Inverse Translation Quality: A Comparative Analysis between Human Translation and Post-editing

Muhamad Alif Haji Sismat
Faculty of Arabic Language
Universiti Islam Sultan Sharif Ali
alif.sismat@unissa.edu.bn

Abstract

This paper aims to analyse the quality of inverse translation and to see whether or not trainee translators, such as undergraduate language students, can produce translations between foreign languages, and whether or not post-editing machine translation and translation memories, have any effect on the Malay students' performance. Through error analysis approach, this paper also aims to reveal the contributing factors to the mistakes the students did in their translations and uncover the nature of Google Translate by identifying the recurring types of errors in the MT outputs. Results revealed that the translation technologies, particularly in the post-editing modified translation memories and machine translation tasks, helped the students improved the quality of their translations, suggesting that non-native speakers can be highly skilled professional translators with years of experience and proper training. Based on the error analysis, syntactic and lexical errors seem to be problematic in Google Translate in both Arabic-English and English-Arabic translations, implying that proper guidelines are crucial in post-editing so that post-editors can be aware of the potential recurrent errors and not overlook them. Also, the study identified that linguistic interference might have significantly influenced the students' performance as the three languages differ from one another in many aspects.

Keywords: inverse translation, translation quality, post-editing, machine translation, computer-assisted translation tool.

1. INTRODUCTION

Translation directionality is often neglected and frowned upon as traditionally translators are preferred to translate into their mother tongue. However, the translation industry has come a long way and evolved due to increasing demands. Therefore, more translators are required to meet the growing demands of clients. In some parts of the world, translation service companies lack native speakers. Consequently, translation into foreign languages, often known as inverse translation, has become a common practice (IAPTI, 2015). Even translation training has also changed as being native speakers is no longer essential as long as the translators are near-native speakers, or the target language is of habitual use (Newmark, 1988).

Therefore, there is a need to evaluate the quality of inverse translations and determine whether or not non-native speakers can be trained to meet the professional standards of quality. There is no doubt that translation technologies, such as machine translation (MT) and translation memory (TM), have significantly advanced over the years, but whether or not these technologies can help non-native translators produce good quality translation, it still remains debatable because many factors can contribute to the findings of the existing studies (Guerberof, 2012; Koponen, Aziz, Ramos & Specia, 2012; Koponen & Salmi, 2015) such as language pairs, text genres, sentence length, and MT users.

As a result, this study also aims to reveal the patterns of errors using the error-based approach as this type of information would essentially help Malay learners of Arabic and English improve their translation skills and give developers and researchers some insight into the types of errors they should 'fix' because correcting recurrent errors particularly can be exhausting and time-consuming. Moreover, there is scarce research in inverse translation, particularly in relation to the role translation technologies in affecting the non-native speakers' translations as reported by Sánchez-Gijón and Torres-Hostench (2014, p.7) as most studies tend to generalise their findings, ignoring the fact that native language (L1) translations differ from inverse translations.

2. METHODS

2.1 Participants

The present study focuses on the quality of inverse translation as different language pairs, and translation direction may provide different results (Toral and Sanchez-Cartagena, 2017; Castilho et al., 2017). Therefore, the research was conducted with Malay students, who are non-native speakers of Arabic and English, at Sultan Sharif Ali Islamic University. The undergraduate students majored in Arabic linguistics and literature, and most of them learned

both Arabic and English at a young age. Therefore, their knowledge background and proficiency in both Arabic and English are sufficient to make them eligible for carrying out translation tasks in this study.

2.2 Text materials and tasks

Initially, there is a list of texts prepared for the research project, but within the given six weeks, the students only managed to translate 11 source texts: 6 English and 5 Arabic texts, ranging from 116-311 words. In comparing the results of human translation and post-editing, the students were required to complete three different tasks: translation from scratch (human translation or HT), post-editing machine translation (PEMT), and post-editing modified translation memory and machine translation (PETM+MT). The three tasks were designed to see whether or not translation technologies used for this study affect the students' translations by identifying the pros and cons.

2.3 Translation technologies

The present study opts for Google Translate (GT) and Microsoft's Bing Translator (BT) as they are considered among the best free MT engines available on the internet. These MT systems can also be integrated into computer-assisted translation (CAT) tools through plug-ins, which enhance the translation workflow. It is also worth noting that the study only analyses the post-edited MT outputs generated during the project and only focuses on Arabic-English language pair.

As this study also analyses post-edited translation memories, it opts for MemoQ 2014 as it offers many features, such as MT integration and translation memory (TM), which are crucial for the study. It is also worth noting that translation memories typically contain previously translated texts so that they can be 'recycled' when translating similar texts. As the nature of TM is of good quality, the study deliberately modified the TM outputs by including errors to increase the task difficulty and challenge the students' critical thinking skills.

2.4 Translation quality assessment and error analysis

Another feature, such as linguistic quality assurance (LQA), helps users annotate errors and generate an automated translation quality assessment (TQA) report by using the TQA models available in the program or creating a new model from scratch. Therefore, this study uses the MeLLANGE error typology (as shown in Figure 1) as the TQA model for annotating errors and evaluating the quality of the inverse translations.

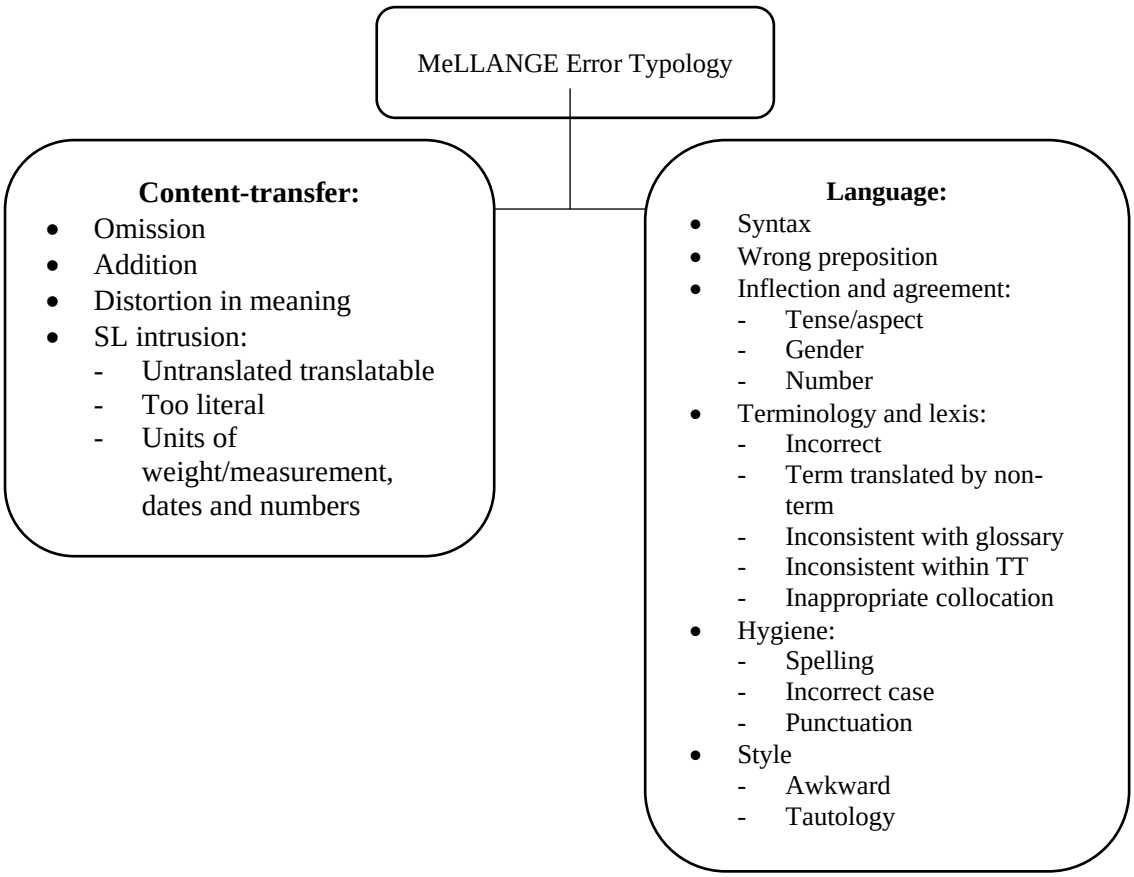

Figure 1: MeLLANGE Error Typology

In assessing the quality, the present study penalises major and minor errors with 5 points and 1 point respectively. MemoQ's LQA then automatically evaluates the normalised score of the students' translations in all three tasks. Two pass-mark thresholds are used to determine the quality of the translations: the 0.90 or 90% threshold set by the European Commission (Paspartu, 2016) and the 0.85 or 85% threshold Temizöz (2013) opted for assessing students' translation quality. The main reason for choosing the two pass-mark thresholds is to determine whether or not non-native speakers can produce translations that meet the standards of trainee and professional translators.

3. RESULTS

3.1 Translation quality assessment

As the focus of the study is analysing and comparing the quality of the inverse translation in all three different tasks, the students were required to work with both translation directions: English-Arabic (EN-AR) and Arabic-English (AR-EN). The main reason is to see whether or not the directionality and type of task affect the translation quality.

Table 1 presents the normalised score of the students' translations. The scores in blue font represent the pass marks according to the European Commission's standard quality, whereas the scores in green font represent the pass marks, according to Temizöz (2013). Meanwhile, the translations that failed to score above the pass mark threshold are marked in red.

Translator	HT		PEMT		PETM+MT	
	AREN	ENAR	AREN	ENAR	AREN	ENAR
1	0.43	0.37	0.73	0.78	0.85	0.88
2	0.54	0.77	0.71	0.75	0.89	0.92
3	0.42	0.65	0.77	0.8	0.91	0.9
4	0.53	0.77	0.73	0.84	0.94	0.89
5	0.46	0.6	0.72	0.74	0.81	0.93
6	0.54	0.41	0.74	0.69	0.85	0.93
Arithmetic mean	**0.49**	**0.6**	**0.73 (+49%)**	**0.77 (+28.3%)**	**0.88 (+79.6%)**	**0.91 (+51.7%)**

Table 1: The normalised score of the translations in different tasks and translation directions

Based on the results, the students performed better when translating from English into Arabic, indicating that their proficiency in Arabic and background knowledge positively affected their performance. In the PETM+MT tasks, the results show that overall the students managed to pass the quality threshold set by European Commission in the English-Arabic translation, indicating that the non-native speakers can produce good quality translations when using both MT and TM as their resources. Note that this study did not provide other types of references, such as access to web searches. Therefore, having additional suitable resources may increase their translation quality. Meanwhile, in the Arabic-English translations, overall, the students only managed to score 0.88. Although they did not score above the threshold for professional translators, they managed to pass the threshold set for translation students, indicating that they can be highly skilled professionals in Arabic-English translation with years of training and work experience.

Even though none of them passed the quality threshold set by the European Commission and Temizöz (2013) in both HT and PEMT tasks, the overall quality of the English-Arabic translation improved by 28.3% and 51.7% in the PEMT and PETM+MT tasks respectively when compared to their average score in the HT tasks. The improved scores indicate that the

non-native speakers benefitted from the post-editing tasks, particularly with the help of translation memory, even though it was deliberately modified with errors. This result supports the findings of the previous similar studies (Daems et al., 2013; Garcia, 2011), implying that the translation quality increased through PEMT tasks, and post-editing helped non-native speakers produce inverse translations that are of good quality.

In the Arabic-English translations, the overall quality of the students' translations also increased by 49% and 79.6% in the PEMT and PETM+MT tasks, respectively. The considerable increase in the students' performance indicates that post-editing can also help them improve their language proficiency and translation skills even though their translation quality did not even score above-average marks in the HT tasks. This improvement shows that non-native speakers are capable of producing good translations if appropriately trained.

3.2 Error analysis:

In this section, the present study also focuses on investigating the types of errors most commonly found in three different tasks. The findings of the analysis aim to help non-native speakers, mainly those who work with Arabic-English language pair when post-editing machine translation. Also, it can raise awareness of the common errors among Malay learners of Arabic and English when translating from Arabic into English and vice versa.

3.2.1 Frequently found errors in HT tasks

Based on the results shown in Figure 2, lexical errors were identified as the most problematic in the Arabic-English translation, accounting for 18.4% of the errors. These errors may be due to the student's unfamiliarity with the scientific terms in the text, even though it contains several religious terms. Other types of errors most commonly found in the Arabic-English translations are incorrect grammatical numbers, distortion of meaning, too literal, awkward style and syntactic errors.

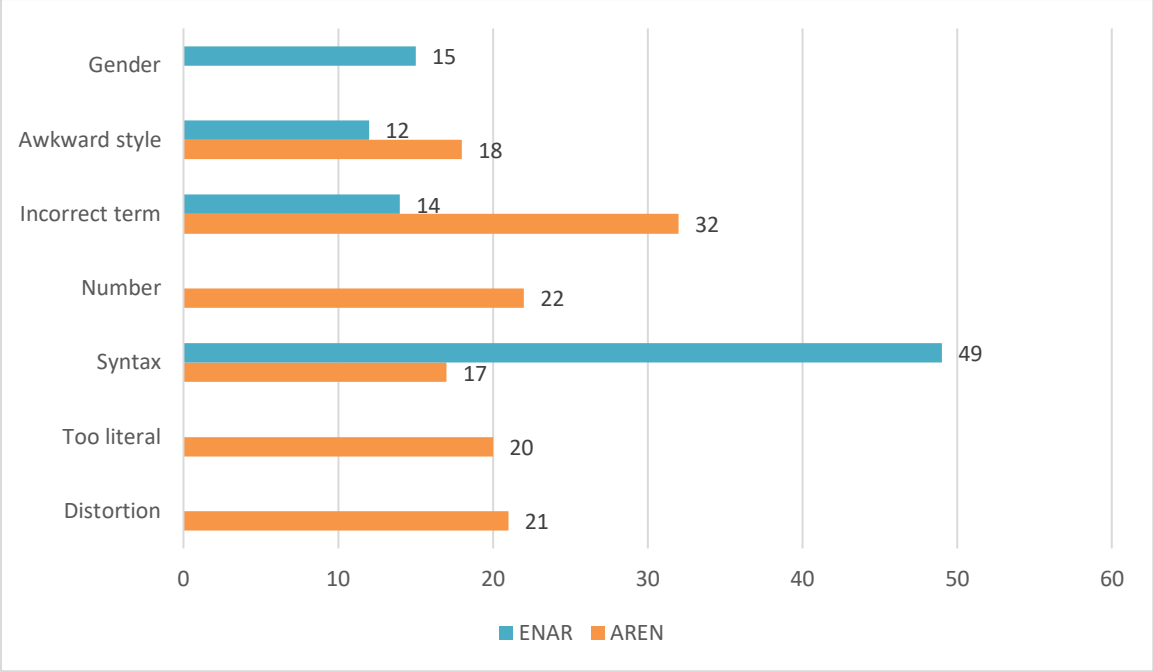

Figure 2: The types of errors commonly identified in the HT tasks.

On the other hand, the common types of errors found in the English-Arabic translations are all language-related, indicating that the students' translations are readable, and did not have any problems in transferring the English contents into Arabic. Syntactic errors were identified as the most problematic as they account for 37.1% of the total errors, suggesting that the students need to pay attention to syntactic errors or improve their syntactic skills when translating from English into Arabic.

3.2.2 Frequently found errors in PEMT tasks

Figure 3 presents the top common errors in PEMT tasks. It is worth noting that these errors are valuable for MT users or post-editors as they could avoid making recurrent MT errors. Also, developers could benefit from the list as they can focus on improving the MT quality. The results show that syntactic and lexical errors contributed to the highest number of errors in both Arabic-English and English-Arabic directions.

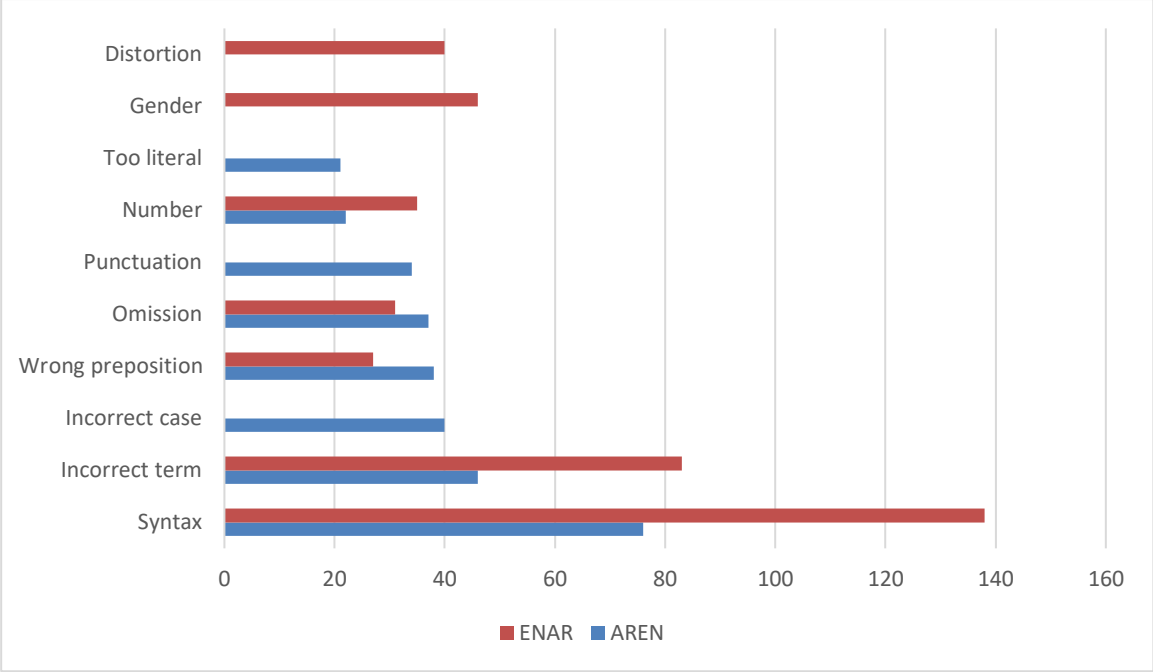

Figure 3: The types of errors commonly identified in the PEMT tasks.

- Syntactic errors:

  Syntactic errors account for 19.3% and 29.9% of the total errors, respectively. The analysis also revealed that the students did not pay attention to three types of syntactic errors: articles, the conjunction "و" and word order. All three types of syntactic are more likely to be related to L1 interference, as syntactically Malay differs from Arabic and English in certain aspects.

  - Articles

  The definite articles "the" and "ال" are equivalent to "itu" which means "that" in English and "تلك" in Arabic, if it indicates femininity. The indefinite articles in Malay are typically indicated with quantity words or classifiers, which do not necessarily have equivalent terms in Arabic or English. Even if they do have equivalence in Malay, typically they would sound awkward. For example, the Malay term "sebiji", which is a classifier for "epal" (apple or تفاحة). It has no direct equivalent term in Arabic and English, but it can only be compensated with the indefinite article "an" (an apple) in English or omitting "ال" from the noun "تفاحة" in Arabic.

  - The conjunction "و" (wa)

  Meanwhile, the conjunction "wa" accounts for 46.4% of the syntactic errors in the English-Arabic translations. A possible explanation for this is that Malay and

English hardly use the conjunction “dan” and “and” to start sentences in academic writing. On the contrary, the conjunction “wa” is “frequently used at the beginning of the sentences and paragraphs but not the first” (Fareh, 1998).

- Word order

On the other hand, word order accounts for 14.5% of the syntactic errors in the English-Arabic translations. Word order differs in the three languages. Hence, the differences may lead to syntactical errors if the students do not thoroughly check their sentence structure.

- Incorrect gender:

  The present study also revealed that gender is also a significant issue in English-Arabic translations. Again, this type of error can be attributed to linguistic interference. In Arabic, most words must indicate gender, but Malay is gender-neutral in most cases. Similar to Malay, English lacks grammatical gender. Therefore, Malay speakers may not notice their mistakes or gender errors in the MT outputs if they are not careful.

- Incorrect number:

  Another commonly found type of error that can be related to linguistic interference is grammatical numbers because Arabic has singular, dual and plural forms for nouns, verbs, and adjectives, depending on the context, for example: “professional workers” can be translated as “العمال المحترفون”. Both noun and adjective are in plural form, but in English, only the noun is plural.

  It is also worth noting that grammatical numbers can pose problems for Malay speakers when translating from and into Arabic and English because plurality in Malay is usually indicated by reduplication, such as “pegawai-pegawai” which means “officers” in English and “الموظفون” in Arabic, and sometimes quantity and classifier comes before a singular noun to indicate plurality: “5 orang pegawai”, which means “5 officers” in Arabic, and “خمس موظفين” in Arabic.

- Lexical errors:

  As previously mentioned, lexical errors are the top common errors in both Arabic-English and English-Arabic translations, accounting for 9.5% and 17.2% of the total errors, respectively. The main reason is likely due to the students’ unfamiliarity with

specialised technical terms and lack of semantic competence. Hence, the students would not know if the MT engine has mistranslated the terms.

- Too literal:

  The analysis also revealed that too literal frequently occurred in the Arabic-English translations. A possible explanation for this is that naturally Arabic sentences are longer than English sentences, as an average sentence length in Arabic is 20-30 words and may exceed 100 words as stated by Al-Taani, Msallam, and Median (2012, p.109). MT systems do not know how to split sentences. Therefore, the MT outputs tend to be of the same length and do not sound natural in English. Therefore, the study suggests that a long sentence should be split into several sentences to get better results.

3.2.3 Frequently found errors in the PETM+MT tasks

When compared to the number of errors in the PEMT tasks, the overall total of errors in the PETM+MT is relatively low. Therefore, this section only focuses on discussing the errors with the high number of errors.

- Syntactic errors:

  The results in Figure 4 indicate that syntactic errors contributed the most errors in the Arabic-English translations, accounting for a total of 63 errors, in contrast with only 13 errors in the EN-AR translations. The difference in the number of syntactic errors supports the fact that the Malay students are more proficient in Arabic than in English.

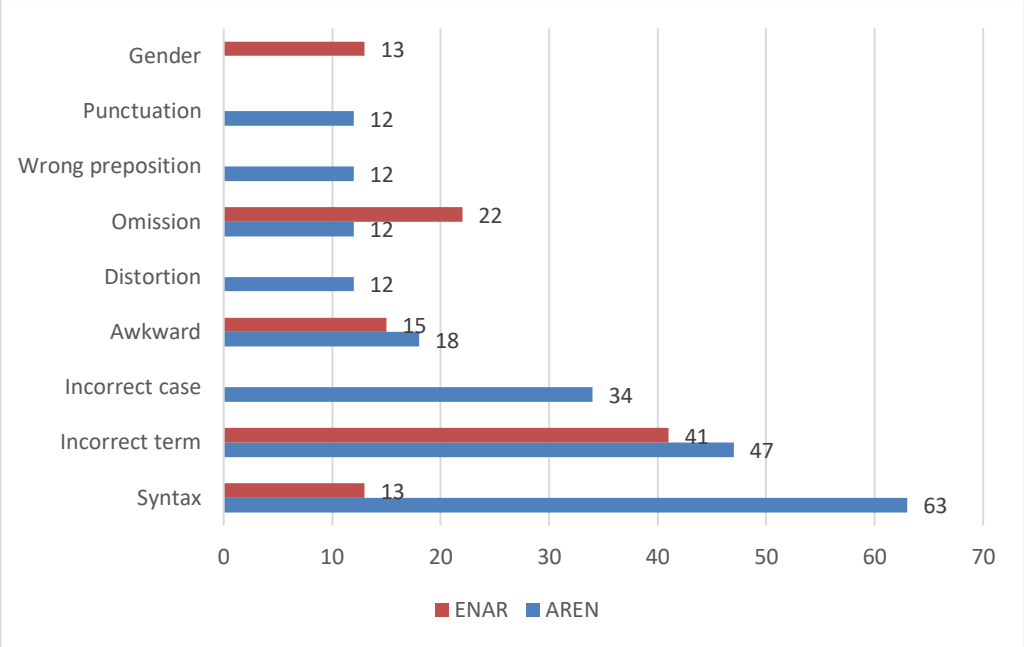

Figure 4: The types of errors commonly identified in the PETM+MT tasks.

- Lexical errors:

  Based on the results in Figure 4, lexical errors contributed to the most errors in EN-AR translation. Apart from the students' unfamiliarity with the technical terms, the students may have overlooked the lexical errors included in the TM. Another possible reason is that the errors might have come from the MT outputs, as previously mentioned in Section 3.2.2.

- Incorrect cases:

  As we can see in Figure 4, incorrect cases are one of the top common errors in the Arabic-English translations. Unlike English, Arabic does not have capital letters. Based on the analysis, GT tends to use upper cases for words that can be commonly associated with proper nouns, such as *forum* and *states*. However, this is not always the case. GT also use small letters for the word *department* when translating دائرة السياحة والتسويق التجاري as it should be rendered with a capital letter as Tourism and Commercial Marketing Department. Even though incorrect cases are minor errors, the students should check the cases thoroughly when translating proper nouns into English because the number of incorrect cases in the Arabic-English translations is relatively high. Hence, it can affect the naturalness of the sentences.

- Omissions:

  Based on the results, one of the top common errors in the AR-EN translation is omissions. A possible explanation for this is that the students may have occasionally overlooked the missing words that MT failed to translate or those that are deliberately omitted from the TM. Their occasional failure to detect the errors suggests that MT and TM can only assist them to a certain extent. If they are not careful, these errors are bound to happen.

4. DISCUSSION & FUTURE WORK

The overall findings suggest that inverse translation is an acceptable practice. Results show that even though the non-native speakers' translations were only average, they managed to improve the quality of their translation with the help of translation technologies, such as translation memory and machine translation. The considerable increase in quality indicates that the post-editing tasks are practical and suitable for training non-native novice translators, as these tasks test their language proficiency and challenge their critical thinking skills. The findings also suggest that non-native novice translator can be trained to become professionals that produce good quality translations. However, this would take years of experience and proper training.

Syntactic errors and lexical errors were identified the top common errors in both English-Arabic and Arabic-English translations. There are several possible explanations for this:

- The students were unfamiliar with the technical terms when translating specialised texts.
- The students did not pay attention to the errors and consequently, overlooked them.
- The differences between the languages, including their mother tongue Malay, may have significantly influenced the students' performance. In linguistics, this influence is called "linguistic interference", which is evident in the high number of errors in articles, the conjunction "wa", word order, grammatical genders, and numbers.

Therefore, this present study suggests that trainee translators, particularly non-native speakers, should pay attention to the listed types of errors and use them as a guideline to avoid making recurring mistakes. The students should also be trained to check their translations thoroughly as minor errors, such as incorrect cases and punctuations, can be recurrent and overlooked even though they can be corrected easily.

It is also important to point out that translators and post-editors should have excellent research skills so that they know where to find suitable sources of references when checking the facts contained in the source text and also when looking for the equivalent terms in the target language, such as trusted and accredited websites, specialised dictionaries and academic books, because different target readers and cultures typically have different technical terms.

The present study only focuses on the quality of the inverse translation. Its scope can be extended to various aspects. For example, the translators' productivity (speed and edit distance) should be investigated as nowadays there is a growing demand in the translation industry. Hence, whether or not inverse translations are acceptable, we cannot deny the fact that the number of native translators is insufficient to meet the high demands in the industry. Moreover, in some parts of the world, translation companies and agencies lack native translators for the target language. Hence, they require non-native translators.

Another aspect that is somewhat popular at the moment is the cognitive approach to the translation and post-editing processes, using advanced technologies such as eye-tracking and keystroke logging. This method is useful for understanding how the non-native speakers carry out inverse translations and also identifying the linguistic features that they struggle to translate.

REFERENCES

Al-Taani, A.T., Msallam, M.M., & Wedian, S.A. (2012), A top-down chart parser for analysing Arabic sentences. *The International Arab Journal of Information Technology*, 9(2), 109-116. Retrieved from https://eis.hu.edu.jo/Deanshipfiles/pub109914508.pdf

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, 108(1), 109-120. Retrieved from https://www.degruyter.com/downloadpdf/j/pralin.2017.108.issue-1/pralin-2017-0013/pralin-2017-0013.xml

Daems, J., Macken, L., & Vandepitte, S. (2013). Quality as the sum of its parts: A two-step approach for the identification of translation problems and translation quality assessment for HT and MT+PE. In *Proceedings of MT Summit XIV Workshop on Post-Editing Technology and Practice*, 2, 63-71. Retrieved from https://pdfs.semanticscholar.org/d845/3786f35a746fffcda098a1702f5f0b9759a2.pdf

Fareh, S. (1998). The functions of and and wa in English and Arabic written discourse. In: *Papers and Studies in Contrastive Linguistics*, 303-312.

Garcia, I. (2011). Translating by post-editing: is it the way forward? *Machine Translation*, 25, 217-237. Retrieved from https://link.springer.com/article/10.1007/s10590-011-9115-8

Guerberof, A.A. (2012). *Productivity and quality in the post-editing of outputs from translation memories and machine translation* (Doctoral dissertation). Universitat Rovira I Virgili, Tarragona, Spain.

IAPTI. (2015). Translation into a non-native language. Retrieved from: https://www.iapti.org/files/surveys/2/IAPTI_non-native_report.pdf

Koponen, M., Aziz, W., Ramos, L., & Specia, L. (2012). Post-editing time as a measure of cognitive effort. In *Proceedings of WPTP*, 11-20. Retrieved from http://www.mt-archive.info/AMTA-2012-Koponen.pdf

Koponen, M., & Salmi, L. (2015). On the correctness of machine translation: A machine translation post-editing task. *The Journal of Specialised Translation*, 23, 118-136. Retrieved from http://www.jostrans.org/issue23/art_koponen.pdf

Newmark, P. (1988). A textbook of translation. New York: Prentice Hall.

Paspartu. (2016). Quality Assurance. Retrieved from https://www.paspartu.gr/en/1/quality-assessment-c92.html

Sánchez-Gijón, P., & Torres-Hostench, O. (2014). MT Post-editing into the mother tongue or into a foreign language? Spanish-to-English MT translation output post-edited by translation trainees. In: O'Brien, S., Simard, M., & Specia, L. Eds. *Thirds Workshop on Post-editing Techniques and Practice*, 5-19. Retrieved from http://www.mt-archive.info/10/AMTA-2014-W2-Sanchez-Gijon.pdf

Secară, A. (2005). Translation evaluation: A state of the art survey. In *Proceedings of the eCoLoRe/MeLLANGE workshop*, 39-44. Retrieved from https://pdfs.semanticscholar.org/e5b3/a34db96b2e4ebb4d621bc4f6b8a9735e8f68.pdf

Specia, L, & Farzindar, A. (2010). Estimating machine translation post-editing effort with HTER. In *Proceedings of the Second Joint EM+/CNGL Workshop Bringing MT to the User: Research on Integrating MT in the Translation Industry (JEC 10)*. 33-41. Retrieved from https://pdfs.semanticscholar.org/6410/e3bf9c780bef4ada5a8eaac7532c9297d082.pdf

Temizöz, Ö. (2013). Post-editing machine translation output and its ervision: Subject-matter experts versus professional translators (Doctoral dissertation). Universitat Rovira I Virgili, Tarragona, Spain.

Toral, A., Sánchez-Cartagena, V.M. A multifaceted evaluation of neural versus phrase-based machine translation for 9 language directions. Retrieved from https://pdfs.semanticscholar.org/7b77/61e0c3c35278a8104994d8bd63fb0b91bb86.pdf