ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 15, No. 6, pp. 1864-1873, November 2024
DOI: https://doi.org/10.17507/jltr.1506.12

An Assessment of the Quality of Post-Edited Text From CAT Tools Compared to Conventional Human Translation: An Error Analysis Study

Hind S. Alsaif
Decision Support Centre-Royal Court, Riyadh, Saudi Arabia

Ebtisam S. Aluthman*
Department of Applied Linguistics, College of Languages, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

***Abstract*—This experimental study aims to evaluate the quality of post-edited texts, originally translated using computer-assisted translation (CAT) tools, in comparison with traditional human translation. This study investigates the quality of post-editing (PE) compared to traditional translation from scratch (TFS) in the context of Arabic–English translation, utilizing the Phrase CAT tool. The main hypothesis posits that PE yields a final product whose quality is similar or equivalent to that of TFS. The participants' scores and error frequencies were evaluated using the American Translators Association framework for standardized error marking, and terminology, word choice, mistranslation, addition/omission, spelling, punctuation, case, inconsistency, style, and grammar in both approaches were compared. Data from nine professional Saudi translators showed that PE generally outperformed TFS in terminology, spelling, punctuation, and case, whereas TFS exhibited strengths in consistency, style, grammar, and literal translation. Statistical analysis confirmed the similarity in overall error rates between PE and TFS. The difference in mean error numbers between TFS and PE was not statistically significant. Thus, the disparity in means likely resulted from random chance and might not indicate substantive differences between the two groups. These results imply that PE yields quality that is comparable or equivalent to that of TFS, proving the aforementioned hypothesis. The implications highlight the need for CAT tool training and PE skills among translators to meet the demands of evolving translation technologies. Furthermore, this study underscores the importance of integrating PE training into translation curricula and organizing workshops to improve CAT tool usage.**

***Index Terms*—post-editing, traditional human translation, computer-assisted translation (CAT) tools, ATA framework, error analysis**

I. Introduction

The development of computer-assisted translation (CAT) tools represents a significant advancement in the field of translation, primarily aimed at enhancing the speed and efficiency of the translation process while maintaining the integrity of the translated material, as noted by Esselink (2000). Nevertheless, these technological advances have not entirely met the high expectations set for them, encountering obstacles arising from technological and cognitive constraints (Garcia, 2012). Furthermore, the development of translation technology, dating back to 1952, was met with skepticism from pioneers in the field of translation, such as Bar-Hillel, who doubted the feasibility of such endeavors. In addition, many translators felt threatened by the possibility of these technologies replacing them (Mossop, 2017). Despite these concerns, it is imperative to recognize that these tools are not intended to replace human translators but to enhance their capabilities. This is evident in their primary objective of reinforcing human productivity. Once it was established that translation technology, regardless of its quality, would never attain the quality of human-edited text and would largely require human intervention, the concept of post-editing (PE) came into existence (Allen, 2001, 2003).

The emergence of PE in the domain of translation has been a crucial development, addressing the shortcomings inherent in translation technologies. PE tools are integrated with various CAT tools, providing a practical solution. O'Brien (2011) highlighted the procedural and qualitative aspects of PE. He described it as a process in which human translators enhance the output of machine translation (MT) in accordance with specific guidelines and quality measurements. This enhancement of technology-generated translations has been substantiated by a growing body of research (Jia et al., 2019) claiming that the application of these technologies can significantly enhance the translation process in terms of productivity. Such enhancement is achieved without compromising the quality of the output, and it concurrently reduces

* Corresponding Author. Email: esaluthman@pnu.edu.sa

© 2024 ACADEMY PUBLICATION

the overall effort required in the translation process. Despite these advancements, a degree of skepticism persists regarding the reliability of the quality of post-edited texts generated by CAT tools. In relation to the context of the present study, this reluctance to trust translation technologies has been observed among Arab translators. Previous studies (Alanazi, 2019; Al-Jarf, 2017; Alotaibi, 2017), as cited in Alkhatnai (2021, p. 73), highlight that, although CAT tools and other digital tools are being utilized more frequently worldwide, there is a sense of reluctance among researchers, particularly in the Arab world, with regard to adopting these modern technological tools.

This experimental study is designed to evaluate the quality of post-edited texts that have been translated through CAT tools, in comparison with traditional human translation, within the Arabic–English translation context. This comparison seeks to contribute valuable insights to the ongoing discourse in translation studies, particularly concerning the usability of integrating technology into the translation process and its implications for the quality of the output, with a specific focus on Arabic–English translation. The experiment in this study addresses the question of whether performing PE on CAT tool-generated text results in quality that is similar or equal to that of human translation. The data analysis for comparing CAT translation and traditional translation comprises three steps: assessing the total scores achieved by each participant in both approaches, examining the occurrence of errors across various criteria in both approaches, and tallying the overall count of errors and subsequently computing the means in both approaches.

II. Literature Review

A. *Translation Quality Assessment*

In the early stages of translation studies, Nida (1969) proposed four tests to determine the translation quality: the cloze test, reactions to alternatives, verbal reading of the text, and explanation of the content. However, House (1997, p. 136) criticized these tests for their lack of specificity and failure to reveal qualitative differences between translations. In the late 1980s, Newmark (1995) introduced an assessment approach that involved specific steps for analyzing a text. These steps included evaluating the extent of deviation from the original, determining if the translator misrepresented the author, assessing the deculturization of the text, and evaluating semantic and pragmatic accuracy (Newmark, 1988; as cited in El-Zeini, 1994). However, Newmark's approach faced criticism for its vagueness and potential subjectivity (El-Zeini, 1994).

Another significant development came from House (1997), who proposed a model for translation quality assessment that focused on identifying "mismatches" or "errors" (Munday, 2016, p. 147). This approach aligned closely with the "error typology" approach, which aims to reduce subjectivity by utilizing consistent classification models (O'Brien, 2012). The error typology approach for translation assessment is currently the most widely used method for evaluating a translation. Its significance lies in reducing the subjectivity of evaluation processes through methodical and consistent classification models (Secară, 2005). The field of translation quality assessment has witnessed various approaches over time. These developments have contributed to the ongoing exploration and refinement of translation quality evaluation methods.

In the present study, the researchers applied the American Translators Association (ATA) framework for standardized error marking for quality assessment to evaluate the translation and PE output of the participants. This framework refers to the guidelines provided by the ATA for marking errors in translated texts. The ATA is a professional association that sets standards and provides resources for translators and interpreters. This framework outlines a systematic approach to identifying and categorizing errors in translations, helping translators and reviewers ensure the accuracy and quality of their work. It typically includes categories such as grammatical errors, mistranslations, omissions, additions, inconsistencies, and stylistic issues. Each category may be further divided into subcategories to provide more detailed feedback.

B. *The Quality of PE and Traditional Human Translation*

PE has been defined and conceptualized by various scholars over time, each adding depth and perspective to its understanding. Veale and Way (1997; as cited in Allen, 2003) offered an initial definition of PE, describing it as "the term used for the correction of machine translation output by human linguists/editors" (p. 297). This definition emphasizes the role of human linguists or editors in refining the output produced by MT systems. Allen (2001) provided a more comprehensive description. He characterized PE as an integrated process within MT, involving professional human translators. These translators engage in correcting machine pre-translated texts and fuzzy matches, which are translations generated from the translation memories (TM) that bridge an original language (source language [SL]) to a translated language (target language [TL]). This process aims to produce translations of higher quality in less time, highlighting the efficiency gains achieved through human–machine collaboration. Further adding to the discourse.

Fiederer and O'Brien (2009) explored whether post-edited MT output is inherently inferior to human translation. Their study involved a comparison of 30 sentences evaluated for clarity, accuracy, and style by 11 qualified raters. Out of these, three versions of each sentence were assessed: one translated version and two post-edited versions. The findings revealed that the PE of machine-translated content yielded higher clarity and accuracy, whereas human translations were superior in terms of style. Guerberof (2009) conducted research on the speed and quality of performing PE on TM or MT outputs. The study involved nine professional translators with experience ranging from 1 year to over 10 years. For the experiment, 791 words were used, divided into new segments needing translation, segments from Trados translation memories (fuzzy

© 2024 ACADEMY PUBLICATION

matches), and segments from Language Weaver MT. A web-based PE tool was employed to time each task, and the Localization Industry Standards Association (LISA) standard was used for error measurement and classification. The study found that post-edited TM outputs had the highest error count, namely 91% higher than the MT segments and 26% more than the human-translated segments; this was attributed to the less apparent nature of errors in naturally flowing language compared to MT, in which errors are more conspicuous.

Garcia (2010) conducted a study to evaluate the benefits of MT in segments that received a “no match” result in the Google Translation Toolkit, particularly focusing on the quality of translations and their utility for translators. Instead of professional translators, this study involved trainee translators who were directed to use the Google Translation Toolkit to translate passages from English to Chinese, either from scratch or by filling in the blanks using Google Translate. The translation quality was assessed by experienced raters using the criteria set by the Australian National Accreditation Authority for Translators and Interpreters (NAATI). The comparative analysis revealed that, in 33 out of 56 cases, the passages translated via Google Translator Toolkit and then edited by humans were rated more favorably. These results suggest that performing PE on MT output could be a more effective and advantageous method in the field of translation.

Elming et al. (2014) conducted a case study focusing on the comparative quality of post-edited MT segments and segments translated by humans. The results indicated that the human-translated segments contained a higher number of errors. Notably, stylistic errors were three times higher in the human-translated segments than in the PE segments. This finding contrasts with the results of Fiederer and O’Brien (2009) at Dublin City University, which suggested that human translation achieved better style than post-edited MT. Yang et al. (2020) carried out a study comparing human translation and PE approaches. Unlike other studies conducted in professional settings, this experiment was carried out in an educational context, aiming to assess the effectiveness of PE. The researchers employed the multidimensional quality metrics (MQM) framework to objectively analyze translation errors in both human translation and MT, focusing on accuracy and fluency. They found that the participants scored higher in PE tasks compared to human translation, with a tendency for more accuracy errors than fluency errors in post-edited outputs.

Samman’s (2022) study, one of the few conducted in Saudi Arabia, aimed to evaluate the efficacy of MT PE training within a female undergraduate translation program. The study, adopting a mixed-method design and based on the Kirkpatrick model of learning evaluation, compared MT PE with human translation in terms of quality and other factors. The translations, which were from English to Arabic, were evaluated using the DipTrans Examiners’ Mark Sheet. The analysis revealed that MT PE was effective in reducing deletion and technical errors but increased errors in accuracy, register, grammar, comprehension, mistranslation, word order, and overall text organization. Despite showing potential for improvement, the study highlighted challenges in achieving accurate and fluent translations. PE was found to be beneficial in addressing errors in capitalization, punctuation, numbering, and agreement. The error count analysis showed only a slight difference in scores between the human translation and MT PE groups, suggesting that PE can achieve results comparable to human translation but does not significantly outperform it. The study underscored the potential benefits of PE for inexperienced translators and non-professionals, emphasizing the need for improved Arabic MT output and specific skill development for effective MT PE.

The present study, motivated by the significant progress in CAT tools and the evident scarcity of scholarly research on their application in the context of Arabic translation, is designed to fill this gap in the existing literature. The main hypothesis of this study posits that the PE of texts translated using CAT tools yields a final product whose quality is similar or equivalent to that of traditional human translation. The key metric for assessing quality in this context is the frequency of errors in the translated texts. This research seeks to ascertain whether performing PE on CAT tool-generated translations can produce a final product with fewer errors and, hence, of higher quality compared to translations completed entirely by human translators without the aid of such technology. This comparison illustrates the effectiveness of CAT tools in enhancing translation quality and the overall accuracy of the final translated product.

III. METHODOLOGY

A. Sample, CAT Tools, and Text

The quality of PE performed on texts originally translated using CAT tools, in comparison with traditional human translation, was evaluated based on the performance of professional translators, with data collected from nine Saudi translators holding bachelor’s or master’s degrees in translation and possessing 5–12 years of experience. The translators were familiar with both traditional translation and PE and had used various CAT tools, such as Trados, MemoQ, Matecat, and Memsource. They were asked to translate an Arabic educational text into English using both approaches. The researchers utilized the Phrase CAT tool to produce the text for PE and to analyze the quality of the final output. A screenshot of Phrase is illustrated in Figure 1 below.

© 2024 ACADEMY PUBLICATION

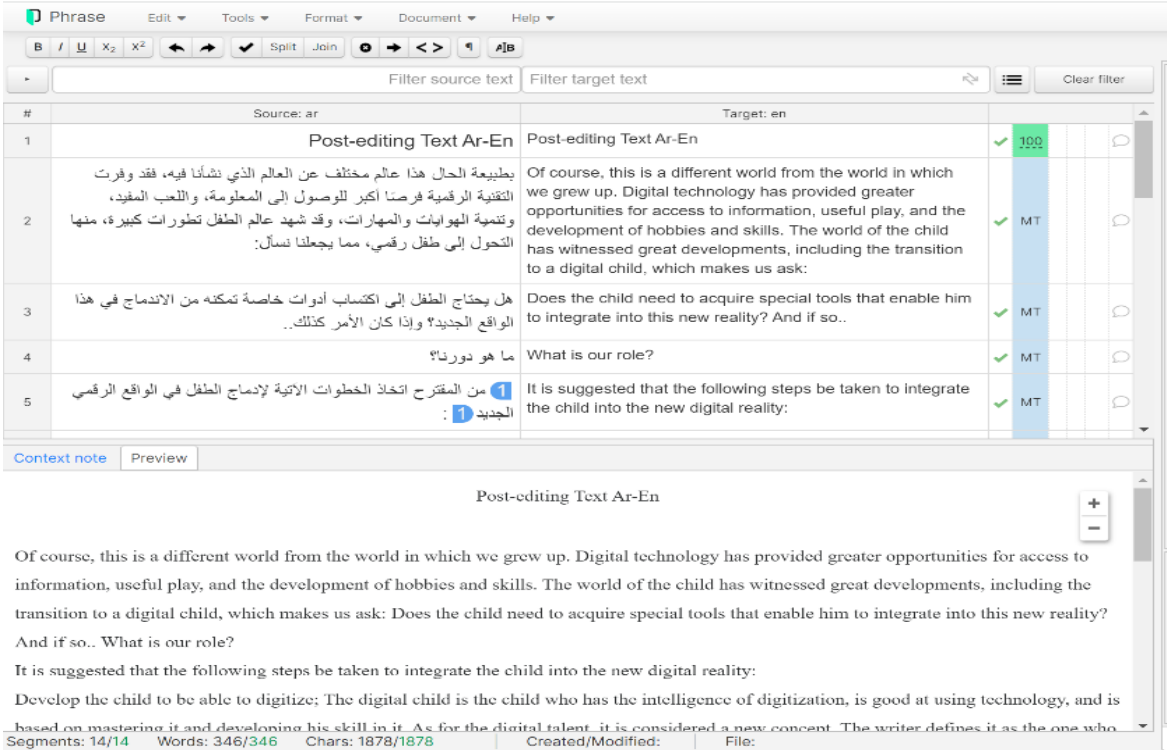

Figure 1. Screenshot of the Phrase Platform

The participants were requested to participate in a language lab, where they used devices to record the process of the approaches on Translog-II in both translation from scratch (TFS) and PE. The selected text was chosen for its suitability in evaluating the translators’ performance in handling non-literary, unambiguous educational content.

B. *Evaluating Rater Agreement Using Cohen’s Kappa*

Cohen’s Kappa (1960), a statistical measure of inter-rater reliability, was employed to evaluate the consistency and agreement between assessments made by the rater and an inter-rater. This approach ensured that the error assessments were dependable and coherent across the raters. The kappa value was calculated using the following formula: $k = (Po – Pe) / (1 – Pe)$, where Po stands for the observed agreement and Pe represents the expected agreement by chance. The Kappa values were interpreted using a scale to determine their agreement levels. It generates a value between -1 and 1, with different ranges indicating varying levels of agreement:

- 0: by-chance agreement
- 0.1–0.2: slight agreement
- 0.21–0.4: fair agreement
- 0.41–0.6: moderate agreement
- 0.61–0.8: substantial agreement
- 1: perfect agreement

C. *ATA’s Framework for Standardized Error Marking*

In this study, the ATA framework for standardized error marking (2002) was utilized to assess the quality of the final texts produced by the participants in TFS and PE. This framework is one of the most commonly adopted error typologies (Doyle, 2003; Koby, 2015; Phelan, 2017). ATA certifications are among the most respected certifications in the world, and this framework is its grading system for certification exams to assess the language skills of potential translators. To identify translation errors and assign points to each error, ATA evaluators refer to a list of error categories. The evaluation of translation quality is conducted by applying the ATA framework for standardized error marking. In this study, the total scores obtained by each participant, errors across various error criteria, and the score means in both approaches were calculated. However, some criteria, such as the false cognate, accent, and other diacritical markers, were not applicable to this study’s language pair (Arabic–English). Furthermore, the incomplete passage criterion was not included in the assessment since the participants were instructed to complete the passages and were not restricted by a time limit.

IV. Results

This section presents the quality analysis investigating the hypothesis that performing PE on a CAT tool-generated text yields quality that is similar or equal to that of human translation. The ATA framework for standardized error marking (Doyle, 2003) was used to assess the quality. Subsequently, the participants’ final scores in both approaches were

© 2024 ACADEMY PUBLICATION

determined by counting the number of errors, taking into account the severity level of the errors. Considering the severity level of errors is essential, as two translations may have an equal number of errors, but one may contain more significant errors, resulting in lower quality. This is followed by a detailed examination and description of the error types and their frequencies in each approach. In addition, this section offers a comprehensive analysis of the total errors found in the TFS and PE approaches.

The first part of the quality analysis compared each participant's final scores in the two approaches. The score calculation involved deducting one point for minor errors, two for major errors, and none for neutral errors. The total points determined the final scores. Figure 2 illustrates those six participants (P2, P3, P4, P5, P6, and P9) performed better in PE, only one participant (P8) achieved equal scoring (P8), and two performed better in TFS (P1 and P7). The TFS scores ranged from 54% (P3) to 86% (P9), while the PE scores ranged from 76% (P4 and P5) to 86% (P9).

Figure 2. Participants' Scores in TFS and PE

An inter-rater reliability analysis was conducted using Cohen's kappa to assess the level of agreement between two raters in the evaluation of error criteria. It was conducted with two participants: P1 and P3. For P1, the calculated kappa value for the observed agreement (Po) was found to be 17 divided by 22, which means that the two raters agreed on 17 error criteria out of 22. This yielded a Kappa value of 0.77, indicating a substantial level of agreement beyond what would arise from random chance. This suggests that the two raters generally concurred in their assessments of error categories.

To further evaluate the agreement, the expected agreement by chance (Pe) was calculated and found to be 0.194. The Cohen's kappa statistic was then calculated using the formula (Po - Pe) / (1 - Pe), resulting in a kappa value of approximately 0.715. This value signifies a substantial level of agreement between the raters, suggesting that their evaluations of error categories were consistent and likely not due to random chance.

For P3, the calculated kappa value for the observed agreement (Po) was determined to be 0.95, which was calculated by dividing the number of agreed-upon criteria (21) by the total number of criteria (22). The expected agreement by chance (Pe) was also calculated and yielded a value of 0.83. Subsequently, Cohen's Kappa statistic was calculated using the formula (Po - Pe) / (1 - Pe), resulting in a kappa value of approximately 0.897. This kappa value signifies a high level of agreement between the raters, affirming the consistency and reliability of the assessments for P3 as well.

Figure 3 illustrates the frequency of errors in each criterion and reveals that the TFS approach was most susceptible to errors related to terminology and word choice, with a total of 85 errors, representing 21.8% of all TFS errors. In contrast, the PE approach had 38 errors of this type, accounting for 15% of all PE errors. This suggests that the number of terminology errors made by the participants was 44% higher in TFS compared to PE. The results in Figure 3 highlight the prevalence of linguistic errors in both TFS and PE. Notably, TFS had a higher frequency of errors in categories such as mistranslation, addition/omission, and register. Specifically, the data show that TFS had 12 more mistranslation errors than PE, with 36 total TFS errors compared to 24 PE errors. Furthermore, TFS had more errors related to addition and omission, with 28 total errors compared to 19 in PE. Despite these discrepancies, the figure also demonstrates that PE was effective in improving the translation quality for these categories, with improvements of 32.1% and 33.3% for addition/omission and mistranslation errors, respectively.

In terms of inconsistency, style, grammar, and literal translation errors, the data highlights that TFS outperformed PE. The numbers show that PE had a higher frequency of errors in these four areas, with 43, 38, 33, and 16 errors, respectively, while TFS had significantly lower error rates, with only 6, 27, 19, and 10 errors, respectively. In addition, the data highlights a significant difference between the two approaches regarding inconsistency errors, with PE having 43 errors compared to 6 errors in TFS. These errors can lead to confusion and misinterpretation, resulting in negative impacts on

© 2024 ACADEMY PUBLICATION

the quality of the translated text. Hence, addressing these errors is crucial for preventing inaccurate translations and improving the quality of translated texts in both approaches. Furthermore, there was a substantial difference in spelling, capitalization, and punctuation errors between the two approaches. In particular, spelling errors occurred 77 times in TFS, the second-highest total after terminology, but only 5 times in PE. Similarly, case errors were more prevalent in TFS; it had 37 errors, compared to only 14 errors in PE, representing a 62.16% reduction. TFS encountered more difficulties in punctuation as well, with 30 errors compared to only 25 errors in PE. These findings suggest that TFS and PE have different strengths and weaknesses, and that both approaches require improvements to enhance the overall translation quality.

The results suggest that PE has a significant positive impact on non-linguistic errors, such as spelling, capitalization, and punctuation, in addition to improving translation accuracy by aiding the translator in avoiding omission/addition and mistranslation. However, PE does not seem to have a noticeable effect on issues such as literal translation, grammar, style, and inconsistency.

© 2024 ACADEMY PUBLICATION

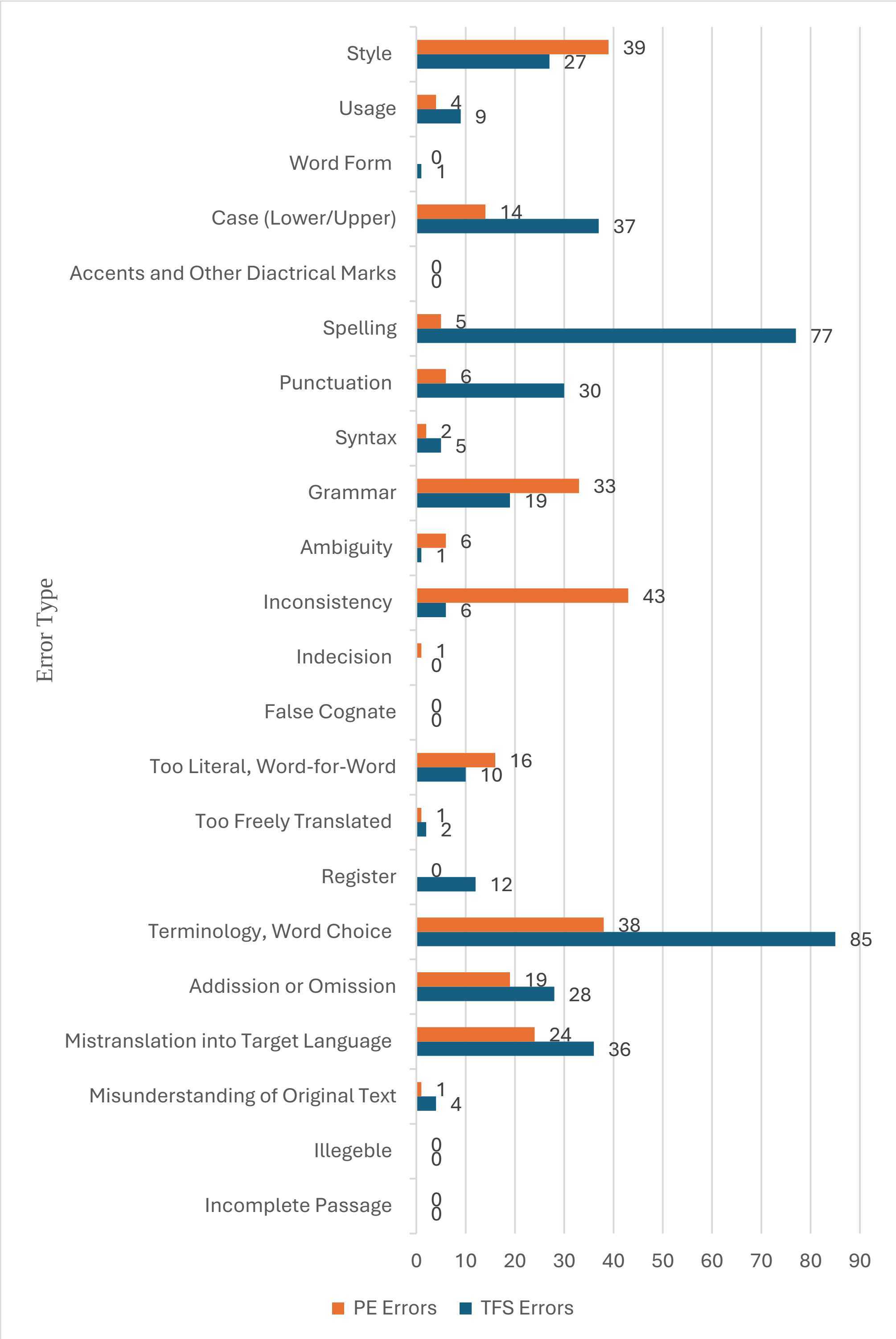

Figure 3. Participants' Error Frequencies in TFS and PE

The final step in the quality analysis was calculating the total number of errors and then finding their means for the TFS and PE approaches, as presented in Figure 4. To obtain this total, the researcher counted the frequency of errors according to the ATA error criteria for each approach and subsequently determined the means of the errors for each approach. The mean number of errors was 17.9 for TFS and 11.5 for PE. A *p*-value of 0.289961119 was obtained, suggesting that the difference in mean error numbers between TFS and PE was not statistically significant at the conventional significance level of $p < 0.05$. This suggests that the disparity in means likely resulted from random chance and might not be indicative of substantive differences between the two groups. In essence, these results imply that PE yielded quality comparable or equivalent to that of TFS, proving the hypothesis that performing PE on CAT tool-generated text results in quality that is similar or equal to that of human translation.

Figure 4. Means of Participants Errors in TFS and PE

V. Discussion and Conclusions

In this study, the quality of the texts produced by the participants in the TFS and PE approaches was evaluated by the researcher using the ATA framework for standardized error marking. To compare the errors found in both approaches, the researcher assessed the quality of these approaches in terms of terminology, word choice, mistranslation, addition/omission, spelling, punctuation, case, inconsistency, style, and grammar. Overall, the number of errors in TFS was higher than in PE: 389 errors and 252 errors, respectively. The results showed that the quality of terminology and word choice was better in PE than in TFS. In TFS, 85 errors (21.8% of total errors) were found in this criterion, whereas in PE, only 38 errors (15% of total errors) were identified. This suggests that translators may struggle with choosing the right terminology, whereas the CAT tool used in PE automatically provides appropriate equivalents most of the time.

Furthermore, regarding errors, this study found that mistranslation and addition/omission errors were less frequent in PE compared to TFS. Mistranslation accounted for 24 errors (9.5% of total errors) in PE and 36 errors (9.25% of total errors) in TFS. This could be caused by different factors, such as misunderstanding the source text, expressing meaning in the wrong manner, or misreading the source text. These are factors mostly related to and caused by human cognition and linguistic expression. Addition/omission errors comprised 19 errors (7.54% of total errors) in PE and 28 errors (7.19% of total errors) in TFS. While exceptions may apply, it is generally preferred to avoid inserting material for clarification in translation and to avoid omitting necessary information. In terms of spelling mistakes, TFS had significantly more errors than PE. In TFS, 77 errors (19.79% of total errors) were identified, while in PE, only 5 errors (1.98% of total errors) were found. This suggests that translators may rely on automatic correction when translating, but since the software used in this study did not correct or indicate spelling mistakes, the participants were often unaware of their errors. In contrast, in PE, the translation was provided by a CAT tool, which reduced the occurrence of spelling mistakes. Compared to PE, TFS exhibited a higher frequency of punctuation errors, with 30 errors (7.7% of total errors); in contrast, PE had only 6 errors (2.38% of total errors) in this criterion. In addition, PE showed a better performance in case errors, with only 14 errors (5.6% of total errors), whereas TFS had 37 errors (9.5% of total errors) in this category. These discrepancies may be attributed to the inherent differences between Arabic and English regarding cases and the use of punctuation. The results indicate that PE and TFS exhibit similar quality, with room for further improvement in PE. This observation aligns with the high-quality MT offered by the Phrase CAT tool. Notably, PE showcased notable enhancements in errors related to spelling, case, punctuation, and translation fidelity. It also helped translators avoid omission/addition and mistranslation. However, errors related to literal translation, grammar, style, and inconsistency did not show significant improvement in PE. This could be due to the natural flow of errors in a high-quality CAT tool translation, which may prevent translators from noticing these errors, as previous research has suggested (Guerberof, 2008, 2009; Yamada, 2019).

The implications of the results are multifaceted and provide valuable insights into the efficacy of PE compared to TFS, as well as the role of CAT tools in translation processes. The findings suggest that PE yields better quality in terms of terminology and word choice compared to TFS. This implies that the CAT tools used in PE are effective in automatically providing appropriate equivalents, potentially enhancing translation accuracy and consistency. Furthermore, PE demonstrates a reduction in mistranslation and addition/omission errors compared to TFS. This indicates that the use of CAT tools in PE may help mitigate errors caused by human cognitive and linguistic factors, such as misunderstanding or

© 2024 ACADEMY PUBLICATION

misreading the source text. TFS exhibits a significantly higher frequency of spelling mistakes compared to PE. This highlights the importance of CAT tools in reducing spelling errors by providing automatic correction, thereby enhancing the overall quality of translated texts. Moreover, PE demonstrates better performance in punctuation and case errors compared to TFS. The discrepancies observed could be attributed to inherent differences between Arabic and English, as well as the use of CAT tools, which may facilitate more accurate punctuation and case usage. Overall, the results suggest that PE with CAT tools offer advantages over traditional TFS in terms of accuracy, consistency, and efficiency. However, they also underscore the need for translators to be adequately trained in utilizing CAT tools and PE techniques effectively. In addition, the findings emphasize the importance of periodically reviewing and adapting translation training curricula to incorporate technological advancements and address specific linguistic challenges, such as those related to Arabic–English translation.

There is a growing demand for translation technologies due to increased translation volumes and the need to increase productivity. This demand requires translators to be familiar with translation technologies and how to interact with them through PE (Bowker, 2015). This is particularly true at present, with advanced translation technologies being developed while the demand for them rises. Furthermore, limited experience with CAT tools will not allow translation students to gain a realistic understanding of the functioning of these tools (Bowker, 2015). This can also be applied to PE training, which is an intrinsic part of using translation technology in general; as mentioned before, these technologies cannot replace human translators and will always require human intervention to match the quality of human-edited text. Nevertheless, nearly all translators—especially local translators—need to receive training in PE. In a study on PE, Koby suggested that "the translator must be trained in post-editing" (Krings & Koby, 2001, p. 12). Furthermore, McElhaney and Vasconcellos (1988) argued that, since translation and PE are varying processes, translators are most suitable for undertaking this task, as they can identify linguistic errors and have rich knowledge about cross-language transfer. Therefore, in line with Al-Rumaih's (2021) implication, CAT tool course plans should be reviewed yearly to ensure they are up to date with technological advancements. Furthermore, PE training should incorporate CAT tool courses. Workshops for students should sufficiently cover translation technology and ensure the practical use of these tools, in addition to providing PE training.

References

[1] Alanazi, M. (2019). *The use of computer-assisted translation tools for Arabic translation: User evaluation, issues, and improvements* [Doctoral dissertation, Kent State University].
[2] Al-Jarf, R. (2017). Technology integration in translator training in Saudi Arabia. *International Journal of Research in Engineering and Social Sciences*, *7*(3), 1–7.
[3] Alkhatnai, M. (2021). Perceptions, skills, and technologies for the revitalization of translation industry in the post COVID-19 era: An empirical evidence from Saudi Arabia. *Journal of Foreign Language Teaching and Translation Studies*, *6*(3), 71–96. https://doi.org/10.22034/EFL.2021.306093.1122
[4] Alotaibi, H. M. (2017). Arabic-English parallel corpus: A new resource for translation training and language teaching. *Arab World English Journal*, *8*(3). https://dx.doi.org/10.24093/awej/vol8no3.21
[5] Alotaibi, H. M. (2020). Computer-assisted translation tools: An evaluation of their usability among Arab translators. *Applied Sciences*, *10*(18). https://doi.org/10.3390/app10186295
[6] Allen, J. (2001). Postediting: An integrated part of a translation software program. *Language International Magazine*, *13*(2), 26–29.
[7] Allen, J. (2003). Post-editing. In H. Somers (Ed.), *Computers and translation: A translator's guide* (pp. 35). John Benjamins.
[8] Al-Rumaih, L. A. (2021). The integration of computer-aided translation tools in translator-training programs in Saudi universities: Toward a more visible state. *Arab World English Journal for Translation & Literary Studies*, *5*(1). http://dx.doi.org/10.2139/ssrn.3802984
[9] Automatic Language Processing Advisory Committee. (1966). *Machines: Computers in translation and linguistics* (Publication 1416). Division of Behavioral Sciences, National Academy of Sciences, National Research Council.
[10] Bowker, L. (2015). Computer-aided translation: Translator training. In S. Chan (Ed.), *Routledge encyclopedia of translation technology* (pp. 88–104). Routledge.
[11] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.
[12] Do Carmo, F., & Moorkens, J. (2020). Differentiating editing, post-editing and revision. In *Translation revision and post-editing* (pp. 35–49). Routledge.
[13] Doyle, M. S. (2003). Translation pedagogy and assessment: Adopting ATA's framework for standard error marking. *The ATA Chronicle*, *32*(11), 21–28.
[14] Elming, J., Balling, L. W., & Carl, M. (2014). Investigating User Behaviour in Post-editing and Translation using the CASMACAT Workbench. In S. O'Brien, L. W. Balling, M. Carl, M. Simard, & L. Specia (Eds.), Post-editing of Machine Translation: Processes and Applications (pp. 147-169). Cambridge Scholars Publishing. Newcastle upon Tyne.
[15] El-Zeini, N. T. (1994). *Criteria for the evaluation of translation: A pragma-stylistic approach*. Cairo University.
[16] Esselink, B. (2000). *A practical guide to localization*. John Benjamins.
[17] Fiederer, R., & O'Brien, S. (2009). Quality and machine translation: A realistic objective. *The Journal of Specialised Translation*, *11*(11), 52–74.
[18] Garcia, I. (2010). Is machine translation ready yet? *Target – International Journal of Translation Studies*, *22*(1), 7–21. https://doi.org/10.1075/TARGET.22.1.02GAR
[19] Garcia, I. (2012). A brief history of postediting and of research on postediting. *Revista Anglo Saxonica*, 291–310.

© 2024 ACADEMY PUBLICATION

[20] Guerberof, A. (2008). Productivity and quality in the post-editing of outputs from translation memories and machine translation. *Localisation Focus – The International Journal of Localisation*, *7*(1), 11–21.
[21] Guerberof, A. (2009). *Productivity and quality in MT post-editing* [conference paper]. Universitat Rovira I Virgilli, Spain. Retrieved July 17, 2021, from http://www.mt-archive.info/MTS-2009-Guerberof.pdf
[22] House, J. (1997). *Translation quality assessment: A model revisited*. Gunter Narr.
[23] Jia, Y., Carl, M., & Wang, X. (2019). How does the post-editing of neural machine translation compare with from-scratch translation? A product and process study. *The Journal of Specialised Translation*, *31*(1), 60–86.
[24] Krings, H. P., & Koby, G. S. (2001). Repairing texts: empirical investigations of machine translation post? Editing processes. Kent, Ohio: Kent State University Press. https://doi.org/10.7202/008026AR
[25] McElhaney, T., & Vasconcellos, M. (1988). The translator and the postediting experience. *Technology as Translation Strategy*, *2*, 140–148. https://doi.org/10.1075/ata.ii.28mce
[26] Morado Vázquez, L., Rodriguez Vazquez, S., & Bouillon, P. (2013). Comparing forum data post-editing performance using translation memory and machine translation output: A pilot study. In *Proceedings of Machine Translation Summit XIV* (pp. 249–256).
[27] Moreno, M. D. (2020). Translation quality gained through the implementation of the iso en 17100: 2015 and the usage of the blockchain: The case of sworn translation in Spain. *Babel*, *66*(2), 226–253.
[28] Mossop, B. (2017). Conflict over technology in the translation workplace. *Multi-Languages Annual Conference 2017*.
[29] Munday, J. (2016). *Introducing translation studies: Theories and applications*. Routledge.
[30] Newmark, P. (1995). *A textbook of translation*. Longman.
[31] Nida, E. A. (1969). *Toward a science of translating: With special reference to principles and procedures involved in Bible translating*. Brill Archive.
[32] Nida, E. A. (2001). *Language and culture-contexts in translating*. Shanghai Foreign Language Education Press.
[33] Nord, C. (2001). *Translating as a purposeful activity-functionalist approaches explained*. Shanghai Foreign Language Education Press.
[34] O’Brien, S. (2007). An empirical investigation of temporal and technical post-editing effort. *Translation and Interpreting Studies – The Journal of the American Translation and Interpreting Studies Association*, *2*(1), 83–136.
[35] O’Brien, S. (2011). Towards predicting post-editing productivity. *Machine Translation*, *25*(3), 197–215. https://doi.org/10.1007/s10590-011-9096-7
[36] O’Brien, S. (2012). Towards a dynamic quality evaluation model for translation. *The Journal of Specialised Translation*, *17*(1), 55–77. Retrieved September 1, 2022, from https://aclanthology.org/www.mt-archive.info/10/JOST-2012-OBrien.pdf.
[37] Phelan, M. (2017). Analytical assessment of legal translation: A case study using the American Translators Association framework. *The Journal of Specialised Translation*, *27*, 89–210.
[38] Samman, H. M. (2022). *Evaluating machine translation post-editing training in undergraduate translation programs—An exploratory study in Saudi Arabia* [Doctoral dissertation, University of Southampton].
[39] Secară, A. (2005). Translation evaluation: A state of the art survey. In *Proceedings of the eCoLoRe/MeLLANGE Workshop* (pp. 39–44). St. Jerome Publishing.
[40] Somers, H. (Ed.). (2003). *Computers and translation: A translator’s guide*. John Benjamins.
[41] Yang, Y., Wang, X., and Yuan, Q. (2020). Measuring the usability of machine translation in the classroom context. *Translation and Interpreting Studies*, *16*(1), 101–123. https://doi.org/10.1075/tis.18047.yan
[42] Yamada, M. (2019). The impact of Google neural machine translation on post-editing by student translators. *The Journal of Specialised Translation*, *31*, 87-106.

Hind S. Alsaif earned a Master of Translation from Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia in 2022.

She has extensive professional experience as a translator, with a particular focus on translation technologies, Computer-Assisted Translation (CAT), and specialized translation fields. Prior to her current position, she held various roles that expanded her expertise in both the practical and academic aspects of translation. She currently works at the Decision Support Centre-Royal Court, located in Riyadh, Saudi Arabia.

Alsaif is actively involved in professional societies related to her field. Her contributions to translation studies and related technologies underscore her commitment to advancing this field. Her work has been recognized in several professional circles, although specific awards and committee roles are not mentioned.

Ebtisam S. Aluthman earned a Ph.D. in Applied Linguistics from the University of Manchester, United Kingdom. Additionally, she obtained her Master’s in Applied Linguistics and a Bachelor of Arts in English Language from Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

She is currently working as the Secretary-General of the Princess Nourah bint Abdulrahman Award for Women’s Excellence and Advisor to the President of Princess Nourah bint Abdulrahman University. Dr. Aluthman is an esteemed member of several academic societies and is evidently interested in the intersection of computing and linguistics. This may involve developing or studying software and tools that support language learning and processing, such as natural language processing (NLP) applications.

© 2024 ACADEMY PUBLICATION