

Students' Linguistic Challenges in Post-Editing of Machine Translation

A Case Study of News Translation

Sri Harto*, Fuad Abdul Hamied, Bachrudin Musthafa, Sri Setyarini

Universitas Pendidikan Indonesia, Bandung, Indonesia

*Corresponding author. Email: harto@upi.edu

ABSTRACT

This research seeks to investigate linguistic challenges encountered by students in translating news from English into Indonesian through a process of post-editing of machine translation (PEMT). Since the research on this particular language pair has not been widely explored by Indonesian researchers, therefore, this study should be thoroughly undertaken to provide the students with a comprehensive description of the linguistic problems faced in their translating activities. A qualitative method with a case study design was implemented to portray 58 students who voluntarily participated in the PEMT process. Research data were obtained from students' survey questionnaires, students' in-depth interviews, and students' translation documents taken from some online media individually selected by students in a Practice of Translating (PoT) course. The data were analyzed by using a thematic analysis with regard to the students' translating activities and their obstacles found during PEMT practices. Three dominant themes drawn from the data, including: students' insufficient subject knowledge in translating proper names; particular grammatical problems found in the use of pronouns and in the differences of language structures between English and Indonesian; and mechanical errors identified in the use of punctuations, capitalization, spelling, and inconsistency in using specific terminologies. These research findings have recommended that appropriate teaching techniques should be created by the lecturers to optimally reach the objectives of PEMT teaching.

Keywords: *Linguistic challenges, news translation, post-editing of machine translating.*

1. INTRODUCTION

In the era of translation technology, the use of machine translations (MTs) in doing translation works cannot be avoided. As one of machine translations, Google Translate (GT) has become one of the tools used by some students as novice translators and even by some professional translators in their translating activities both for personal needs and even for professional purposes. The use of GT, as one of the MTs, is considered to be a solution for the demands of the translation industry which cannot be accomplished by human translators (Jia, Carl, & Wang, 2019).

In the context of Indonesia, the use of translation technologies, such as MTs, to many undergraduate students, even to some particular translators, is an effort of making initial drafts as pre-translated texts of their translation assignments. They are assured to use the MTs since the quality of MT outputs is now improving very significantly (Vieira, Alonso, & Bywood, 2019). As a result, the students and the translators are confident to use

the MT for a variety of purposes, from translating academic papers that should be submitted in English to translating journal articles that will be published in international journal publications.

Despite the significant quality development of MT, the students and translators using the MT for their translation tool should be reminded that there are some errors found in the MT outputs. They cannot rely too much on the MT since there are at least three categories of errors that can be identified from the MT outputs, i.e. grammatical errors, lexico-semantic errors, and syntactic errors (Sycz-Opon & Galuskina, 2017; Yamada, 2019; Zaretskaya, Vela, Pastor, & Seghiri, 2016).

Further consideration should also be made when the students and translators are using MT for drafting their translation works. They have to realize that MT is not always relevant to all text types. Particular texts such as literary texts, legal texts, and creative works may not be translated by MT, but it can be applied to non-critical

materials such as web articles (Schaler, Way, & Carl, 2003 in Quah, 2006).

Although news articles are considered to be parts of the web-contents, the translation of news item texts cannot be done alone by MT. The translators are then required to be aware of the errors resulting from the MT outputs and they have to revise the errors to improve the quality of the translated texts. When some typical errors are still found in the texts, they cannot be accepted by end-users, therefore, the post-editing of machine translation (PEMT) process is required (Jia, Carl, & Wang, 2019).

In the Indonesian higher education context, the final texts resulted from the students' PEMT practices can be analyzed to see whether or not the post-edited texts are found to have some errors. The students' errors resulted from PEMT can be analyzed using Rasmussen's skill-based (SB) performance, rule-based (RB) performance, and knowledge-based (KB) performance framework (Rasmussen, 1983). However, this initial framework is implemented by different experts in a variety of fields.

One of the implementations of the Rasmussen's (1983) framework was made by Nitzke (2019) particularly applying it in the field of post-editing of machine translation (PEMT). Through adapting the SB, RB, and KB in the students' translation activities, their errors in post-editing (PE) practices can be analyzed based on the three categories. The errors shown by the students in the three categories, including the skill-based, rule-based, and knowledge-based errors are parts of the students' linguistic challenges in doing their translation activities that should be carefully taken care of to improve the quality of their translation works. As one of the characteristics of texts that can be translated through the use of MT (Schaler, Way, & Carl, 2003 in Quah, 2006), news item texts are considered to be relevant for the students' translating practices through the process of PEMT. Since this particular PEMT practice implemented in news translation in the English-Indonesian language pair has not been widely explored by researchers in the Indonesian research context, therefore, this topic is worth investigating.

2. METHOD

2.1. Research Design

This research applied a qualitative method using a case study design as suggested by Creswell (2012), Malik and Hamied (2016), and Yin (2018). It involved 58 undergraduate students majoring in English language education in a prominent public education university in Indonesia. The students took the Practice of Translating (PoT) course in their fourth semester as one of the compulsory courses that should be taken by the students within the semester.

2.2. Data Collection

The research data were collected from students' survey questionnaires (SQs), students' in-depth interviews (Intvs), and students' translation documents (Docs) in the forms of machine translation (MT) outputs which were then post-edited by the students. Other students' documents resulted from their translation activities collaboratively done by groups of five students through a translation circle (TC). The TC was initiated to provide the students with a discussion forum consisting of five students where one of the students' translated works was collaborated with four of their colleagues in the group. They played roles as collaborators who provided comments, inputs, corrections, and even questions related to the translation works made by one of the members of the TC group. Collaborators were encouraged to find out whether or not their peers' translation works were found to show some problems with grammar, terminology, and accuracy of contents, and mechanics (Massardo, Meer, O'Brien, Hollowood, Aranberri, & Drescher, 2016; Sin-wai, 2017).

2.3. Type of Texts Used by Students for PE Practice

In the Practice of Translating (PoT) course, students practiced translating through both translation from scratch (TfS) and post-editing (PE) of machine translation (MT). With regard to PEMT, since the students' translating practices were assisted by the use of MT, this research was then focused on post-editing of machine translation (PEMT). PEMT here was applied to news item texts in which these were classified into middle and low levels of difficulty (Schaler, Way, & Carl, 2003 in Quah, 2006). Various news articles derived from different media were individually selected by students for the students' post-editing practices. In this research context, the students taking the PoT course were given a chance to translate news text items through the use of Google Translate (GT, 2020) as one of the most selected machine translations (MTs), and the MT outputs were then post-edited by the students to improve the quality of the translation.

2.4. Data Analysis

The thematic analysis (TA) formulated by Braun and Clarke (2006, 2014) and Clarke and Braun (2018) was implemented to identify the research data and classify them into some general categories. The categories of data were then further identified by referring them to Rasmussen's (1983) framework with regard to three levels of human performance model including skill-based, rule-based, and knowledge-based performances. This framework was then adapted by Nitzke (2019)

which was specifically focused on analyzing the translators' errors found in the translation of texts through the post-editing (PE) of machine translation (MT) activities, which in this research context is referred to as PEMT.

In Nitzke's (2019) adaptation, skill-based performance was shown in the ability to apply the appropriate technical writing and avoid errors in the typing process. When the technical errors and writing errors were identified, it did not mean that the students did not have any skills in undertaking the work. These kinds of errors can be identified in the form of punctuation, capitalization, and spelling errors. In addition, rule-based performance was seen from the students' performance in implementing rules of grammar in the texts being edited. For instance, in post-editing of the machine translation (MT) outputs from English into Indonesian and vice versa, errors in the use of tenses in a particular sentence were considered to be rule-based errors. In this particular context, the students were considered to implement inappropriate tenses with regard to the context of time provided in the sentence. Finally, the knowledge-based performance was viewed from the students' knowledge to identify information in more detail and their understanding of the textual meaning of the information in particular (Nitzke, 2019).

3. RESULTS AND DISCUSSION

3.1. Online media selected by students for PE practice

Based on the data of the students' selection on the online media for their translating and post-editing (PE) practices, there were eighteen different international English media purposively selected by 58 students. Each was selected by a different number of students depending on their own interests. The online media selected by students were alphabetically listed to see the names of the media and the number of students who chose the same media for their translating and PE practices, i.e. Antara News (3), BBC (16), BuzzFeed News (1), CBS News (1), Channel News Asia (4), CNN (8), Crash (1), Fox News (1), Independent (1), Japan Today (1), Literary Hub (1), New York Post (1), The Guardian (2), The Herald News (1), The Jakarta Post (13), The New York Times (1), Variety (1), and Vox (1). When the students were given the freedom to choose their own online articles, they had a variety of choices. The details of the online media selected by the students for their PE practices are presented in Figure 1.

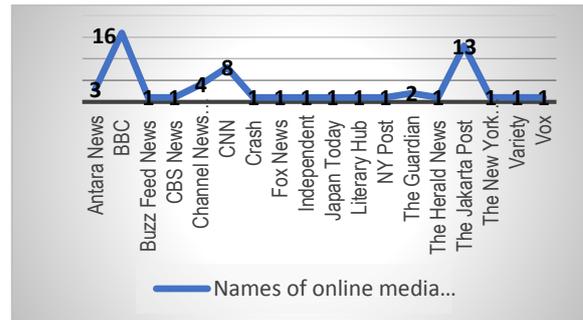


Figure 1 The online media selected by students for PE practice.

3.2. Skill-based performance

Data classification obtained from the documents of post-editing practices undertaken by 58 undergraduate students indicated that post-editing of machine translation (PEMT) was a new activity for the students taking the Practice of Translating (PoT) course. Viewed from the three levels of the students' performance, there were some errors found in the students' works in post-editing practices. These errors included the three levels of performances such as inappropriate use of punctuation, problems in using capital letters for personal names, places, and cities. In addition, problems were also found in writing particular words with appropriate spellings, writing a publisher's name, and inappropriate use of terminologies.

3.2.1. Punctuations

The ways of using punctuations were not the same between the source language and the target language. In order to avoid errors in using punctuations, students were supposed to know ways in writing punctuations. Data analysis derived from the students' post-editing practices, for instance, indicated that a full stop (.), a colon (:), and a comma (,) were three different punctuations commonly neglected by the students in the practice of post-editing (PE). Errors in using a full stop (.) were oftentimes found in the students' post-editing works when the full stop should actually be used to end a completed sentence and a number indicating the number of units. For instance, an error in using a full stop to notify the end of a sentence (Docs, S15) can be seen in Quotation [1].

Quotation [1]

... Amid the restriction, visitors are still allowed to enjoy the views of Borobudur temple from the eighth level. (... Ditengah pembatasan. Pengunjung masih diizinkan untuk menikmati pemandangan candi Borobudur dari teras delapan [without a full stop to end the sentence]. [Docs, S15] (The Jakarta Post, February 18, 2020).

Quotation [1] indicates that there was an error found in using a full stop that should actually be put at the end of the sentence functioning to notify that the sentence has represented a completed idea or thought (Merriam-Webster Dictionary: <https://www.merriam-webster.com/dictionary/sentence>, 2021). Errors in the use of punctuations were also found in the students' post-editing works (Docs, S25), particularly in using a colon (:) to write time units as indicated in Quotation [2].

Quotation [2]

a. Jakarta/Jumat, 21 Februari, 2020/8:00 malam

b. Jakarta / Jumat, 21 Februari 2020 / 10:06 pagi [Docs, S25] (The Jakarta Post, February 21, 2020).

In quotation [2], the use of a colon (:) in writing time units were considered inappropriate since according to the general guidelines for the Indonesian language spelling, in which in the Indonesian contexts, it is known as *Pedoman Umum Ejaan Bahasa Indonesia* (PUEBI), indicating that a full stop was used to notify differences of an hour, minute, and second in the number of time units (PUEBI, 2016). Errors in the use of punctuations in students' post-editing practices were not only found in using a full stop and a colon but also identified in applying a comma and an apostrophe. Errors in using a comma happened when the students post-edited a sentence consisting of a description. For instance, errors in using a comma in a sentence consisting of a description of several words as stated by student-45 (S45) through an interview (Intv-S45) and the results of collaborative works through a translation circle (TC) made by student-33 (TC-S33).

3.2.2. Capital letters

The results of data analysis taken from the students' post-editing practices in the translation of news articles from English into Indonesian indicated that the students made some mistakes in using capital letters. The most common mistakes made by the students in using capital letters occurred in writing the titles of the news articles. The errors made by the students could be identified from the use of capital letters at the beginning of every word written in the title of the news articles including prepositions and connectors used in the title of the news articles. Samples of these errors obtained from the students' works (Docs, S35) can be seen in Quotations [3] and [4].

Quotation [3]

Delhi riots: City tense after Hindu-Muslim clashes leave 23 dead. (Kerusuhan Delhi: Ketegangan kota setelah bentrokan Hindu-Muslim menewaskan 23 orang). [Docs, S35] (Kaumba, 2020).

Quotation [3] shows that the title of the news article was only written using capital letters at the beginning of the sentence. Capital letters were used in every first letter of the words in the title of a book, an article, an essay, a paper, or a magazine except for prepositions, conjunctions, or disjunctions not located at the beginning of a sentence such as "di", "ke", "dari", "dan", "yang", and "untuk" (PUEBI, 2016). In contrast to quotation [3], quotations [4] and [5] are identified to show students' errors in using capital letters that occurred in writing the title of news articles, particularly in writing prepositions using capital letters at the beginning of words.

Quotation [4]

21 of The World's 30 Cities with The Worst Air Pollution are in India. (21 dari 30 Kota di Dunia Dengan Polusi Udara Terburuk Ada di India). [Docs, S40] (Regan, 2020).

Quotation [4] is an example of errors made by students in translating an article title. In this case, the student made a mistake by writing the word "dengan" using a capital letter in the first letter of the word which should be actually written in a lowercase since the word "dengan" in Indonesian is included in a preposition. In contrast to quotation [4], quotation [5] is an example of the student's error which is considered to be more serious than that of quotation [4]. In quotation [5] the student did not write capital letters at the beginning of each word written in the article title not included in prepositions and conjunctions such as "satu", "hari", "paling", "menguntungkan", "tahun", "ini", "didedikasikan", "anak", and "perempuan" which are parts of the title of a news article as indicated in Quotation [5].

Quotation [5]

Hina Matcuri (Doll Festival): One of the year's most auspicious days dedicated to girls. (Hina Matsuri (Festival Boneka): Salah satu hari paling menguntungkan tahun ini yang didedikasikan untuk anak perempuan). [Docs, S45] (Shoji, 2020).

Referring to some examples of the cases quoted above, errors in the use of punctuation and capital letters in writing articles made by the students can be influenced by the students' inaccuracy in editing the texts. In this case, it did not mean that they did not have any insight or skills in writing in the Indonesian language since Indonesian was actually their native language. However, through an interview with one of the students (Intv-S20) involved in the post-editing practice, S20's idea is indicated in Quotation [6].

Quotation [6]

Selama ini, masalah yang dihadapi dalam penerjemahan dari Inggris ke Indonesia adalah kaidah-kaidah dalam Bahasa Indonesia, kadang-kadang itu terlupakan. (So far, the problems encountered [by the students] in translating [texts] from English into Indonesian, among others, are

ignoring the principles of the Indonesian language [writing]) (Intv-S20).

3.2.3. Spelling errors

Spelling errors made by students in writing particular words were mainly caused by the missing of any letters. As a result, the words were not completely written. Errors in spelling particular words might cause different meanings in a sentence. This error would be very serious if it occurred in writing people’s names, names of places, names of institutions, names of days, numbers, and other specific things. These errors could be objectively seen in accordance with the translated source language. Spelling errors were the errors that often occurred in the machine translation (MT) outputs. The causes of the spelling errors found in the MT outputs were influenced by the provision of additional software, differences in grammatical rules applied in the source and target language, and careless practices in students’ typing activities (Nitzke, 2019).

Data derived from students’ post-editing practices in the translation of news articles from English into Indonesian were identified to show that the students were relatively inaccurate in translating specific terminologies with regard to days of the week, dates of the month, prepositions followed by time units, and names of publishers or institutions incompletely written by the students. An example of a spelling error, as identified by Yamada (2019) in the students’ post-editing practice, is shown in Figure 2.

Figure 2 showed that the student was not careful in writing the word “pembalap” since the letter “m” was missing, therefore it was written “pebalap”. In contrast to the case in Figure 2, Figure 3 shows that the students failed to write the word “mekarnya” appropriately by adding the letter “r” before the letter “k”, therefore, it became “merkarnya”. As the consequence, this word was not properly translated since it was not found in the Indonesian language dictionary commonly used by

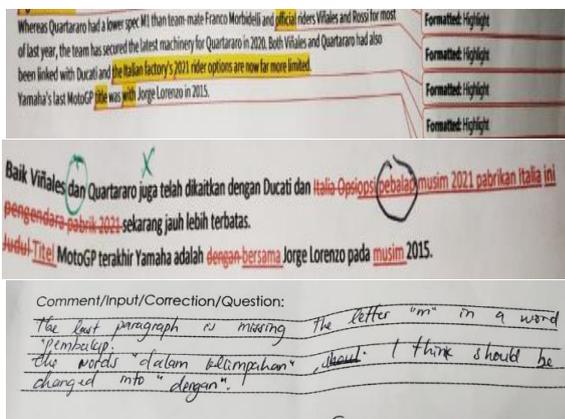


Figure 2 Sample of a spelling error found in students’ PE practice.

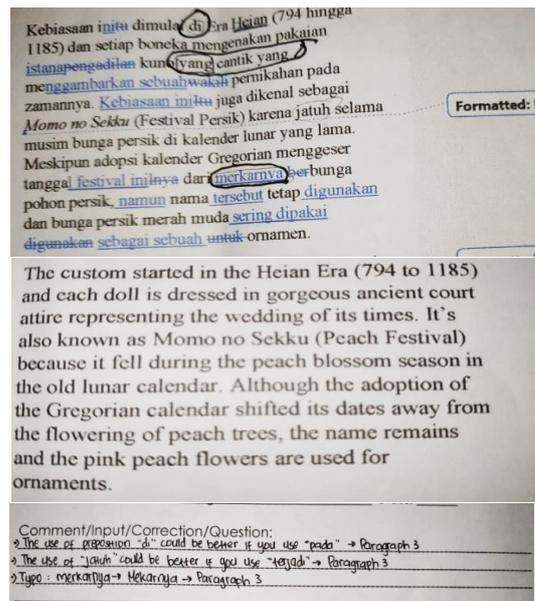


Figure 3 The other sample of a spelling error identified in student’s PE practice (Text was taken from Shoji, 2020).

Indonesian people (PUEBI, 2016) in their everyday life. This was commented on by one of the collaborators in the TC activity as indicated in Figure 3.

The more serious student error in post-editing practice was found in writing the name of a publisher. The publisher’s name “The Jakarta Post” was a proper name, therefore, it cannot be changed through, for example, omitting the word “The” in its proper name. This can be identified in Quotation [7].

Quotation [7]

The Jakarta Post, Jakarta / Fri, February 21, 2020 / 10:06 am
 (The Jakarta Post, Jakarta / Jumat, 21 Februari 2020 / 10:06 pagi) [Docs-S37] (The Jakarta Post, February 21, 2020).

Students’ inaccuracies in editing the MT outputs can be seen objectively by directly comparing the source language and its target language. This does not mean that the students did not know the correct writing rules in the target language. The students might not be careful in editing the texts. To overcome the skill-based errors, the students were advised to use other software such as automatic spelling checkers. In addition, the students needed to give careful attention to whether they had insufficient knowledge of a particular field (Nitzke, 2019) such as the one in writing a publisher’s name, i.e. The Jakarta Post. In this case, there is an assumption that students did not know that the name of the publisher was The Jakarta Post, as a consequence, the word “The” was omitted. In more specific cases, the level of performance required by the students is higher to produce accurate final translation results.

3.2.4. Inconsistency in using specific terminologies

Quality translation results were demonstrated by the use of appropriate and consistent terminologies. In the practice of post-editing, a post-editor must be careful in editing the use of terms or words that were repeated. The use of different words or terms with reference to the same object in one text could lead to different meanings. Student translators who were considered to be inexperienced translators were commonly identified to encounter this problem (Daems, Vandepitte, Hartsuiker, & Macken, 2017). In the PEMT practices, therefore, the students as post-editors were advised to maintain consistency in using specific terminologies (Harto, Musthafa, & Setyarini, 2020). Examples of the inconsistent use of terms found in the students' post-editing practice (Text: The Jakarta Post, February 21, 2020) including the use of words such as "masker" and "masker wajah" which are actually referring to the same object, "face mask". Examples are presented in Figure 4.

3.3. Rule-based performance

The students' errors were classified into the rule-based performance consisting of personal pronouns and problems in differentiating active and passive voices. The rule-based performance in data analysis in this study was focused on errors analysis performed by the students in applying the skills and knowledge of grammatical rules from the source language to the target language. Inappropriate use of grammatical rules in the students' post-editing practices was found to be one of the challenges encountered by the students (Harto, Musthafa, & Setyarini, 2020). In this research context, the source language texts were English and the target language texts were Indonesian. The texts translated by the students were English news purposely selected by the students from various online news articles based on the students'

individual interests. Data of students' post-editing practices were identified in forms of documents indicating that the rule-based errors made by students in post-editing practices were in forms of inaccuracies in translating personal pronouns, noun phrases, verb phrases, conjunctions, active and passive voices, formal versus informal terms, idioms, and tenses. Further descriptions of these cases are presented in the following sections.

3.3.1. Personal pronouns

The personal pronoun is referred to by Huddleston and Pullum (2006) to indicate the first person such as "I" and "we", second person "you", and third person which includes "he", "she", "it", and "they". With regard to the translation of texts from English into Indonesian, translators should be careful to use personal pronouns since there are several personal pronouns that have multiple meanings in the Indonesian language, such as the personal pronoun "we" in which it can be translated into "kita" (referring to the speaker and hearers) and "kami" (referring to the speaker without including the hearers). The phenomena of incorrect pronouns in post-editing have been identified by (Sycz-Opon & Galushkina, 2017), therefore, post-editors have to carefully identify possible errors found in using MT. In this context, the personal pronoun "we" indicates two different types of personal pronouns. Examples of errors in translating the personal pronoun "we" made by student-37 (S37) in post-editing practice are shown in Quotation [8].

Quotation [8]

"However, Nadiem also said it might be difficult to realize his dream. "I'm not sure if we can achieve this goal but we have to dream big," Nadiem said." (Namun, Nadiem juga mengatakan mungkin sulit mewujudkan keinginannya. "Saya tidak yakin apakah kami dapat mencapai tujuan ini, tetapi kami harus bermimpi besar," kata Nadiem) [Docs, S37]. (The Jakarta Post, February 21, 2020).

In case [8], the student translator should correctly translate the word "we", namely us, because the word "we" refers to Nadiem and the readers. Nadiem said in the media to greet the public, therefore, the personal pronoun "we" in this context referred to "kita" instead of "kami". The other example of error was encountered by student-54 (S54) in translating the word "May" which was actually referring to a person's name, May, who was 27 years of age [... "May, 27," ...], but it was translated into Indonesian words [... "27 Mei" ...] meaning the 27th day of May referring to the date of the month. In the case that the student translator (S54) understood it, "May" as a lady's name, Aleali May, could be represented by a personal pronoun "she" or "her" [Docs, S54].

The other problem was found in the quotation [9] indicating that a person's name "May" was translated into "Mei" referring to the fifth month of the year. The

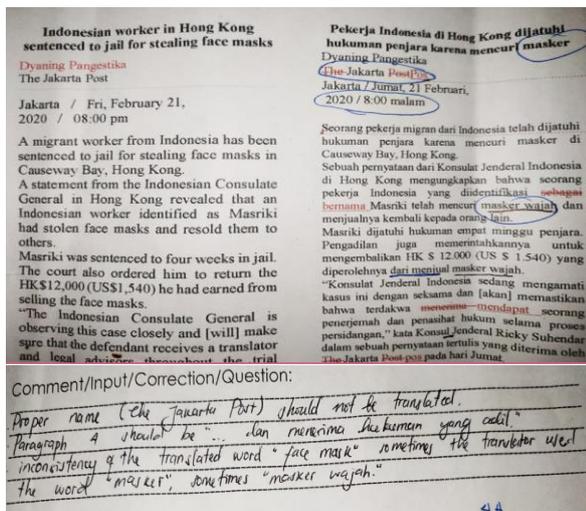


Figure 4 Student's inconsistency in using terminologies.

complete description of the source language and its target language is presented in Quotation [9].

Quotation [9]

“As of Friday, the photo, posted earlier this month, collected more than 34,000 “likes” from May’s 439,000-some followers.” “(Hingga hari Jumat, foto tersebut, yang diposting awal bulan ini, telah mengumpulkan lebih dari 34.000 “suka” dari 439.000 pengikut Mei.)” [Docs, S54] (Dorn, 2019).

3.3.2. Active and Passive Voices

Passive sentences in the Indonesian language could be represented in two forms of sentences, namely the sentences using affixes “ter—” and “di—” added to the words representing the predicate of the sentences (PUEBI, 2016). Although the two affixes could form active sentences, they had different meanings. In a post-editing process, student-3 (S3) should be able to correctly choose these affixes to make the passive sentences relevant to the intended meaning. As future post-editors, the students should also be trained on these particular lexico-grammatical patterns (Martikainen, 2019). These differences could lead to errors potentially made by the students in post-editing practice as indicated in Quotation [10].

Quotation [10]

Her purple nails spelled out the names “Kobe” and “Gigi”. (Kuku ungunya dituliskan nama “Kobe” dan “Gigi”). [Docs, S3] (Jamieson, 2020).

The MT output indicated its translation as “Kuku ungunya menyebutkan nama “Kobe” dan “Gigi”. Based on this context, the word “menyebutkan” was not appropriate, therefore, the student-3 (S3) as the post-editor was supposed to revise it, for example, into “Kuku ungunya dituliskan nama “Kobe” dan “Gigi”. However, the phrase “spelled out” which was translated by machine translation (MT) into “menyebutkan” and it was post-edited by S3 into “dituliskan”, but the two translations were not considered to be natural, instead it would be more appropriate when it was translated into “tertulis(kan)”, so the sentence would become “Kuku ungunya tertulis(kan) nama “Kobe” dan “Gigi”.

3.4. Knowledge-based performance

The students’ errors were found in the knowledge-based performance including failure in understanding information in more detail and failure in understanding textual contexts. Knowledge-based performance is usually represented by the ability of students to produce final translation results relevant with the messages in the source texts (Nitzke, 2019). Knowledge-based errors in this level of performance can be seen by the number of uses or selection of words irrelevant with the expected meaning in the source language. This kind of error occurred since the students failed to understand the context of the sentence and the language being translated.

Analysis of data in this study showed that the knowledge-based errors made by the students in the post-editing process were interpreted through its lexical meaning (Daems, Vandepitte, Hartsuiker, & Macken, 2017), while the true translation should try to understand the contextual meaning according to the time, events, socio-cultural background, and of course its contexts. Some descriptions of the knowledge-based errors encountered by the students include the failure in understanding information in more detail and failure in understanding textual contexts.

3.4.1. Failure in Understanding Information in More Detail

Detailed information in a message could be made in the form of definite information that could not be changed since the information was already related to the unity of a number. It could be in the form of a person’s name, date, name of the month, the year, the number of units referring to the exact counting units. Post-editing practice documents (Docs) showed that the students failed to understand the word “May” as the name of a person. It was translated by MT into “May” as the name of the month and it was not revised by S54. However, ways in writing a person’s name and the person’s age were different between English as the source language texts and Indonesian as the target language texts. In this case, S54 did the same thing in writing the person’s name and the person’s age between its source language and its target language. This phenomenon is identified in Quotation [11].

Quotation [11]

“People die here. People starve here. People go missing here. Ain’t nothing cool about that,” raged Auset Semhar, 28, among those ripping May, 27, for using “poverty as a prop.” (“Orang-orang mati di sini. Orang-orang kelaparan di sini. Orang-orang hilang di sini. Tidak ada yang keren tentang hal itu, “amuk Auset Semhar, 28, di antara mereka yang merobek May, 27, karena menggunakan” kemiskinan sebagai latar.”) [Docs, S54]. (Dorn, 2019).

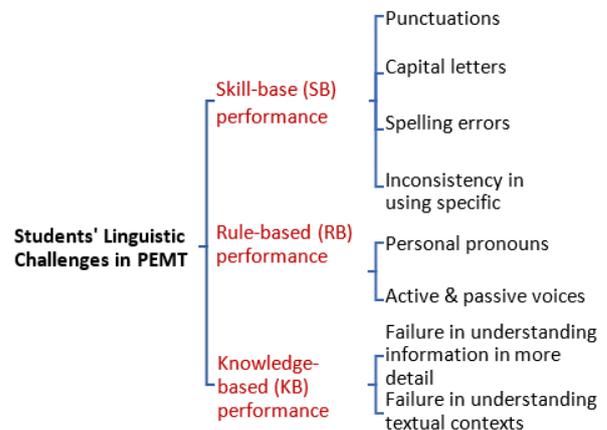


Figure 5 Students’ linguistic challenges in PEMT.

Based on quotation [11], “May, 27,” was translated by MT into “May 27” and it was corrected by S54 into “May, 27,”. “May” is the name of a person, followed by 27 as an indication of age. In the Indonesian language, the writing of the age is made in brackets. So, it would be written “May (27)” instead of “May, 27,”.

The other error was made by the student by omitting the article “The” in the publisher’s name of “The Jakarta Post”. In this case, the student intentionally omitted the article “The” which indicated that the student did not know that there was a newspaper publisher named “The Jakarta Post” so that the student assumed that the article “The” was an article that should be followed by the words “Jakarta Post”. Therefore, knowledge competence was needed by the students to appropriately revise the errors resulted from the MT outputs without making the other mistakes during the process of post-editing practices.

3.4.2. Failure in understanding textual contexts

Understanding the exact meaning of the source language requires a deep understanding of the context. An editor, in this particular context, was the student who practiced post-editing, should be sensitive and have a good understanding of the subject knowledge of the field (Neubert, 2000). Errors in understanding the context of the sentences or the context of the texts can lead to errors in using the right terminologies (Harto, Musthafa, & Setyarini, 2020). This error would make the meaning of the word ambiguous and give a new meaning. For example, data of students’ post-editing practices showed that there were some mistakes made by students in understanding the context of the sentences so that the edited sentences were still found to be inaccurate. In quotation [12], the phrase “his memorial service” was translated into “*layanan peringatannya*”. In this case, the student’s translation was considered to be inappropriate in using the word “*peringatan*” since the context of this sentence was remembering someone who passed away. Since it was about a memorial service, therefore, the word “*peringatan*” could be changed into “*pemakaman*”. So, a more appropriate translation would be “*layanan pemakaman*”. See the Quotation [12] for more details.

Quotation [12]

Beyonce Sang One of Kobe Bryant’s Favorite Songs during His memorial Service (Amber Jamieson). (*Beyonce Menyanyikan Salah Satu Lagu Favorit Kobe Bryant Selama Layanan Peringatannya*) (Amber Jamieson). [Docs, S46] (Jamieson, 2020).

Quotation [13]

The abbreviation of date units, for example, in the translation from English into Indonesian, its English initials “(d/m/y)” are referring to “(date/month/year)”, but in the Indonesian language, the abbreviation of date units “(tgl/bln/thn)” is referring to “(tanggal/bulan/tahun)” [Intv-S13].

Different from quotation [12], quotation [13] was an example of a student’s error in understanding an acronym. In this particular case, the student-13 (S13) did not carefully learn the differences between the English acronym and its Indonesian one. For instance, “(d/m/y)” referring to “(date/month/year)” and its Indonesian version was “(tgl/bln/thn)”. These differences were shared by S13 through an interview (Intv-S13) as it is presented in quotation [13]. Based on the quotation [13], the students as translators have to be aware of some differences, particularly in making acronyms both in the source language and in the target language so that they can make the right decision to write the relevant ways in writing them. Based on the aforementioned elaborations, the students’ linguistic challenges in post-editing of machine translation (PEMT) process, particularly in translating news item texts, are summarized in Figure 5.

4. CONCLUSION

Students have to build their translation competences and skills by practicing translation in various texts and in particular subjects based on their fields of interest. An investigation of students’ linguistic challenges in the translation of news item texts undertaken by the undergraduate students through post-editing of machine translation (PEMT), particularly in the English-Indonesian language pair, is made to provide the students with comprehensive descriptions of their linguistic problems resulted from their real translating activities. Due to some differences in the language structures between English and Indonesian, therefore, the students have to identify particular grammatical problems and mechanical errors resulted from the MT outputs and make relevant corrections and adjustments in order to achieve more accurate and acceptable translations.

REFERENCES

- Borobudur temple temporarily closes level nine, 10 to public. (2020, February 18). *The Jakarta Post*. <https://www.thejakartapost.com/travel/2020/02/18/borobudur-temple-temporarily-closes-level-nine-10-to-public.html#>.
- Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Braun, V., and Clarke, V. (2014). What can “thematic analysis” offer health and wellbeing researchers? *International Journal of Qualitative Studies on Health and Wellbeing*, 9(1), doi: 10.3402/qhw.v9.26152.
- Clarke, V., and Braun, V. (2018). Using thematic analysis in counselling and psychotherapy research: A critical reflection. *Counselling and Psychotherapy Research Journal*, 18(2), 107-110.

- Creswell, J. W. (2012). Collecting qualitative data. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research. Fourth ed.* Boston: Pearson, 204-35.
- Daems, J., Vandepitte, S., Hartsuiker, R.J., and Macken, L. (2017). Identifying the machine translation error types with the greatest impact on post-editing effort. *Frontiers in Psychology*, 8(1282), 1-15.
- Dorn, S. (2019, November 16). Poverty is not a prop: A controversy over Instagram shoot at Queensbridge houses. *New York Post*. <https://www.google.com/amp/s/nypost.com/2019/11/16/poverty-is-not-a-prop-controversy-over-instagram-shoot-at-queensbridge-houses/amp/>.
- Harto, S., Musthafa, B., & Setyarini, S. (2020, December). Investigating Distinctive Problems Observed in Post-editing of Machine Translation Output Made by Indonesian Undergraduate Students. In *4th International Conference on Language, Literature, Culture, and Education (ICOLLITE 2020)* (pp. 687-693). Atlantis Press.
- Huddleston, R., and Pullum, G.K. (2006). *A student's introduction to English grammar*. Cambridge: Cambridge University Press.
- Jamieson, A. (2020, February 24). Beyonce sang one of Kobe Bryant's favorite songs during his memorial service. *Buzz Feed News*. <https://www.google.com/amp/s/www.buzzfeednews.com/amphtml/amberjamieson/beyonce-kobe-gianna-bryant-memorial>.
- Jia, Y., Carl, M., and 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, 60-86.
- Kaumba, M. (2020, February 26). Delhi riots: City tense after Hindu-Muslim clashes leave 23 dead. *ZNBC News*. <https://www.znbc.co.zm/news/delhi-riots-city-tense-after-hindu-muslim-clashes-leave-23-dead/>
- Malik, R.S., and Hamied, F.A. (2016). *Research methods: A guide for first time researchers*. Bandung: UPI Press.
- Martikainen, H. (2019, September). Post-Editing neural MT in Medical LSP: Lexico-Grammatical patterns and distortion in the communication of specialized knowledge. In *Informatics* (Vol. 6, No. 3, p. 26). Multidisciplinary Digital Publishing Institute
- Massardo, I., Meer, J.V.D., O'Brien, S., Hollowood, F., Aranberri, N., and Drescher, K. (2016). *MT post-editing guidelines*. Amsterdam: TAUS Signature Editions.
- Merriam Webster Online Dictionary. Retrieved on 2 September 2021 from <https://www.merriam-webster.com/dictionary/sentence>.
- Neubert, A. (2000). Competence in language, in languages, and in translation. In C. Schaffner., and B. Adab (eds.). *Developing translation competence* (pp. 3-18.). Philadelphia: John Benjamins Publishing Company.
- Nitzke, J. (2019). *Problem solving activities in post-editing and translation from scratch: A multi-method study*. Berlin: Language Science Press.
- PUEBI. (2016). *Pedoman umum ejaan bahasa Indonesia (PUEBI) (4th ed)*. Jakarta: Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan dan Kebudayaan.
- Quah, C.K. (2006). *Translation technology*. New York: Palgrave Macmillan.
- Rasmussen, J. (1983). Skills, rules, and knowledge: Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics*, 13(3), 257-266.
- Regan, H. (2020, February 25). 21 of the world's 30 cities with the worst air pollution are in India. *CNN*. <https://www.google.com/amp/s/amp.cnn.com/cnn/2020/02/25/health/most-polluted-cities-india-pakistan-intl-hnk/index.html>
- Shoji, M. (2020, February 26). Hina matsuri (dolls festival): One of the year's most auspicious days dedicated to girls. *Japan Today*. [https://japantoday.com/category/features/lifestyle/Hina-Matsuri-\(Dolls-Festival\)-One-of-the-year%E2%80%99s-most-auspicious-days-dedicated-to-girls](https://japantoday.com/category/features/lifestyle/Hina-Matsuri-(Dolls-Festival)-One-of-the-year%E2%80%99s-most-auspicious-days-dedicated-to-girls)
- Sin-wai, C. (2016). *The future of translation technology: Towards a world without Babel*. Routledge.
- Sycz-Opon, J., and Galuskina, K. (2017). Machine translation in the hands of trainee translators: An empirical study. *Studies in Logic, Grammar and Rhetoric*, 49(62), 195-212.
- Vieira, L.N., Alonso, E., and Bywood, L. (2019). Introduction: Post-editing in practice - process, product and networks. *The Journal of Specialised Translation*, 31, 2-13.
- We have to dream big: Nadiem wants Indonesian to become Southeast Asia's common language. (2020, February 21). *The Jakarta Post*. <https://www.thejakartapost.com/news/2020/02/21/we-have-to-dream-big-nadiem-wants-indonesian->

[to-become-southeast-asias-common-language.html#](#)

Yamada, M. (2019). The impact of google neural machine translation on post-editing by student translators. *The Journal of Specialised Translation*, 31, 87-106.

Yin, R. K. (2018). *Case study research and applications*. Sage.

Zaretskaya, A., Vela, M., Pastor, G.C., and Seghiri, M. (2016). *Measuring post-editing time and effort for different types of machine translation errors*.