word.
	Bing Microsoft	وملكة ذات وجه عادل	3	عادل is an incorrect word that needs to be replaced by another. "Fair" is closest to beautiful in this context.
	Google Translate	وملكة ذات وجه عادل	3	عادل is an incorrect word that needs to be replaced by another. "Fair" is closest to beautiful in this context.

Table 4: MT errors of excerpt (3) – *Eat, Pray, Love*

English Text	MT System	Error	Rank	Explanation
Eat, Pray, Love	IBM Watson	الأكل ، براي ، حب	3	The English word Pray is transliterated in Arabic as براي. Error correction requires using a different word.
	Bing Microsoft	تناول الطعام، الصلاة، الحب	-	The two verbs "pray" and "love" are translated as nouns. Error correction requires using another form for each word.
	Google Translate	كل صلى حب	-	No error
I refused that thought	IBM Watson	وأنا رفضت الفكر	1	الفكر *al-fikr* (literally means thinking) needs to be replaced with الفكرة *al-fikra* (literally means idea). Error correction requires using another form.
	Bing Microsoft	رفضت تلك الفكرة	-	No error
	Google Translate	رفضت هذا الفكر	1	Correct word, incorrect form. هذا الفكر needs to be replaced with هذه الفكرة
It would consume me	IBM Watson	تستهلك لي	2	"Consume" is translated literally while the adequate translation of the phrase "to be consumed by an idea" is تتمكن مني or تتملكني. "Me" is wrongly translated as لي *lī* (literally means for me). Error correction requires using other words.
	Bing Microsoft	تستهلك لي	2	"Consume" is translated literally while the adequate translation of the phrase "to be consumed by an idea" is تتمكن مني or تتملكني. "Me" is wrongly translated as لي *lī* (literally means for me). Error correction requires using other words.
	Google Translate	سيستهلك مني	2	"Consume" is translated literally while the adequate translation of the phrase "to be consumed by an idea" is تتمكن مني or تتملكني. Error correction requires using another word.
How could I be such a criminal jerk	IBM Watson	كيف يمكن أن أكون مثل هذا المجرم الأحمق	2	"Criminal jerk" is wrongly translated literally as المجرم الاحمق *al-mujrm al-aḥmaq*. Error correction requires using the correct word.
	Bing Microsoft	كيف يمكن أن أكون أحمقاً إجرامياً	2	"Criminal jerk" is wrongly translated literally as المجرم الاحمق *al-mujrm al-aḥmaq*. Error correction requires using the correct word.

English Text	MT System	Error	Rank	Explanation
	Google Translate	كيف يمكنني أن أكون رعشة إجرامية	2	"Criminal jerk" is wrongly translated literally as المجرم الاحمق *al-mujrm al-aḥmaq*. Error correction requires using the correct word.
to proceed this deep into a marriage	IBM Watson	ليمضي هذا بعمق في الزواج	3 + 1	"This deep" is literally translated. Error correction requires using the correct word. ليمضي is the correct word, but in the wrong format.
	Bing Microsoft	لأمضي في هذا العمق في الزواج	3	"This deep" is literally translated. Error correction requires using the correct word.
	Google Translate	لأستمر في هذا الزواج	-	No error
Hadn't I loved it	IBM Watson	ألم يكن لي أن أحب ذلك ؟	4	يكن لي أن are extra and unnecessary words. Correction requires deleting them.
	Bing Microsoft	ألم أحبذلك؟	1	Typographic mistake. The demonstrative article is joined to أحب. Error requires correcting the form.
	Google Translate	ألم أحبها؟	1	"It" is wrongly translated as referring to a feminine entity. The word is with the wrong ending, and requires changing the form.
So why was I haunting its halls every night, howling like Medea	IBM Watson	لذا لماذا كنت أطارد قاعاتها كل ليلة ، يعوي مثل ميديا ؟	1 + 3	يعوي is the correct word, but in the wrong form. "Medea" is usually translated into Arabic as ميدوسا *mīdosa*. Error correction requires using another word.
	Bing Microsoft	إذن لماذا كنت أطارد قاعاتها كل ليلة ، أعوى مثل (ميديا)؟	3 + 1	"Medea" is usually translated into Arabic as ميدوسا *mīdosa*. Error correction requires using another word.
	Google Translate	فلماذا كنت أطارد قاعاتها كل ليلة ، عويلًا مثل المدية؟	3 + 1	"Medea" is usually translated into Arabic as ميدوسا Error correction requires using another word. عويلا *'auela* is the correct word, but in the wrong form.
Wasn't I proud of all we'd accumulated	IBM Watson	ألم أكن فخورا	1	فخورا *faḵurn* is the correct word but in the wrong form as the speaker is a female.
	Bing Microsoft	لم أكن فخورا	1	فخورا *faḵurn* is the correct word but in the wrong form as the speaker is a female.

English Text	MT System	Error	Rank	Explanation
	Google Translate	لم أكن فخوراً	1	فخورا *faḵurn* is the correct word but in the wrong form as the speaker is a female.
the prestigious home	IBM Watson	المنزل المرموق	2	Wrong collocation. Correction requires using a synonym.
	Bing Microsoft	منزل المرموقة	2 + 1	Wrong collocation. Correction requires using a synonym. منزل *manzil* requires adding the definite article. Correction requires using the correct form.
	Google Translate	المنزل المرموق	2	Wrong collocation. Correction requires using a synonym.
… some box-shaped superstore	IBM Watson	بعض المتاجر الكبرى	5	Box shaped was deleted. Correction requires adding the missing words.
	Bing Microsoft	بعض سوبرستور مربع على شكل	3 + 9	"Superstore" is transliterated in Arabic, and box shaped is translated literally. Correction requires using different words. The word order is also wrong in مربع على شكل
	Google Translate	بعض الصناديق ذات الشكل المربع متجر	3 + 9	Wrong words and wrong word order.
buying ever more appliances on credit	IBM Watson	شراء المزيد من الأجهزة للائتمان	3	"On credit" is mis-translated. Error correction requires using another word.
	Bing Microsoft	وشراء من أي وقت مضى أكثر الأجهزة على الائتمان	9 + 3	"On credit" is mis-translated. Error correction requires using another word. There is also a problem with the word order.
	Google Translate	وشراء المزيد من الأجهزة من أي وقت مضى على الائتمان	5	"On credit" is mis-translated. Error correction requires using another word.
Why did I feel so overwhelmed with duty	IBM Watson	لماذا شعرت بأني مغمور بالواجب	2	"Overwhelmed" is mis-translated as مغمور. *Maġmur* (literally means nameless). Error correction requires using a synonym.
	Bing Microsoft	لماذا شعرت بالإرهاق الشديد من الواجب	-	No error

English Text	MT System	Error	Rank	Explanation
A Cognitive Approach to Machine Translation Post–editing of English–Arabic Literary Texts				
	Google Translate	لماذا شعرت بالإرهاق من واجبي	-	No error
the dog-walker	IBM Watson	الكلب	5	"Walker" was deleted. Correction requires adding the missing word.
	Bing Microsoft	الكلب ووكر	1	"Walker" was transliterated in Arabic. Correction requires using another word.
	Google Translate	وممشي الكلاب	3	"Dog walker" was translated literally. Error correction requires using a different word.
a writer	IBM Watson	كاتب	1	A word with a wrong ending as the speaker is a female.
	Bing Microsoft	كاتبة	-	No error
	Google Translate	كاتبة	-	No error

4.1 IBM Watson

According to the comments of the post-editors, the majority of errors with IBM Watson of *Animal Farm* fell under the category of "incorrect word" that should be replaced by either a different word, a synonym, or the same word but in a different form. Other errors included inability to recognize idiomatic expressions, and wrong word order. There were no punctuation errors.

The MT of *A Tale of Two Cities* included three types of errors. Two belonged to the category of incorrect word that requires a replacement by either a completely different word or a synonym. One single error was due to wrong word order.

Similar errors occurred in the translation of *Eat, Pray, Love* generated by IBM Watson. Most of the errors were due to using an incorrect word. There were two instances of a missing word.

4.2 Bing Microsoft

Six types of errors appeared in the MT of the first excerpt. Again the biggest number of errors was "incorrect word" that require replacement by a different word, a different form, or a synonym. There were also two errors in the word order: one requires moving single words, while the other is a major one that requires moving entire phrases.

A fewer number of errors appeared in the MT of the second excerpt. The majority were those related to "incorrect word" that require

using a different word or a synonym. There was one "extra word" error that requires deletion.

There were many errors in the MT of the third excerpt. Again most of them require selecting a different word or a synonym. There were also a number of "wrong word order" errors, one "extra word", and one error due to failure to recognize an idiomatic expression.

4.3 Google Translate

Moving to the MT generated by Google translate, there was one word order error type in the first text, while all the other errors were at the morphological and lexical level: an incorrect word, a word with an incorrect form, and incorrect synonym.

Google translation of the second text showed only two incorrect word errors that require replacement with different words.

The type of errors in the MT of the third text included one missing word, one wrong word order, one wrong translation of an idiomatic expression, one extra word, and a number of incorrect words.

5. Findings and Discussion

The results of analyzing the errors in the MT generated by the three systems for the selected literary texts indicate that the biggest number of errors occur at the morphological level (i.e., incorrect form) and the lexical level (i.e., incorrect style synonym, or incorrect word). Failure to recognize idiomatic expressions is a problem with the three systems, while errors related to extra word, missing word, and word order are mainly evident in Bing Microsoft MT. There are no punctuation errors in the three MT tools. It seems there is a correlation between the nature of the prose and the type of errors made by the MT system. Most cases of the incorrect word error type appear in the MT of the first text which includes words with rare usage (e.g. popholes, scullery), and words with synonyms (e.g. Major, old), the least number of word order errors occurs in the MT of the second text which consists of short and simple sentences, while most of the errors in the third text MT are those related to word order and form (feminine vs masculine form ending).

Referring to Temnikova's (2010) cognitive error ranking table, the highest number of errors could be grouped under the categories that are cognitively easy to correct by post-editors. Most of the errors require "activation in memory of previous representations and mental vocabulary look-up" (2010, p. 3488).

The findings of this study agree with those of Besacier and Schwartz (2015) who point out that the quality of MT is acceptable with some post-editing effort (p. 120). They also agree with the findings of Toral and Way (2018) who report improvement in the quality of MT

(p.263). The findings contradict with those of Omar and Gomaa (2020) who maintain that "literary translation is not a job for which MT systems have been designed" (2020, p. 232). Similarly, the results disagree with those of Guerberof-Arenas and Toral (2022) who conclude that a post-edited MT is poor.

With that being said, the problems of MT of literary texts could be summarized as follows:

- Problems due to the use of words that have more than one meaning (e.g. plain, fair).
- Problems due to the use of idiomatic expression (e.g. word had gone round, Mr. Jones went out of the way).
- Problems due to grammatical gender (e.g. proud, writer) and number (e.g. hen-house).
- Problems due to the inability of the system to translate all the words (e.g. Google failed to translate "popholes", and IBM Watson skipped "walker" in "dog-walker").
- Problems due to selecting wrong prepositions (e.g. of the Manor farm, a queen with a fair face).
- Problems due to following the word order of the source language (e.g. with the ring…, box shaped superstore).
- Problems due to wrong selection of collocations (e.g. made his way, prestigious home).
- Problems due to inserting words in the source language (e.g. scullery), and translating the same word twice (e.g. it was the season of darkness). Both of these errors were with the MT generated by Bing.

The fact that the three programs might produce different types of errors when translating other excerpts of the same three texts or even other texts from 20th century, is worth investigating in future studies. It is hoped that taking these problems into consideration when developing MT systems can improve their performance to a reasonable extent.

6. Conclusion and Future Direction

As MT seems to be moving from the peripheries of the translation practice closer to the center, it becomes important to investigate the potential of this technology in dealing with literary texts. The study findings showed that MT of literary texts generates sufficient quality that can be enhanced by careful post-editing process. Most of the MT errors of literary texts can easily be detected and corrected by post-editors. Since

MT can increase the productivity and speed of translators in various domains, it is time to develop more effective MT systems that can handle the challenges of translating literary texts so that putting effort in the MT post-editing stage becomes worth it. Integrating MT tools into the workflow of literary translators through post-editing can speed up their work, and help ensure consistency in cases of repetitiveness in the text. Definitely there will remain a number of stylistic features peculiar to literary texts that make MT unable to be on par with professional human translation. It is hoped that this contribution can be useful to build MT tools with more adequate output for literary translation.

Finally, with the growing interest in computer-aided translation, the coming few years should witness more researches that aim at suggesting innovative approaches to produce more accurate, error-free MT that can deal with a variety of text types. Relevant research should also be expanded to include building automated post-editing software for translation to and from Arabic that can reduce cognitive load in post-editing.

References

Almanna, Ali, & Jamoussi, Rafik. (2022). NMT verb rendering: A cognitive approach to informing Arabic-into-English post-editing. *Open Linguistics*, *8*(1), 310-327.

Al-khresheh, Mohammad H., and Shahab A. Almaaytah. (2018). English proverbs into Arabic through machine translation. *International Journal of Applied Linguistics and English Literature,* 7,158-166.

Almahasees, Zakaryia Mustafa. (2017). Machine translation quality of Khalil Gibran's *the Prophet*. *AWEJ for Translation & Literary Studies,* 1(4), 151-159.

Besacier, Laurent and Lane Schwartz. (2015). Automated translation of a literary work: A pilot study. *Proceedings of the Fourth Workshop on Computational Linguistics for Literature*, 114–122.

Herbig, Nico, Santanu Pala, Mihaela Vela, Antonio Krüger, and Josef van Genabith, Josef. (2019). Multi-modal indicators for estimating perceived cognitive load in post-editing of machine translation. *Machine Translation*, *33*(1), 91-115.

Chen Fang, Jianlong Zhou, Yang Wang, Kun Yu, Syed Z. Arshad, Ahmad Khawaji, and Dan Conway. (2016). *Robust multimodal cognitive load measurement*. Springer, New York.

Dickens, Charles. (2007). *A Tale of Two Cities*. London: Penguin.

Fonteyne, M., Tezcan, A., & Macken, L. (2020). Literary machine translation under the magnifying glass: Assessing the quality of an NMT-translated detective novel on document level. *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC),* 3783-3791.

Gilbert, Elizabeth. (2007). *Eat, Pray, Love: One Woman's Search for Everything*. London: A&C Black.

Guerberof-Arenas, Ana, and Antonio Toral. (2022). Creativity in translation: Machine translation as a constraint for literary texts. *Translation Spaces*. doi: https://doi.org/10.1075/ts.21025.gue

Kenny, Dorothy, and Marion Winters. (2020). Machine translation, ethics and the literary translator's voice. *Translation Spaces*, 9(1), 123-149. doi: https://doi.org/10.1075/ts.00024.ken

Moorkens, J. (2018). Eye tracking as a measure of cognitive effort for post-editing of machine translation. *Eye tracking and Multidisciplinary Studies on Translation*, 55-69.

Omar, Abdulfattah, and Yasser Gomaa. (2020). The machine translation of literature: Implications for translation pedagogy. *International Journal of Emerging Technologies in Learning*, 15(11), 228-235.

Orwell, George. (1945). *Animal Farm*. London: Faber & Faber.

Poibeau, Thierry. (2017). *Machine translation*. Massachusetts: MIT Press.

Taivalkoski-Shilov, Kristiina. (2019). Ethical issues regarding machine (-assisted) translation of literary texts. *Perspectives*, 27(5), 689-703.

Temnikova, Irina. (2010). Cognitive evaluation approach for a controlled language post-editing experiment. *Proceedings of the Seventh International Conference on Language Resources and Evaluation,* 3485 – 3490).

Toral, Antonio, and Andy Way. (2018). What level of quality can neural machine translation attain on literary text?. In Joss Moorkens, Sheila Castilho, Federico Gaspari, and Stephen Doherty (eds.), *Translation Quality Assessment: From Principles to Practice*, (pp. 263-287). Springer: Cham.

Vilar, David, Jia Xu, Luis Fernando d'Haro, and Hermann Ney. (2006). Error analysis of statistical machine translation output. *Proceedings of the fifth international conference on language resources and evaluation.* http://www.lrec-conf.org/proceedings/lrec2006/pdf/413_pdf.pdf

Voigt, Rob, and Dan Jurafsky. (2012). Towards a literary machine translation: The role of referential cohesion. *Proceedings of the NAACL-HLT 2012 Workshop on Computational Linguistics for Literature*, 18-25.

Yamin, Boulenouar Mohamed. (2014). A comparative study of machine and human translation: The case of English-Arabic literary translations. *RML*, 9, 154-160.