

Investigating Distinctive Problems Observed in Post-editing of Machine Translation Output Made by Indonesian Undergraduate Students

Sri Harto*, Bachrudin Musthafa, Sri Setyarini

Department of English Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

*Corresponding author. Email: harto@upi.edu

ABSTRACT

This research aims to explore the distinctive problems encountered by students in the process of post-editing of machine translation output in the translation of academic texts. Fifty-eight undergraduate students majoring in English language education sitting in fourth semester and taking Practice of Translating course were purposively recruited to participate in this study. Data of this qualitative case study were collected through classroom observation, students' translation documents and students' reflective notes. Data were then analyzed on the basis of translation and post-editing theoretical framework. Results of the research revealed that students faced some distinctive problems in terms of terminology, grammar, particularly in translating the texts from Indonesian into English, choice of words, and inconsistency in the use of technical terms which actually referred to the same contexts. In addition, student translators also experienced significant problems with collocations, inadequate subject knowledge of the translated texts and source language texts that should actually be controlled to result in more understandable outputs. Then, the students should also be encouraged to engage themselves more intensively in the post-editing activities through providing them with some relevant and sufficient post-editing practices and experiences. These activities are intended to uncover the obstacles and distinctive problems faced by the undergraduate students to achieve high quality translation results through post-editing process.

Keywords: *Distinctive problems, machine translation output, post-editing*

1. INTRODUCTION

In the digital and technological era, world documents are distributed in a variety of languages based on the targeted readers through the process of translation. It cannot be denied that the demands of translation works are growing very rapidly (Çetiner & İşısağ, 2019; Čulo & Nitzke, 2016) due to the global growth of translation industry. In response to the needs of the translation industry, post-editing (PE) of machine translation (MT) output is now considered relevant to be implemented by translators as PE has become an integral part of the translation industry (Nitzke, Hansen-Schirra, & Canfora, 2019). This has changed the people's mindset particularly human translators who always think of doing translation works through translation from scratch (TfS) without using the translation machine (MT) in which the MT actually can

help the translators to make their first translation draft and complete the translation works much faster.

The process of post-editing (PE) of machine translation (MT) output through the final touch of a human translator (HT) (Daems et al., 2017) is believed to produce large quantity of works with high quality translation. The post-editing of machine translation (PEMT) aims to answer the recent phenomenon complaining the high productivity of MT output with lower quality and high quality of human translators (HTs) with lower productivity by combining the MT output with the work of HT (MTHT) through post-editing process. In order to achieve the high productivity (HP) and high quality (HQ) process of translation (herein referred to as HPHQ), therefore, PEMT output, particularly in the language pairs of English-Indonesian and Indonesian English, is worth investigating. The investigation of the PEMT output in these language pairs is done, for instance, through taking a look at

various difficulties and problems encountered by student translators taking translation courses at the undergraduate program majoring in English language education who are prospective to be future translators.

In the process of post-editing there are some problems that might be found and these problems should be closely seen to make sure that the problems are clearly identified. The clear identification of the post-editing problems faced by students will help the lecturers to find out the appropriate solutions for the identified problems. The problems cannot be avoided since these will remain problems when the solutions have not been found. This research is needed to find out the patterns of the existing post-editing problems faced by students in the process of translating. Then, based on the patterns of these problems, they are classified into several categories and solutions will be made possible on the basis of data obtained from the fields with regard to the students' experiences in doing the post-editing process. An investigation on the distinctive problems faced by the undergraduate students as future translators is very crucial to be implemented in order to find out ways out of the students' problems. The research on the post-editing of machine translation output is, therefore, worth considering.

2. METHODS

The present qualitative case study investigates the distinctive problems observed in post-editing of machine translation output made by undergraduate students in a prominent public university in Bandung, Indonesia. There were 58 students at the undergraduate program majoring in English language education voluntarily recruited to be participants in this study. These fourth semester students took Practice of Translating (PoT) course worth two credit hours during the semester. Data of the research were obtained from three main instruments, i.e. classroom observation, students' translation documents and students' reflective notes. During the classroom observation, the teaching and learning activities were observed to see the materials taught to students, the students' learning strategies, media used in the process of teaching and kind of tasks assigned to students both as classroom tasks and as home assignments. Other important points to note during the classroom observation were done through taking a look at the patterns of interaction between lecturers and students and between one student and another.

Students' translation documents and students' reflective notes were identified to see the progress of students' learning through their translation documents. The identification of the documents was also important to see the obstacles and challenges faced by students in the process of post-editing activities. The students were required to translate an abstract from English into

Indonesian with the use of a Google Translate (GT) and they post-edited the output of the GT to result in a quality translation. The process of post-editing was intended to produce quality post-edited texts (Vieira, 2020) done by the undergraduate students who were studying Practice of Translating (PoT) course. The post-edited texts were expected to obtain accurate, faithful, grammatical, idiomatic (if any) & informative target texts (Sin-wai, 2016). In addition, a number of errors found in the target texts that should be corrected, according to Sin-wai (2016), among others, were linguistic, referential, stylistic, syntactical, terminological, and typological errors. These can be used as guidelines to analyze the data resulted from this research.

3. FINDINGS AND DISCUSSION

Findings of the research were classified into some patterns with regard to students' problems observed in the post-editing process. There were seven distinctive problems found in their translation practices, i.e. terminological problems, inappropriate use of grammatical rules, irrelevant word choices, inconsistent use of technical terms, violations in translating collocations, inadequate subject knowledge of the translated texts and uncontrolled source language texts. Details of each finding are described in the following sections.

3.1. Terminological Problems

The accuracy of translating terminologies in particular areas of interest was needed and the translators were required to have some competences and skills on the subjects of the translation. The techniques in doing the translation were sometimes different from one subject to another. For instance, the translation of natural sciences would be different from that of social sciences texts. The terminologies used in translating technical materials were usually standardized terms which were commonly used in the relevant fields. In this particular task, during the research, students were required to translate an abstract of a journal article from English into Indonesian entitled "The Representation of Indonesian Cultural Diversity in Middle School English Textbooks" written by Parlindungan, Rifai, and Satriani (2018) and published by an international journal publisher indexed with Scopus. The results of the students' translation works indicated that there were some technical terminologies commonly used in writing the academic texts, such as pedagogically, linguistic knowledge, and semiotic approach. The outputs of the machine translation (MT) were pedagogically ("*secara pedagogis*"), linguistic knowledge ("*pengetahuan linguistik*"), and semiotic approach ("*pendekatan semiotik*") consecutively. The students have already had the awareness on these terms, as the consequence, they

did not post-edit them since they had an understanding that the translation resulted from the MT has been acceptable and they, therefore, did not need to post-edit them. Viewed from some students' reflective notes (for instance: R35, R49, R53) indicated that the students actually encountered problems to find relevant terminologies used in this field of study.

3.2. Inappropriate Use of Grammatical Rules

Problems with grammar found in the students' post-editing practices were closely related with the translators' linguistic competences and skills of both the source language and the target language. Translators' grammatical competence could be identified when the source language texts were read and they were compared to those of the machine translation outputs. It would be a hard work for translators to work on if many differences of grammatical rules were found between the source language and the target language texts. Some of the differences, among others, were positions of subject, predicate, object, and complement in each language and the combinations of some elements of noun phrase which were very different between English and Indonesian. In the case of post-editing of machine translation outputs in the academic texts, some inappropriate use of grammatical rules was still found in the students' translation from English into Indonesian. For example, "cultural sensitivity" was translated a student (R9) into "*budaya yang sensitif*" in which it was correctly machine-translated by using Google Translate (GT, 2020) into "*sensitivitas budaya*" (Translate.Google.Com). It indicated that as native speakers of Indonesian, the students still found it difficult to use Indonesian language grammar appropriately. Therefore, the student translators were required to master the grammatical rules of both English and Indonesian in order to be professional translators.

3.3. Irrelevant Word Choices

Acceptable translation equivalent would determine readers to easily understand the content of the translated texts made by the student translators. To result in acceptable translation works, the translators were not only expected to truly understand the subject knowledge of source language texts, but also required to understand the contexts and target language readers. The determining factors to achieve acceptable translations from source language to target language were made through keeping the meanings (Catford, 1965; Nida & Taber, 1982; Larson, 1984; Newmark, 1988; Machali, 2000; Suryawinata & Hariyanto, 2016; House, 2017) of the translated texts and through selecting the most appropriate choice of words in the target language.

These could be influenced by several factors, among others, were the competences and skills of the

translators, the translators' understanding on the subject knowledge, and contexts of the translated texts. To some novice translators, these competences and skills had to be sharpened through some translation practices and experiences in selecting relevant words to represent ideas that were shared by source language writers. Some samples were found to be significant to explore the post-editing (PE) of machine translation (MT) output process where the MT output was basically resulted from the vocabularies recorded in the system. As the result, many of the MT outputs were not relevant with the context of the source texts. Some of the problems encountered by students as novice translators in the process of post-editing of machine translation (PEMT) output are indicated in the following cases. For instance, the MT output of a phrase "critical issue" was "*masalah kritis*" which was not post-edited by all student respondents (R1-R58) with the appropriate words. For example, "...become a critical issue because..." was translated into "...*menjadi masalah kritis karena* ...". In this particular case, the phrase "critical issue" was not adequately translated into "*masalah kritis*". The novice translators were actually supposed to give the final touch to the phrase through selecting the appropriate words on the basis of contexts. This phrase could be post-edited by changing the phrase into "*isu penting*" or "*masalah penting*". The word "*kritis*" in its original phrase "*masalah kritis*" in Indonesian language was not considered relevant by the Indonesian dictionary known as *Kamus Besar Bahasa Indonesia* (KBBI) (Indonesia, 2016) since "*kritis*" is an adjective to signify a condition identified to be very serious, dangerous, and showing an emergency situation.

Example of the second case was the English word "exposure". The MT output of this word was "*paparan*" which was then post-edited by some students into "*pemaparan*" (R17, R28, R33, R34, R48, R50, R54), "*pengungkapan*" (R20), "*penyebaran*" (R27, R55), "*pengaruh*" (R29), "*pengajaran*" (R39), and "*penggambaran*" (R42). Viewed from its contextual aspect and its subject of discussion, the word "exposure" did not mean to say the words written by students. In addition, viewed from the context that the textbook was used to refer to the media for the teaching of English and its subject was the textbook, therefore, the students' post-edited texts were not considered appropriate in terms of its lexical meaning. These words would be more adequate to be translated into "*media pendukung*". As the consequence, when the choices of words were used by students, some differences of meaning found in the same sentence using the word "exposure". The differences can be found in Table 1.

T1: In Indonesia, exposure to the English language is...

Table 1. Varieties of post-edited word choices

English sentence	In Indonesia,	exposure to the English language	is...
Indonesian sentence-1	Di Indonesia,	pemaparan bahasa Inggris (R4, R9, R12, R16, R17, R18, R21, R22, R23, R26, R29, R33, R34, R45, R48)	umumnya dimediasi oleh...
Indonesian sentence-2		pengungkapan bahasa Inggris (R20)	
Indonesian sentence-3		penyebaran bahasa Inggris (R27)	
Indonesian sentence-4		pengaruh bahasa Inggris (R29)	
Indonesian sentence-5		pengajaran bahasa Inggris (R39)	
Indonesian sentence-6		penggambaran bahasa Inggris (R42)	

Among the six sentences in Table 1, it can be identified that the sentences have differences in meaning and they even have far differences in meaning from the sentences shown in the following when the word “exposure” is translated into “*media pendukung*”. So, it will be written “*Di Indonesia, media pendukung untuk bahasa Inggris umumnya dimediasi...*”. The words “*media pendukung*” in this sentence will be more acceptable and the readers may understand it that the textbook is a media which is used for the teaching of English.

3.4. Inconsistent Use of Technical Terms

3.4.1. Inconsistency in Making Meaning

Consistency in translation is needed to understand the overall message of the source language texts. Consistency usually happens in the texts which repeat words or phrases both in words level and in the level of phrases or in the form of pronoun when it happens in a noun. Inconsistency in the process of translation is also possible in nouns, verbs, adjectives, and adverbs which are functioning as subject, predicate, object, or adverb in a sentence. In the case of post-editing, post-editor’s awareness to have high consistency in post-editing is very crucial. An example of post-editing practices found in translation class was faced by several students who were inconsistent in editing the texts, particularly in the translation of the word “diversity” which is translated into Indonesian very differently into “*diversitas*”, “*keragaman*”, and “*keanekaragaman*” within the same abstract as it is indicated in the following case (R313A=Respondent-3 I Class 3A).

- R1 “The Representation of Indonesian Cultural Diversity in Middle School English Textbooks” (“*Representasi Diversitas Budaya Indonesia di dalam Buku Bahasa Inggris Sekolah Menengah Pertama*”).
- R2 “... how Indonesian cultural diversity and whose cultures are represented in this textbook ...”.

(... bagaimana keragaman budaya Indonesia dan budaya yang diwakili dalam buku pelajaran bahasa Indonesia...”).

- R3 “...depiction of Indonesian cultural diversity”.
(... penggambaran yang seimbang dan setara dari keragaman budaya Indonesia.”).

In the cases of R1 and R2, two examples were made to describe the inconsistency in post-editing the word “diversity” found in an abstract title (T1) and in a sentence written in an abstract (r2). This inconsistency may interfere readers’ understanding in reading the abstract since there is a tendency that the word “*diversitas*” and “*keragaman*” were referring to the different words in which the two words actually have the same meaning. The word “*diversitas*” is taken from an English word “diversity”, meanwhile the words “*keragaman*” and “*keanekaragaman*” Indonesian words. This inconsistency happened in the use of technical terms and its inconsistency did not happen in the kind of words. Inconsistency also happened in the word “textbook”. In this particular case, “textbook” was translated into “*buku bacaan*” and “*buku teks*” (R4) and “*buku pelajaran*” (R5) where it supposed to be “*buku sumber*” which is used as a reference in students’ learning. An example of the case is indicated in the following (R4I3A).

- R4 “The Representation of Indonesian Cultural Diversity in Middle School English **Textbooks**”.
(“*Representasi Keanekaragaman Budaya Indonesia di Buku Bacaan Bahasa Inggris Sekolah Menengah*.”)
- R5 “... through the use of English textbooks.
(... melalui penggunaan **buku teks** bahasa Inggris.”)

3.4.2. Inconsistency in Using Standard Terms

Inconsistency in using standard terms was done by the translators through creating their own terms which were not found in the Indonesian language dictionary (KBBI) (Indonesia, 2016). In the case of post-editing of machine translation output from English into Indonesian, it is important to note that not all of the adapted words have been registered in the Indonesian dictionary. Moreover, the translation of academic texts required the translators to use a formal language. Examples of inconsistency in using the standard terms can be seen in the translation of abstracts in the Table 2.

Table 2. Source text, MT output and post-edited text

No	Source text	MT output	Post-edited text
1	appropriation	<i>apropriasi</i>	<i>kesesuaian</i> (R10)
2	inclusiveness	<i>inclusivitas</i>	<i>inclusivitas</i> (R1-R58)
3	inclusive	<i>inklusif</i>	<i>inclusive</i> (R1-R58)
4	pedagogically	<i>secara pedagogis</i>	<i>secara pedagogis</i> (R1-R58)
5	sensitivity	<i>sensitivitas</i>	<i>sensitivitas</i> (R1-R58)

The English word “appropriation” in the source language text and the result of MT output was “*apropriasi*” seemed to be correct since the word “*apropriasi*” was considered to be an adapted word from “appropriation”. However, when taking a look at the Indonesian dictionary (KBBI) (Indonesia, 2016), the word was not registered in the dictionary yet. In addition, the MT output “*apropriasi*” alone did not represent the intended meaning of the source language text. To some of the students who have already had good translation skill, they would post-edit the word “*apropriasi*” into “*kesesuaian*”. The word “*kesesuaian*” was considered to be more acceptable compared to that of “*apropriasi*”, but most of the students did not edit the word and accept to use the output “*apropriasi*” as suggested by MT. The same case also happened to the word “inclusiveness” in which its suggested output of the MT was “*inklusivitas*”. This word has not been found in the Indonesian dictionary (KBBI) (Indonesia, 2016), but the students thought that this word was appropriate to use by them.

- R6 “... with cultural sensitivity and appropriation for the inclusiveness of other cultural values ...”
*Dengan sensitivitas dan **apropriasi** budaya untuk **inklusivitas** nilai-nilai budaya lain.*
*Dengan sensitivitas dan **kesesuaian** budaya untuk **inklusivitas** nilai-nilai budaya lain.”*

Different from the above two cases, the three words written in the table, “inclusive”, “pedagogically”, and “sensitivity” have been adapted into Indonesian into “*inklusif*”, “*pedagogis*”, and “*sensitivitas*”. Therefore, when these three words were used in the translation, they can be accepted and they were consistent with the standardized terms used in the Indonesian language dictionary (KBBI) (Indonesia, 2016).

3.5. Violations in Translating (Using) Collocations

A collocation is a habitual co-occurrence of individual lexical items that enter into high-frequency grammatical structures (Newmark, 1988). Newmark (1988) then provides some examples of the collocation, i.e. adjective plus noun, noun plus noun, and verb plus object, to give some adequate description on finding out appropriate connections between nouns and verbs, adjectives and nouns, adverbial groups and verbs, and appropriate connectives or conjunctions. Noun compounds are common collocations found in social sciences (Newmark, 1988). Collocation can be found in terms of grammatical aspect and lexical aspect. In the aspect of grammar, collocation in English is often times found in words followed by prepositions. In the translation practice through post-editing of MT output made by students in a journal abstract it was found out that there were 10 words that can be grouped into prepositional collocation as indicated in the Table 3.

Table 3. Samples of collocations in source text, MT output, and post-edited text

Source text	MT output	Post-edited text
exposure to	<i>paparan</i>	PEMAPARAN (R17, R28, R33, R34, R48, R50, R54) <i>pengungkapan</i> (R20) <i>penyebaran</i> (R27, R55) <i>pengaruh</i> (R29) <i>pengajaran</i> (R39) <i>penggambaran</i> (R42)
represented in	<i>diwakili dalam</i>	<i>ditunjukkan dalam</i> (R10) <i>ditampilkan dalam</i> (R34) <i>direpresentasikan dalam</i> (R19)
drawing on	<i>menggambar</i> <i>menggambar</i> <i>pada</i>	<i>mengutip dari</i> (R9, R12, R13) <i>mengacu pada</i> (R4, R10, R23) <i>melihat pada</i> (R11) <i>sesuai pada pendekatan</i> (R17) <i>penggambaran dalam</i> (R18) <i>sesuai yang digambarkan oleh</i> (R20)

Notes: MT=Machine translation

From the data in Table 3, it can be seen that most of the students did not have good self-confidence in post-editing the phrase “drawing on”. The MT output indicated that “draw on” was translated into “*menggambar pada*”. Viewed from the perspective of collocation, the Indonesian word “*menggambar*” was not supposed to be collocated with “*pada*”, but “*menggambar dengan*” since the activities of drawing used a certain equipment which was supposed to be “*menggambar di*”. Collocation was considered important since it would influence on differences of meaning. Taking a look at the post-edited translation works made by the students, it can be highlighted that not all students did the post-editing works correctly in prepositional collocations. Some of the examples indicating these problems found in the collocations such as “*melihat pada*”, “*sesuai pada*”, “*penggambaran dalam*” which were considered inappropriate. Alternatively, in this particular context, these Indonesian collocations were suggested to be translated into “*melihat dari*”, “*sesuai dengan*”, and “*menggambar di*” in their target language text. Meanwhile, the lexical collocation in translating the abstract was found in the translation of “play a pivotal role” which was then machine translated into its Indonesian phrase “*memainkan peran penting*” and it was post-edited by some students into “*memiliki peran penting*”.

The above case of collocation was in the Indonesian words “*memainkan*” and “*peran yang dianggap*” which have the true relationship that the role has been played. However, when it was translated into “*memiliki peran*”, it means that it was not a collocation since the word “*memiliki*” does not collocate with “*peran*”.

3.6. Inadequate Subject Knowledge of the Translated Texts

Understanding different text types and area of studies are important competences that should be owned by translators. The understanding of the translators is intended to obtain the content of source language texts. In understanding the source language texts, translators were influenced by whether or not they were familiar with the subject knowledge of the source language text and the translators' linguistic competence in understanding the source language text. The target readers of the translation works determined the translation techniques, choices of words, and level of difficulties of the target language text that should be produced by the translators.

3.7. Uncontrolled Source Language Texts

In the case of translating the abstract, most of the students have already identified the characteristics of academic texts; therefore, they could implement the Indonesian formal language to translate the academic text. They have understood that the source language was an abstract which was resulted from the research findings. As the consequence, the students have already used the common words in analyzing the research findings such as "pendekatan", "metode", "hasil penelitian", and "mengutip dari". The students tended to use literal translation to post-edit the output of the MT.

The students found some difficulties when they were supposed to post-edit MT output containing grammatical problems in its source language text. Therefore, when considering using MT output in their translation works, the students were advised to pre-edit the grammatical problems found in its source language text. Pre-editing itself is closely related with the efforts of controlling source language texts through making some necessary revisions before the texts are machine-translated (Nitzke, 2019) to result in quality translation that meets the expectation of the end users of the translation.

4. CONCLUSION

Post-editing (PE) of machine translation (MT) output should be done since the MT output is still found to have some problems. The elements of appropriate word choices, grammatical rules applied in the target language, and the messages conveyed in the texts were some of the aspects to consider in the post-editing process. The translators' competences and skills in the process of post-editing were required to solve these problems. The translators' competences and skills can be enhanced through the translation practices giving emphasis not only on the translation from scratch (TfS) but also on the use of machine translation output through post-editing (PE) process to result in better

quality translation. The learning process is designed to encourage and to improve the students' skills in post-editing the output of machine translation (MT) through the touch of human translators. By obtaining these skills the students will have some competences to do the translation works and they know when to do them. Therefore, the students' awareness in finding out the problems found in MT output such as inappropriate word choices, problems with collocations, and inappropriate use of grammatical rules can be continuously improved.

REFERENCES

- Catford, J.C. (1965). *A linguistic theory of translation*. Oxford: Oxford University Press.
- Çetiner, C., & İşısağ, K. U. (2019). Undergraduate level translation students' attitudes towards machine translation post-editing training. *International Journal of Languages' Education and Teaching*, 7(1), 110-120.
- Čulo, O., & Nitzke, J. (2016). Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation. In *Proceedings of the 19th Annual Conference of the European Association for Machine Translation*. 106-114).
- Daems, J., Vandepitte, S., Hartsuiker, R. J., & Macken, L. (2017). Identifying the machine translation error types with the greatest impact on post-editing effort. *Frontiers in Psychology*, 8, 1282.
- Google Translate (GT). (2020). Translation from English into Indonesian. Machine-translated on 25 September 2020. Available online at: Translate.Google.Com.
- House, J. (2017). *Translation: The Basics*. Routledge.
- Indonesia, K. B. B. (2016). Badan Pengembangan Dan Pembinaan Bahasa, Kementerian Pendidikan Dan Kebudayaan Republik Indonesia.
- Larson, M. L. (1984). *Meaning-based translation: A guide to cross-language equivalence*. 366. Lanham, MD: University press of America.
- Machali, R. (2000). *Pedoman bagi penerjemah*. PT. Grasindo.
- Newmark, P. (1988). *A textbook of translation* (Vol. 66). New York: Prentice hall.
- Nida, E. A., & Taber, C. R. (Eds.). (1982). *The theory and practice of translation*. 8. Brill Archive.
- Nitzke, J. (2019). *Problem solving activities in post-editing and translation from scratch: A multi-method study*. Language Science Press.

- Nitzke, J., Hansen-Schirra, S., & Canfora, C. (2019). Risk management and post-editing competence. *The Journal of Specialised Translation*, 31, 239-259.
- Parlindungan, F., Rifai, I., & Safriani, A. (2018). The representation of Indonesian cultural diversity in middle school English textbooks. *Indonesian Journal of Applied Linguistics*, 8(2), 289-302.
- Sin-wai, C. (2016). *The future of translation technology: towards a world without Babel*. Taylor & Francis.
- Suryawinata, Z., & Hariyanto, S. (2016). Translation: Bahasan teori & penuntun praktis menerjemahkan (edisi revisi). *Malang: MNC*.
- Vieira, L. N., O'Hagan, M., & O'Sullivan, C. (2020). Understanding the societal impacts of machine translation: a critical review of the literature on medical and legal use cases. *Information, Communication & Society*, 1-18.