

Enhancing Machine-Translated Tourism Texts: Post-Editing Practices of Indonesian Students

Dewi Kesuma Nasution*, and Bambang Panca Syahputra

Faculty of Teacher Training and Education, Universitas Muhammadiyah Sumatera Utara,
Medan, Indonesia

*Corresponding Author: dewikesuma@umsu.ac.id

Received: 1st June 2024, Revised 13th June 2024, Accepted 30th June 2024

ABSTRACT

This study investigates practices among students majoring in the English Education Study program at Universitas Muhammadiyah Sumatera Utara Indonesia in post-editing machine-translated tourism texts. Utilizing a descriptive qualitative approach, we analyzed the post-editing practices of these students. The texts postedited were authentic Indonesian tourism texts sourced from lonely planet website. The findings revealed machine translation errors in three main areas: lexico-semantic errors, grammatical errors, and syntactic errors. Lexico-semantic errors often involved incorrect word choices and mistranslations of nuanced terms. Grammatical errors were prevalent in the misuse of tense, agreement, and article usage. Syntactic errors included improper sentence structure and word order, leading to awkward or unclear translations. As a result of these errors, it was observed that students did some corrections to polish the machine translated text, from the most dominant type of correction to the least dominant, including correcting lexical choices, correcting word form, correcting word order, omissions, additions and deleting. It is recommended that future studies involve more diverse participants with different text types so as to generate a more comprehensive finding.

Keywords: Post-editing, tourism translation, machine translation

1. INTRODUCTION

Machine Translation (MT) system, such as Google Translate, has significantly enhanced the speed and accessibility of translation services. However, despite their progress, this system often produces output that lacks the accuracy and nuance required for high-quality translations. Despite its drawbacks, MT has some advantages. With MT, learners can also devote more of their working memory to conceptualizing ideas and thinking of rhetorical devices rather than focusing on the literal translation of lexical items (Chon et al., 2021). This has led to the emergence of postediting—a process where human translators review and refine machine-generated translations to ensure they meet the necessary standards.

Post-editing (PE) has become a critical skill in the translation industry, blending technological proficiency with linguistic expertise. PE has begun to attract considerable attention due to the quality of raw MT output that may not completely represent the meaning reflected in the source text or the inaccuracy of the MT output based on the linguistic rules of the target language (Chung, 2020). For university students, especially those studying translation, mastering post-editing is essential. It equips them with the ability to improve machine translations and prepares them for the demands of the modern translation market. Despite its importance, there is limited research focusing on how students, particularly in Indonesia, approach post-editing and the challenges they face in mastering this skill.

This study aims to explore the post-editing practices of Indonesian university students. By analyzing their practices, the research seeks to provide a comprehensive understanding of the current state of post-editing education and practice in Indonesia. The findings are expected to contribute to the development of more effective training programs that can better equip students with the necessary skills to succeed in the translation industry.

2. LITERATURE REVIEW

Despite the significant quality development of machine translation (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 (Harto *et al.*, 2022). 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-Opoń & Gałuskińska, 2017; Yamada, 2022; Zaretskaya *et al.*, 2016).

Postediting of machine translation, which combines machine translation with human translation, is the process of reviewing and adapting raw machine translation output to finalize the target text in accordance with translation brief and established requirements (Yang & Mustafa, 2022). Allen (2003) mentions that the task of the posteditor is to edit, modify, and/or correct pretranslated text that has been processed by a machine translation system from a source language into (a) target language(s). The ISO 18587 standard for postediting of machine translation output, which was codified in 2017, interprets that postediting is performed on MT output for the purpose of checking its accuracy and comprehensibility, improving the text, making the text more readable, and correcting errors.

A number of scholars have attempted to investigate the MT errors and categorize them. Depraetere (2010) analyzed ten post edited texts done by translation students who are not trained on PE and stressed that MT errors should be included in training to avoid full dependence on MT output. In fact, post editing cannot be considered the same as proof reading as the errors in human translation are different from those in MT output. For instance, spelling and typing errors hardly ever occur in MT output; however, syntactic, and lexical ones are frequently observed in MT output not in human translation (Čulo & Nitzke, 2016).

Popović *et al.* (2014) investigated five types of operations: correcting word form, correcting word order, adding omissions, deleting, additions and correcting lexical choices. Their research provides a comprehensive framework for understanding the specific actions required to improve machine-translated texts. By categorizing these operations, Popović *et al.* offer a structured approach that can be systematically applied to various translation tasks. Correcting word form and order addresses fundamental linguistic errors, ensuring that translations adhere to grammatical conventions and syntactic norms. Adding omissions and deleting unnecessary additions enhance the clarity and conciseness of the text, making it more readable and accurate. Correcting lexical choices ensures that the translation captures the intended meaning and nuances of the original text, which is crucial for maintaining the integrity of the message. Therefore, their categorization of post-editing operations is essential for improving the overall quality of machine-translated texts and can serve as a guideline for both novice and experienced post-editors.

Another 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 (Harto et al., 2020).

Some strategies have been employed in post-editing different texts. Nino (2008), for example, revealed that when EFL students were asked to post-edit a raw MT output by consulting different online resources, they were found to make use of PE strategies such as rewriting, paraphrasing, self-correction, guessing, inferencing, reflecting, and using synonyms. In the meantime, Kliffer (2008) discovered that word choice remained the most frequent error category among the PE results. Even after PE, average frequency of total errors between PEd output and student translations was not significantly different. Similarly, to modify the MT sentences, the learners employed PE strategies such as deletion, paraphrase, and grammar correction (Shin & Chon, 2023).

3. METHOD OF THE RESEARCH

This research is descriptive qualitative research in which students' works of tourism text postediting is presented and described. The paper discusses what changes are made and why those changes are made. Additionally, errors in terms of grammatical errors, lexico-semantic errors and syntactic errors are tabulated and their frequency is calculated. By doing so, the research will uncover the dominant and least dominant errors found in Machine translation which were postedited by student translators.

3.1 Source of Data

The text of the translation was taken from the world's renown travel website lonelyplanet. The text was then randomly selected based on the student translator familiarity with the text. One text about Bali was finally adopted and used as the source of the data. Afterwards, the text was machine translated using Google Translate and student translator post-edited them.

3.2 Participants

This study involved a sample of 5 students from English Education Department, Universitas Muhammadiyah Sumatera Utara, Indonesia. These students were purposely selected under the specified criteria set, including having an A grade on translation and having an A grade for Grammar. These criteria limited the population so as to filter the post-editing results and avoid unnecessary edits from low achieving students which were excluded as the participants. As experience of participants in translation represents a significant extraneous factor in this research; as the more experienced or professional the translators are, the more likely the postediting task done is better; even if they do not have previous knowledge of post-editing (Ericsson 2003).

4. RESULTS AND DISCUSSIONS

The following chart illustrates errors committed by Google Translate:

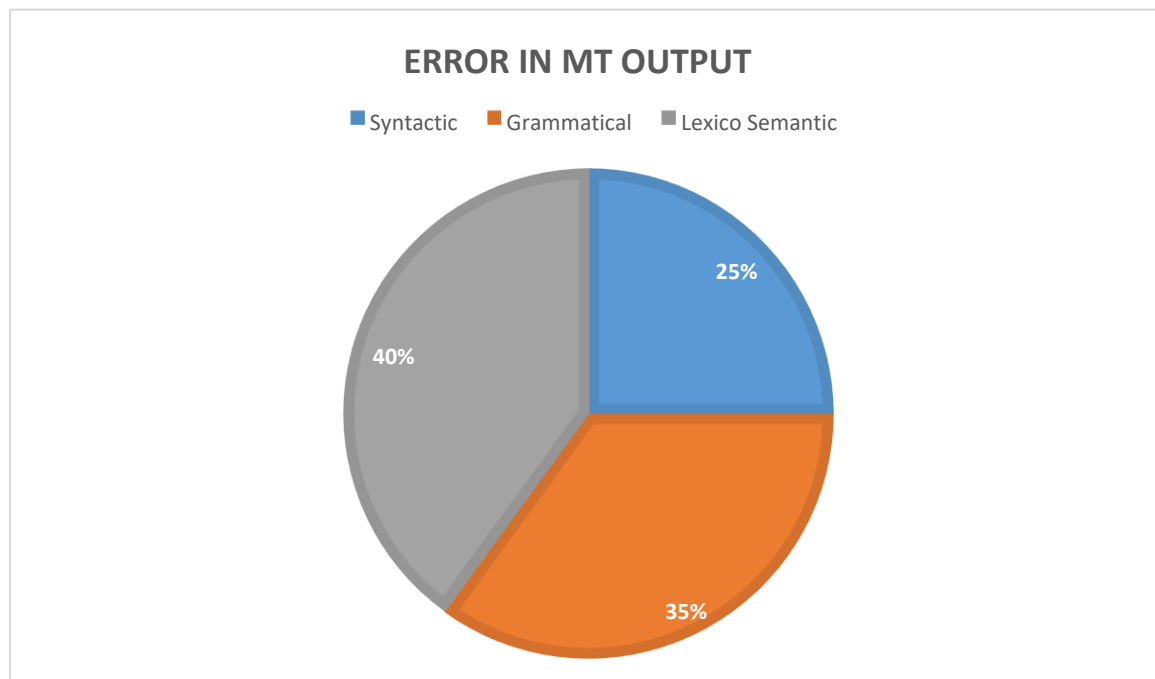


Figure 1. Types of Errors in MT output.

The above chart shows that lexico-semantic errors are the most frequent type of error in this data set, making up 40% of the errors. Grammatical errors are the second most frequent type of error, at 35%. Syntactic errors are the least frequent type of error, at 25%. Lexico semantic errors occur on the cultural texts. Translating text containing elements of a culture is not easy because different regions have different cultures (Supendi, 2017).

It was observed that students did some actions as proposed by Popović *et al.* (2014), including correcting word form, correcting word order, adding omissions, deleting, additions and correcting lexical choices. The following section elaborates on machine translation errors and students' post-editing activities on each sentence of the erroneous text. Code "S" represents "student" and "S1" represents "student number one".

Data 1

ST : If you only visit one museum in Ubud, make it this one

MT: Jika Anda hanya mengunjungi satu museum di Ubud, jadikanlah museum ini.

PE: Jika Anda hanya mengunjungi satu saja museum yang ada di Ubud, maka pilihlah museum ini

It was observed that the above machine translation failed to render the idiomatic expression "make it this one" into the target language. This is one of the drawbacks of MT that it cannot render culturally related expressions into the target language. S2 post-edited the phrase into "maka pilihlah museum ini" which is more local and suited to the target audience. We can see that students corrected a lexis "jadikanlah" into "pilihlah", and added a word "maka" to make it more acceptable.

Data 2

ST: Founder Agung Rai built his fortune selling Balinese artwork to foreigners in the 1970s, and during his time as a dealer he also built one of Indonesia's most impressive private collections of art.

MT: Pendirinya Agung Rai membangun kekayaannya dengan menjual karya seni Bali kepada orang asing pada tahun 1970an, dan selama menjadi pedagang ia juga membangun salah satu koleksi seni pribadi di Indonesia yang paling mengesankan

PE: Pendirinya yaitu Agung Rai telah membangun kekayaannya dengan menjual karya seni Bali kepada orang-orang asing pada tahun 1970an, dan selama menjadi pedagang ia juga membangun salah satu koleksi seni pribadi yang paling mengesankan di Indonesia.

It is observed that the above machine translated text maintains a sufficient level of fidelity to the source text. It accurately translates the key ideas: Founder Agung Rai, selling Balinese artwork, the 1970s timeframe, and the impressive art collection. However, it still needs some changes.

S1 post-edited made some stylistic changes. For example, S1 added "yaitu" to adds specificity, and "orang-orang asing" to emphasize plurality. The machine translation is relatively natural but could benefit from slight improvements for better flow. For instance, "pada tahun 1970an" could be more naturally written as "pada tahun 1970-an" with a hyphen, a minor typographic preference in Indonesian. The post-edited version enhances naturalness. S1 also added the word "telah" for the phrase "built "to provide a clearer verb tense that aligns well with typical Indonesian narrative style.

Moreover, the machine translation handles the cultural references adequately, such as "karya seni Bali" (Balinese artwork) and "koleksi seni pribadi" (private art collection). These terms are directly translated and are understandable in Indonesian. The post-edited translation retains these cultural references and slightly enhances clarity. The use of "yaitu" helps emphasize the founder's identity, which is important in an Indonesian cultural context where titles and roles are often highlighted for respect and clarity.

Data 3

ST: Exhibits include classical Kamasan paintings and Batuan-style work from the 1930s and '40s

MT: Pamerannya meliputi lukisan klasik Kamasan dan karya gaya Batuan dari tahun 1930an dan 40an

PE: Pameran tersebut berisi pameran lukisan klasik Kamasan dan karya gaya Batuan dari tahun 1930an dan 1940an

The results of machine translation above indicates that changes need to be made. For example, in terms of grammatical error, the phrase "the exhibit includes" was translated into "pamerannya meliputi" by machine translation. "Meliputi" is not a proper word in this context. Hence, S4 added the word "tersebut" and changed the lexis "meliputi" into "berisi" which is grammatically correct and is more natural in the target language and semantically clear in this context. Additionally, S4 removed unnecessary possessive "nya" as the use of suffix "-nya" is grammatically awkward.

Data 4

ST: The traditional art gallery is also home to a collection of works by expat artist Walter Spies (1895–1942), who played a significant role in the development of the Ubud painting school.

MT: Galeri seni tradisional ini juga merupakan rumah bagi koleksi karya seniman ekspatriat Walter Spies (1895–1942), yang berperan penting dalam perkembangan sekolah seni lukis Ubud.

PE: Galeri seni tradisional ini juga menyimpan koleksi karya seniman ekspatriat Walter Spies (1895-1942), yang memainkan peran penting dalam pengembangan sekolah seni lukis Ubud.

The lexico semantic error by MT occurred on the phrase "merupakan rumah bagi koleksi". Despite it is lexically correct but semantically slightly off for the context, the phrase "rumah bagi" (home to) is a literal translation that might sound awkward in the Indonesian target language and does not capture the idiomatic sense effectively. In terms of basic sentence structure, MT made no grammatical mistakes, but could benefit from more idiomatic phrasing. The phrase "rumah bagi" was post-edited by S5 into "menyimpan koleksi" where "menyimpan" is more natural and fit the Indonesian culture, thus improving the lexical correctness in the context. Another lexical change made was the phrase "berperan penting". This phrase is accurate and conveys the meaning of "played a significant role." However, S5 post-edited the phrase into "memainkan peran penting" which makes the phrase semantically richer, indicating active involvement and influence and better conveys the source text's nuance.

Data 5

ST: Enter the museum grounds from Kafe Arma on Jl Raya Pengosekan or around the corner at the ARMA Resort entrance.

MT: Masuki area museum dari Kafe Arma di Jl Raya Pengosekan atau di sudut pintu masuk ARMA Resort

PE: Masuki area museum dari Kafe Arma di Jl Raya Pengosekan atau di tikungan pintu masuk ARMA Resort.

It is observed that the phrase "di sudut" is a direct translation of "around the corner," in the above machine translated text. It does not capture the idiomatic meaning in Indonesian, where "sudut" usually refers to a sharp corner or the inside of a corner. Additionally, "di sudut" slightly disrupts the flow due to its less common usage in this context. The phrase "di sudut pintu masuk" is lexically correct but failed to fit for the context. This phrase was corrected by S2 into "di tikungan pintu masuk" which is lexically correct and is a better semantic fit for "around the corner," indicating a turn in the road rather than a sharp corner.

The post-edited translation (PE) shows a marked improvement over the machine translation (MT) by addressing and correcting grammatical, lexico-semantic, and syntactic errors. These changes highlight the effectiveness of post-editing in refining machine translations to ensure they are not only accurate but also contextually appropriate and easy to understand for the target audience.

Data 6

ST: This cultural compound opened in 1996 and displays his collection in two purpose-built gallery buildings

MT: Kompleks budaya ini dibuka pada tahun 1996 dan menampilkan koleksinya di dua bangunan galeri yang dibangun khusus

PE: Pada tahun 1996 dibuka kompleks budaya dengan menampilkan koleksinya di dua bangunan galeri khusus

It is observed that S4 made some changes on the structure of the machine translated text. While the syntactic errors occurred in MT, the syntax is correct but somewhat rigid since it follows the source text closely, which can make it less natural in Indonesian. By starting with "Pada tahun 1996 dibuka kompleks budaya," the sentence flows more naturally and is common for Indonesians to start with the year first when it comes to introducing when something is built.

S4 corrected the phrase “dan menampilkan” into “dengan menampilkan” to improve semantic clarity, linking the ideas more effectively. While at the Lexico-Semantic Errors, MT generates slightly awkward translation of “purpose-built” into “yang dibangun khusus”. At a glance, this machine translated text has normal readability level, however, it can be edited to create a more natural expression. Hence, S4 post-edited the phrase into “bangunan galeri khusus” which effectively conveys the idea of purpose-built galleries more naturally.

Additionally, the post-edited phrase “dengan menampilkan koleksinya di dua bangunan galeri khusus” instead of “dan menampilkan koleksinya di dua bangunan galeri yang dibangun khusus” provides a clearer connection between the clauses, enhancing readability.

The restructured sentence flows better and reads more naturally in Indonesian. It maintains clarity and coherence. The post-edited translation (PE) significantly improves upon the machine translation (MT) by addressing and refining grammatical, lexico-semantic, and syntactic issues. These improvements demonstrate the value of post-editing in enhancing the quality of machine translations to ensure they are accurate, clear, and natural for the target audience.

Data 7

ST: Stand-out works to seek out in the modern art gallery are Green Rice Paddies (1987) by Nasjah Djamin (1924–1999) and Wild Orchids (1988) by Widaya (1923–2002).

MT: Karya menonjol yang dapat dilihat di galeri seni modern adalah Sawah Hijau (1987) karya Nasjah Djamin (1924–1999) dan Anggrek Liar (1988) karya Widaya (1923–2002).

PE: Karya-karya penting yang ditampilkan dalam galeri seni kontemporer antara lain Sawa Hijau (1987) karya Nasja Jamin (1924-1999) dan Wild Angrek (1988) karya Widaya (1923-2002).

It can be observed that a several phrases are corrected by S5. For example, “karya menonjol yang dapat dilihat” was simplified for better readability. S5 post-edited the phrase into “Karyakarya penting”. In the meantime, the phrase “yang dapat dilihat” is a bit redundant and could be simplified. S5 post-edited the phrase into “yang ditampilkan” which sounds more natural and fit to the context of the text. In addition, S5 changed the lexis of “adalah” in the machine translated text into “antara lain” which makes the text more coherent and flows naturally. Overall, data number 7 which was post-edited by student above revealed that S5 made dominant changes of lexis.

5. DISCUSSIONS

The machine translation, in the present research, captures the general nuances of the source text but lacks some of the subtleties and cultural nuances in particular. The post-edited translation (PE) demonstrates a more refined application of fidelity, naturalness, handling of cultural references, and conveyance of nuances compared to the machine translation (MT). Lexical errors can introduce complex or obscure words that are not aligned with this goal. Similarly, complex grammatical structures introduced by errors can create unnecessary challenges for understanding.

The post-editing process carried out by students in the present research improves the machine translation by addressing grammatical, lexico-semantic, and syntactic errors. The post-edited translation (PE) corrects the possessive construction and uses more precise terminology, resulting in a more natural and comprehensible sentence. The present research findings align with the findings (Malang, 2021) which discovered that google translate produces lexical errors dominantly compared to other error types.

Thus, post-editing effort in this particular error is highly needed and crucial. This highlights the importance of choosing the most appropriate words in order to produce high-quality translations that accurately reflect the source text's meaning and are appropriate for the target language audience.

With regard to tourism language, it often uses a simplified vocabulary and sentence structure to cater to a wider audience with varying language proficiency. Tourism text often presents information without a specific author's voice. Syntactic errors caused by machine translation in this research can disrupt this flow by introducing unnecessary emphasis or awkward phrasing. The post-editing process focused on improving the naturalness and readability of the translation. This involved adding clarifying terms, adjusting prepositions, including additional context for better understanding, and changing the tone to be more inclusive and engaging. The edits also aimed to maintain a formal and polite style suitable for an informative and descriptive text about a museum.

6. CONCLUSIONS

In conclusion, machine translation errors can significantly disrupt the core characteristics of tourism language. By mitigating these errors through high-quality tools, human review, and clear writing, tourism businesses can ensure their message resonates effectively with international travelers. Qualified post-editors with the target language culture understanding should be involved in post-editing machine translated texts. In addition, changes to lexis and changes to the sentence structure are among top actions that were performed by student post-editor in the present research which can be adopted by other post-editors when it comes to post-editing tourism related materials. It is recommended that future research involve a more diverse participants so that a more varied results can be obtained from the post-editing practice among Indonesian university students.

REFERENCES

- Allen, Jeffrey H. 2003. "Post-editing", in: Harold Somers (ed.), *Computers and Translation: A translator's guide*. Benjamins Translation Library, Amsterdam: John Benjamins Publishing Company, 297–317. (<https://benjamins.com/catalog/btl.35.19all>).
- Chon, Y. V., Shin, D., & Kim, G. E. (2021). Comparing L2 learners' writing against parallel machine-translated texts: Raters' assessment, linguistic complexity and errors. *System*, 96, 102408. <https://doi.org/10.1016/j.system.2020.102408>.
- Chung, E. S. (2020). The effect of L2 proficiency on post-editing machine translated texts. *Journal of Asia TEFL*, 17(1), 182–193. <http://dx.doi.org/10.18823/asiatefl.2020.17.1.11.182>.
- Čulo, O., & Nitzke, J. (2016). Patterns of terminological variation in post-editing and of cognate use in machine translation in contrast to human translation. *Proceedings of the 19th Annual Conference of the European Association for Machine Translation, EAMT 2016*, 4(2), 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(AUG), 1–15. <https://doi.org/10.3389/fpsyg.2017.01282>.
- Depraetere, I. (2010). What counts as useful advice in a university post-editing training context? Report on a case study. *EAMT 2010 - 14th Annual Conference of the European Association for Machine Translation*, May.
- Harto, S., Hamied, F. A., Musthafa, B., & Setyarini, S. (2022). Students' Linguistic Challenges in Post-Editing of Machine Translation. *Proceedings of the Fifth International Conference on Language, Literature, Culture, and Education (ICOLLITE 2021)*, 595(November). <https://doi.org/10.2991/assehr.k.211119.099>.

- Harto, S., Musthafa, B., & Setyarini, S. (2020). *Investigating Distinctive Problems Observed in Postediting of Machine Translation Output Made by Indonesian Undergraduate Students*. 509(Icollite). <https://doi.org/10.2991/assehr.k.201215.108>.
- Kliffer, M. D. (2008). Post-editing machine translation as an FSL exercise. *Porta Linguarum*, 9, 53–67. <https://doi.org/10.30827/Digibug.31745>.
- Malang, M. I. (2021). *Common lexical errors made by machine translation*. 8(1).
- Napitupulu, S. (2017). Analyzing Indonesian- English abstracts translation in view of translation errors by google translate. *International Journal of English Language and Linguistics Research*, 5(2), 15-23s.
- Niño, A. (2008). Evaluating the use of machine translation post-editing in the foreign language class. *Computer Assisted Language Learning*, 21(1), 29–49. <https://doi.org/10.1080/09588220701865482>.
- Popović, M, et. al. (2014). Relations between different types of post-editing operations, cognitive effort and temporal effort. In *Proceedings of the 17th Annual Conference of the European Association for Machine Translation*, pages 191–198, Dubrovnik, Croatia. European Association for Machine Translation.
- Shin, D., & Chon, Y. V. (2023). *Second language learners' post-editing strategies for machine translation errors*. 27(1), 1–25.
- Sycz-Opoń, J., & Gałuskińska, K. (2017). Machine translation in the hands of trainee translators-an empirical study. *Studies in Logic, Grammar and Rhetoric*, 49(1), 195–212. <https://doi.org/10.1515/slgr-2017-0012>.
- Yamada, M. (2019). The impact of google neural machine translation on post-editing by student translators. *Journal of Specialised Translation*, 31, 87–106.
- Yang, Z., & Mustafa, H. R. (2022). On Postediting of Machine Translation and Workflow for Undergraduate Translation Program in China. *Human Behavior and Emerging Technologies*, 2022, 1–11. <https://doi.org/10.1155/2022/5793054>.
- Zaretskaya, A., Vela, M., Pastor, G. C., & Seghiri, M. (2016). Measuring post-editing time and effort for different types of machine translation errors. *New Voices in Translation Studies*, 15(September), 63–92.