Journal of the University of Babylon for Humanities مَجلّةُ جَامعةِ بَابِل للعُلومِ الإنسَانيّةِ
Vol. 29/ No. 12/ 2021 المجلد 29/ العدد 12/ 2021

Literary Machine Translation: Towards a Post-editing Tool

Farah Abbas Abo Al-Timen
Lecturer at Department of Translation/College of Arts/ Al-Mustansiriyah University
farah86@uomustansiriyah.edu.iq

Submission date: 30/8/2021 **Acceptance date: 22/9/2021** **Publication date: 12 /12/2021**

Abstract

Given the remark able development that took place in machine translation (MT) in recent years, people all over the world tended to rely on MT since it provides a faster result than human translation (HT) and saves effort. Even with these merits, there is still a difference of opinion on whether MT output has reached human equivalence. This study was provoked by the following question "Does post-editing (PE) improve MT output?" The study attempts to answer the question by providing a theoretical approach to PE methods, with practical examples. Data was taken from the Arabicnovel *فرانكشتاين في بغداد Frankenstein in Baghdad* and the application of Google Translate (GT) was chosen for the automatic translation of the data. A professional translator with experience in literary translation was asked to do the task of PE Google output. Then an analysis of the grammatical and semantic differences - that have been found between GT and the PE version- was introduced. The research findings showed that conducting PE enhanced the efficiency of GT output, especially on the semantic level where the percentage of PE on the meanings (85%) was much more than on grammar (15%).

Keywords:Arabic, English, Frankenstein in Baghdad, Google Translate, Literary Text, Machine Translation, Post-editing.

الترجمة الآلية للنصوص الأدبية: نحو أداة تحرير لاحق

فرح عباس أبو التمن
قسم الترجمة /كلية الآداب/الجامعة المستنصرية

المستخلص

نظرًا للتطور الملحوظ الذي حدث في الترجمة الآلية في السنوات الأخيرة، أصبح الناس في جميع أنحاء العالم يعتمدون على هذه الترجمة؛ لأنها توفر نتيجة أسرع من الترجمة البشرية وتتطلب جهداً أقل. لكن برغم وجود هاتين الميزتين، إلا أنه لا يزال هناك اختلاف في الرأي حول ما إذا كانت مخرجات الترجمة الآلية قد وصلت إلى كفاءة الترجمة البشرية. تم إثارة هذه الدراسة من خلال طرح السؤال التالي "هل يؤدي التحرير اللاحق إلى تحسين مخرجات الترجمة الآلية؟" للإجابة على هذا السؤال قدمت الدراسة نهجاً نظريًا لأساليب التحرير اللاحق، فضلًا عن سرد أمثلة عملية.تم أخذ البيانات من رواية "فرانكشتاين في بغداد"، واعتمدت الدراسة على تطبيق ترجمة جوجل لترجمة المقاطع المختارة من الرواية، وطلب من مترجم محترف في مجال الترجمة الأدبية القيام بمهمة اجراء التحرير اللاحق على مخرجات ترجمة جوجل ثم قام الباحث بتحليل أوجه الاختلاف بين الترجمة الآلية والتحرير اللاحق من الناحية النحوية والدلالية. اظهرت نتائج البحث أن إجراء التحرير اللاحق على النص الأدبي عزز من كفاءة ترجمة جوجل، خاصة من الناحية الدلالية حيث كانت نسبة التحرير اللاحق على المعاني (78%) أكثر بكثير مما على النحو (22%).

الكلمات الدالة:الترجمة الآلية، اللغة العربية، اللغة الإنجليزية، ترجمة جوجل، تحرير لاحق، فرانكشتاين في بغداد، نص أدبي.

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

1 Introduction

Since the early 2000s,MTsystems have been accessible through using sources such as GT. The extent of quality of machine-translated output produced by the application of GT has been improved but not to the extent that qualifies it to live up to HT .Many people have tended to rely on GT extensively to perform their translation tasks and count on it without making any modifications, i.e., PE, on the output. Hence, they have an inaccurate translation with many grammatical and semantic mistakes.

PE of machine-translated output is now seen as a successful method of integrating MT with HT with the aim of minimizing time and effort in performing various translation tasks. The process of PE is defined by Wagner[1: 1]as the "correction of a pre-translated text rather than translation from scratch". Somers [2:138] also described PE as a process of "tidying up the raw output, correcting mistakes, revising entire, or, in the worst case, retranslating entire sections". Accordingly, PE means to observe a machine-translated output, detect and correct the errors to make sure that the MT meets a level of quality as that obtained by HT.

The big challenge for a machine system is the extent of its knowledge of a language vocabulary and grammar and how to identify and choose the suitable meaning and the correct structure since the machine system does not have an innate brain to think and act like a human brain does. It cannot, for example, determine the exact structure of sentence not it can choose an appropriate meaning for a word with multiple-equivalences according to the context that the word comes in. In view of that, PE can be of a great benefit in helping MT to give a human-like translation. Vitek [3:2]comments that "it is up to the reader to make sense of those words haphazardly jumbled up together by a non-thinking machine". On the other hand, professional translators can predict the incorrect meanings and structures that are produced by GT and they can take appropriate PE measures to exclude such inaccuracies. Henceforth, this study explores the benefit of applying PE on GT application to see the effect on the final translation result of GT and its adequacy compared to HT.

2.Literary MT and PE: Related Work

A great deal of effort has been devoted to developing a literary MT system with the help of PE in the past years regarding prose work[4], [5], [6], [7]. In the work of Moorkens et al.[4], the literary MT was studied through PE of the translation of the language pairs (English-to-Catalan). The study focused on showing the significance of the cooperative work of MT side by side with HT. Toral et al.[5], included working on an English novel that was translated automatically and then post-edited by professional literary translators. The study showed that PE led to decreasing the number of MT mistakes the increasing translation productivity. The work of Arenas & Toral[6] involved the translation of a story from English into Catalan that focused on showing the impact of PE on literary MT output and concluded that when professional translators post-edited MT, the output quality was the highest. As for the work of Ziganshina et al.[7], it involved assessing human PE to compare the performance of three MT engines - Google

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

one of them - of the language pairs(English to Russian).GT performed the best with highest average quality estimates for its initial MT output, and needed the lowest amount of human PE compared to other MT engines.

On the other hand, PE efficiency can also be measured by time-saving. Many academic studies have identified that PE is faster than translating from scratch, despite the translator's experience or language pairs [8]. Nevertheless, there is no approval on the exact time that can be saved by using PE. Other academic studies reported that time-saving can reach 40% [9]. Professional translators have also reported positive attitudes towards working with PE of MT where it required less effort than editing HT [10].

3.Method

To achieve the research goal of discovering the benefit of using PE on MT output, the methodology goes as follows: First, the Arabic literary text was translated via GT application. Second, a professional translator with experience in literary translation was asked to observe the Arabic source text (ST) and then read the target text (TT)– GT output. Third, the professional translator applied some modifications - PE- where needed on the GT output and afterwards, the PE was analyzed grammatically and semantically. Finally, GT and PE were compared with each other to analyze the similarities and differences found between GT and PE in order to give a comprehensive exploration of the PE process and identify the mistakes found in the GT output.

4. Data Collection and Discussion

The examples of extracts are taken from the Arabic novel *فرانكشتاين في بغداد*/*Frankenstein in Baghdad* by the writer*أحمد سعداوي*/*Ahmed Saadawi*. The application used for the translation of the extracts is GT[أ*].The direction of the study is from Arabic into English. The reason behind choosing this novel is because it has a rich variety of literary expressions and terms that set a good example for developing methods of PE literary machine-translated output.

The data analysis below show that although there are semantic and grammatical similarities observed between the translation of GT and the PE version given by the HT, there are also many significant differences that should be categorized and analyzed as in the following steps.

4.1 The Grammatical PE:

4.1.1:PEwith Subordinate Conjunctions

ST: (12 ،ص) ستسمع أصوات بناتها عبر الهاتف فتنسحب العتمة من صدر ها قليلا.

GT: She would hear the voices of her daughters on the phone, and the darkness withdrew from her chest a little.

PE: When she hears the voices of her daughters, the darkness in her chest will fade away a little bit.

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

The Arabic ST used the subordinate conjunction ف to link the two phrases ستستمع اصوات بناتها عبر الهاتف and فتنسحب العتمة من صرها قليلا in the above extract. GT translated the sentence literally connecting the two phrases by the subordinate conjunction *and* whereas a more natural translation in English is accomplished by introducing the subordinate phrase *ستستمع اصوات بناتها ...* with the subordinating conjunction *when* as done by the HT.

4.1.2:PE of Capitalization

ST: (185،ص)."البلدة" في مركز شرطة ..
GT: .., at the (town) police station.
HT: .., at Town/Balda Police Station.

There are errors of capitalization done by GT in which *مركز*, *شرطة* and *البلدة* had to be capitalized in the TT as done by the HT since *مركز شرطة "البلدة"* is a place title that should be capitalized.

4.1.3: PE by Passive Voice

Arabic and English structures are rather different in which Arabic tends to use active voice and English tends to use passive voice. In the following extract, HT post-edited the translation of GT and turned the ST active voice into passive voice:

ST: (ص، 289-290) ... شاهدت قطها "نابو" يدخل المطبخ الصغير على أثر رائحة الطعام،
GT: She saw her cat "Napo" enter the kitchenette after the smell of food.
PE: She saw her cat "Napo" enter the kitchenette, drawn by the smell of food.

It is clear that Google translation sounded unnatural by rendering *على أثر* into *after*, whereas the PE by passive voice was likely to be used here in this case to make the whole sentence sounds acceptable and natural.

4.1.4: PE of Relative Pronouns

In the following extract, GT couldn't render the relative pronoun *الذي* correctly in which the relative clause was referring to an animate person (*عزرائيل العظيم, الثقب الأسود*) that should be given the animate pronoun *who instead of that*. For that reason, HT resorted to post-edit the relative pronoun by changing it from *that* into *who*.

ST: (ص، 169) إنني الثقب الأسود و عزرائيل العظيم الذي سيبتلع هذا العالم كله.
GT: I am the black hole and Azrael the biggest that will swallow this whole world.
PE: I am the black hole and Great Azrael who will swallow this whole world.

4.2 The Semantic PE:

4.2.1:PEof Metaphorical and Idiomatic Expressions

ST: (ص، 17) رأسي يزدحم بالأفكار المتضاربة.
GT: My head is crowded with conflicting thoughts.
PE: My head is crowded with conflicting thoughts.

ST: (ص، 23)..تسقط في هاوية الضياع،
GT: fall into the abyss of loss, ..
PE: fall into the abyss of loss, ..

The ST metaphor can be retained as a metaphor having the same suitable effect on the TT since it does not have a cultural or idiomatic background. As in the previous two extracts in which both GT and HT rendered the metaphorical expressions *رأسي يزدحم*

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

تسقط في هاوية الضياع and بالأفكار literally. While in the following extract, the metaphor can exert a considerable translation problem due to the presence of an idiomatic expression:
ST: (126 ،إن اسمها حسب كلام المنجمين هو توابع الخوف. (ص
GT: Its name, according to astrologers, is the aftershocks of fear.
PE: Its name, according to astrologers, is the familiars of fear.
Here the rendering of the idiomatic expression توابع الخوف has to had a PE in order to avoid Google literal non-common translation of *aftershocks of fear*. A possible translation suggested by the HT was *familiars of fear*. Such translation coveys naturally the meaning of the local idiom shared among the ST people.

4.2.2: PE of Slang Language

ST: (32 ،كان هادي يسمي ناهم "المكرود"، .. (ص
GT: Hadi called Nahim "Al-Makrood"
PE: Hadi called Nahim "Poor thing/ Misfortunate"

Here, the Iraqi slang term مكرود could not be recognized by GT and it was rendered literally by resorting to transliteration. To avoid this. HT opted to post-edit Google transliteration and suggested the well-known slang idiomatic expression in the TT(*poor thing*) or the word (*misfortunate*) which both mean someone who has *extremely bad luck* and express the Arabic meaning of مكرود. Another example of slang language is found in the following extract:
ST: (21 ،خلي عينك بعين الله.(ص
GT: keep your eye with Allah's eye.
PE: look at the bright side.

It was rendered by GT into *keep your eye with Allah's eye.* This is an unnatural translation and a possible PE equivalent would be: *look at the bright side*. Such a rendering can deliver the ST slang meaning naturally and have the same effect on the TT readers as that found on the ST readers.

4.2.3:PE of Formal Language

ST:(106 ،المجد لله في العلى وعلى الأرض السلام وللناس المسرة. (ص
GT: Glory be to God in the highest and peace on earth and happiness to people.
PE: Glory be to God in the highest, and peace on earth, and goodness to men.

Formality of English language tends to be different from the formality of Arabic language in which in English there are certain words and structures that reflect the formal style. Crystal[11: 186]describes formality as the type of speech that involves certain choice of vocabulary and sentence syntax. In the above extract, GT succeeded in rendering the formality of the Holy Verse in some places and failed in others. The formality of this extract can be easily post-edited into an adequate formal English by substituting the word *happiness* with *goodness* and the word *people* with *men*.

4.2.4: PE by Lexical Substitution

ST: (174 ،حسب تعاليم المجنون الأكبر الذي لف عمامة برتقالية على رأسه وأسبل لحيته. (ص
GT: according to the teachings of the greatest madman who wrapped an orange turban on his head and tied his beard.
PE: according to the teachings of the greatest madman who wrapped an orange turban on his head and had a long beard.

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

Lexicality brings a translation problem because GT can easily fail to recognize the sense of a lexical word according to its context. In the above extract, GT rendered the lexical verb *أسبل* into *tied* while the HT resorted to post-edit the translation into the addition of the verb phrase: *had a long beard*.

4.2.5:PE by Lexical Addition

ST: نعمة و سلام من الله أبينا والرب يسوع الذي أحبنا قبل أن نحبه، .. (ص، 71)

GT: Grace and peace from God our Father and the Lord Jesus, who loved us before we loved Him.

PE: Grace and peace from God our Father and the Lord Jesus Christ, who loved us before we loved Him.

The HT added the word *Christ* although there's no equivalent of *Christ* (i.e., *المسيح*) in the Arabic ST and it can be said that this is an obligatory addition where *Christ* is a familiar idiomatic expression of the related meaning (*Jesus*). A case of addition is said to be obligatory when there is a need to give more information in the TT to protect the cohesion of a phrase, a sentence or a text [12: 206]. This addition is understandably justified to avoid stylistic repetition in the TT and to accomplish a natural and uninterrupted translation.

4.2.6: PE by Lexical Omission

ST: وقواطي معلبات اسماك وبقوليات. (ص، 216)

GT: Qawati, canned fish and beans

PE: .. and tins of fish and beans.

In this extract, the HT opted to omit the word *قواطي* from the TT for the reason that both Arabic words *معلبات* and *قواطي* - Iraqi slang - deliver the same meaning of the word *tins* in the TT. The semantic loss here is compensated for by avoiding a greater loss in idiomaticity that Google literal translation offered by resorting to transliteration (Qawati). Ivacovoni [13:101] mentioned the procedure of omission as being a dropping of a word from the ST when there isno equivalent word in the target culture.

4.2.7:PE of Semantically Related Words by Reduction

ST: ستشعر أم سليم..... باليأس و الإحباط. (ص، 16)

GT:Umm Salim will feel hopeless and frustrated

PE:Umm Salim will feel distressed.

There are cases where reduction is required to avoid redundancy in the TT [14: 228].The procedure of reduction is particularly applied on semantically related words if the source language tends be a redundant language as in the case of Arabic language. Semantically related words like *يأس* and *إحباط* don't generally pose a translation issue and can often be translated by reduction. In the above extract HT resorted to merge the two words into one word (*distressed*) to avoid synonymy or near-synonymy repetition. The same holds true for the following extract, where the HT merged *اعتذار* and *أسف* into *apologies*.

ST: أمطره بكلمات الاعتذار والأسف. (ص، 276)

GT: He showered him with apologies and regrets.

PE: He showered him with apologies.

4.2.8:PE of Common and Non-common Cultural Terms

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

ST: (258 ،ص) طناطل و سعلوات .. هذه كلها مخاوف يخلقها الناس
GT: Tantals and coughs.. these are all fears people create.
PE: Ogres and succubi .. are just manifestations of people's fears.

In the above extract, GT couldn't recognize the common cultural terms *طناطل* and *سعلوات* and rendered the first one literally by resorting to transliteration and mistranslated the other term with the Arabic word *سعال*.The HT resorted to post-edit the unnatural translation of GT with more natural-sounding translation: *Ogres* and *succubi*, which are very common supernatural demons in English that make an excellent equivalence to the Arabic cultural terms*طناطل* and *سعلوات* respectively. Also in the following extract, GT could not be aware of the cultural differences:

ST: (125،ص) قام بتسخير الجان والتوابع.. للعثور على طيف الأسم
GT: He harnessed the elves and minions to find the spectrum of the name.
PE: He harnessed the djinn and familiar spirits to find the spectrum of the name.

For this reason there was also a PE for the two cultural terms: *elves* and *minions*by suggesting the two terms instead: *djinn* and f*amiliar spirits* respectively. It can be noticed from these extracts that Arabic and English are not closely related cultures. Table (1) below shows the semantic and grammatical analysis of PE cases found in the previous examples:

Table (1):The Frequency Analysis of PE Data

No.	Analysis of PE	Frequency Data	Percentage
1	Grammatical PE	4	22%
2	Semantic PE	14	78 %
	Total	18	100 %

Based on the table results in addition to the previous data findings, it can be said that the usage of PE tool on GT is the way forward to enhance GT productivity. In fact, the present research's results support all the findings mentioned in section 2, suggesting that PE helps to enhance GT quality and efficiency.

5 Conclusions

It is evident that GT has an important role in assisting translators to perform their translation faster, nevertheless it does not always give the correct meaning of a word nor it is always grammatically correct. Thus, it would be of great help if it is integrated with human assistance by means of PE. For that reason, this research introduces a framework for post-editing that can be seen as a tool of improving MT through modification by providing the basic knowledge about PE process in order to produce a high-quality translation.

There search findings show that MT proved its adequacy in translating the Arabic literary text via GT, with some failures that required HT assistance to perform PE at the grammatical and semantic levels. In most cases, the translation of GT was literal and the

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

PE on the semantic level was more than on the grammatical level, especially the lexical choices and the metaphorical expressions which means that MT needs to reconsider the design of its system.

The results also highlight the fact that PE should be seen as a detecting and processing tool of language grammatical and semantic errors and how to eliminate them properly in order to produce a human-like translation with the same level of quality and accuracy.

6-Recommendations

Considering the findings obtained earlier, PE has the potential of improving MT efficiency of literary texts because it helps to better recognize the mistakes made by GT and can therefore be used as feedback for researchers and developers to improve the MT system. The findings can also be used for teaching translators the methods involved in PE machine-translated output.

CONFLICT OF INTERESTS

There are no conflicts of interest

References:

[1]Wagner, Emma (1985). "Post-editing Systran: A challenge for Commission translators." *In Terminologie and Traduction* Journal. Vol. 3, pp. 1-6.

[2]Somers, Harold. (2001). *Computers and Translation: A Translator's Guide*. Amsterdam: John Benjamin's Publishing Company.

[3] Vitek, Steve. (2000). "Reflections of a Human Translator on Machine Translation. Will MT Become the "Deus Ex Machina" Rendering Humans Obsolete in an Age When "Deus Est Machina?" *In Translation Journal*. Vol. 4, issue 3, pp. 1-8.

[4]Moorkens, Joss, Antonio Toral and Sheila Castilho (2018)."Translators' perceptions of literary post-editing using statistical and neural machine translation." *In Translation Spaces Journal.* Vol. 7, issue 2, pp. 1-22.

[5]Toral, Antonio, Martijn Wieling and Andy Way (2018). "Post-editing Effort of a Novel with Statistical and Neural Machine Translation." *In Frontiers in Digital Humanities Journal*. Vol. 5, pp. 1-11.

[6]Arenas, Ana & Toral, Antonio (2020)."The Impact of Post-editing and Machine Translation on Creativity and Reading Experience." In*Translation Spaces Journal*, Vol. 9, issue 2, pp.1-28.

[7] Ziganshina, L. Eugenevna, Ekaterina V. Yudina, Azat I. Gabdrakhmanov and Juliane Ried (2021)."Assessing Human Post-Editing Efforts to Compare the Performance of Three Machine Translation Engines for English to Russian Translation of Cochrane Plain Language Health Information: Results of a Randomised Comparison." *In MDPI Journal*. Vol 8, issue 9, pp. 1- 16.

[8]Spence, Green, Jeffrey Heer, and Christopher D. Manning (2013). "The Efficacy of Human Post-Editing for Language Translation." *In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI),* pp. 439–

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq

448.[9]Plitt, Mirko and Francois Masselot (2010). "A Productivity Test of Statistical Machine Translation Post-Editing in A Typical Localization Context." *In Prague Bulletin of Mathematical Linguistics*. Vol. 93, pp. 7–16

[10]Arenas, Ana (2013). "What do professional translators think about post-editing?" *In The Journal of Specialized Translation*, no. 19, pp. 75-95.

[11]Crystal, David (2003). *A Dictionary of Linguistics and Phonetics*. 5th, ed. Oxford: Blackwell Publishing

[12]Baker, Mona (1992). *In Other Words: A Coursebook on Translation*. Routledge: London and New York.

[13]Ivacovoni, Alessio (2009). *Translation by Omission*. Retrieved from: https://iacovoni.wordpress.com/2009/02/01/translation-by-omission/

[14]Nida, Eugene (1964). *Towards a Science of Translation*. Leiden: Brill.

Note:

* The translation of GT was done on 5, Feb, 2021. Retrieved from https://translate.google.com/?sl=en&tl=ar&op=translate

Online ISSN: 2312-8135 Print ISSN: 1992-0652
www.journalofbabylon.com/index.php/JUBH Email: humjournal@uobabylon.edu.iq