

EVALUATING THE STRENGTHS AND LIMITATIONS OF CHAT GPT-GENERATED TRANSLATIONS IN ACADEMIC POST-EDITING WORKFLOWS: A CASE STUDY OF ENGLISH-INDONESIAN ACADEMIC TEXTS

Christanta Rejuna Phanes Sembiring Brahmama, Rudy Sofyan, Umar Mono
Universitas Sumatera Utara, Medan, Indonesia

Received: 2025-01-16

Accepted: 2025-04-07

Published: 2025-06-26

Abstract

This study investigates ChatGPT's strengths and limitations in academic post-editing, particularly in English-Indonesian translations. Utilizing a qualitative case study, we explore how ChatGPT enhances translation efficiency while tackling cultural and contextual accuracy challenges. The study employs comparative data analysis, involving a professional Indonesian-English translator who post-edited ChatGPT-generated translations of academic abstracts. Data collection methods included keystroke logging, screen recording, and structured participant feedback, providing empirical insights into ChatGPT's post-editing strengths and limitations. The findings reveal that ChatGPT-generated translations demonstrate strengths in grammatical accuracy, structural coherence, and workflow efficiency. The post-editor reported a significant reduction in post-editing time compared to traditional machine translation outputs. However, the study also identifies limitations in cultural adaptation, idiomatic expressions, and domain-specific terminology, necessitating considerable human intervention to ensure academic quality. The comparative analysis highlights ChatGPT's potential while emphasizing the essential role of human expertise in refining AI-assisted translations.

Keywords: *academic text translation; artificial intelligence; chatGPT-generated translations; post-editing process; strengths and limitations*

1. Introduction

The translation industry has experienced significant changes due to rapid technological advancements, especially with the emergence of machine translation (MT) systems and artificial intelligence (AI) tools. These developments have significantly impacted the translation process, shifting the role of human translators from primary content creators to collaborative editors who work in tandem with AI systems. Rebeca & Stiegelbauer (2013) note that automatic translation tools, such as Google Translate, have enhanced cross-linguistic communication, thereby diminishing the immediate necessity for learning foreign languages in informal and business communications. Nonetheless, the study underscores that MT cannot substitute human skill, as it often falters in understanding context, grammatical nuances, and cultural subtleties. Although AI-driven tools can improve efficiency, they cannot deeply analyze meanings, especially in complex academic or

specialized contexts. Additionally, the research indicates that dependence on MT might deter people from pursuing foreign language studies, which raises concerns regarding future language proficiency. Despite these challenges, AI-powered translation continues to progress, playing a more vital role in contemporary translation processes, while still necessitating human post-editing to maintain accuracy and contextual relevance.

ChatGPT, an innovative AI model by OpenAI, represents a substantial progression in AI-assisted translation. As a sophisticated language model, ChatGPT has proven capable of delivering translations that are coherent, contextually relevant, and stylistically appropriate (Jiao et al., 2023; Liu et al. (2023)). However, most existing research focuses on general MT outputs, leaving a gap in our understanding of ChatGPT's role in academic post-editing workflows, particularly for English-Indonesian texts. Given the high linguistic precision, structural coherence, and specificity required in academic texts, post-editing is essential to ensure that AI-generated translations meet academic standards. While ChatGPT provides a solid basis for translation tasks, its effectiveness in maintaining academic integrity and cultural nuances is still largely unexplored.

Despite its potential, there are significant research gaps in assessing the efficiency, accuracy, and contextual reliability of translations produced by ChatGPT in academic post-editing. Although prior studies have examined various AI translation tools, there has been little focus on ChatGPT's application in English-Indonesian academic post-editing contexts. This research aims to address this gap by analyzing the pros and cons of translations generated by ChatGPT in academic post-editing, specifically concentrating on English-Indonesian academic texts that pose unique linguistic and cultural challenges.

Previous studies have investigated AI translation tools in academic scenarios. Atarchi, Elemari, & Marouane (2024) explored how AI-driven translation tools equipped doctoral students in Morocco to overcome linguistic barriers when translating research articles from French to English. Their findings highlighted AI's potential for improving academic writing but also noted challenges like linguistic bias and the need for post-editing. Similarly, Mukti et al. (2024) evaluated how effective AI translation tools are among English literature students at UIN Sunan Gunung Djati Bandung, discovering that while AI tools greatly assist academic translation, dependence on them might hinder language skill development. This underscores the necessity of human oversight to ensure contextual accuracy.

Further studies have looked into user preferences and acceptance of AI tools to understand their effect on academic translation processes. Hidayati & Nihayah (2024) found that students outside the English department preferred AI tools like ChatGPT over traditional options like Google Translate for translating academic texts. Their results highlight ChatGPT's ability to grasp and express complex ideas, which is crucial in academic contexts. Ye (2024) expanded on the challenges and suggested improvements for ChatGPT, noting its advanced capabilities for handling complex academic language. Ye concluded that while ChatGPT holds promise, further improvements are necessary to maximize its effectiveness in academic environments where precision and clarity are critical.

Investigating user acceptance further, Wang, Xu, & Liu (2024) utilized the Unified Theory of Acceptance and Use of Technology (UTAUT) to analyze student perceptions of ChatGPT for academic translation tasks. Their results suggest that both performance expectations and social influence significantly impact ChatGPT adoption, indicating that performance quality is vital for acceptance. Similarly, Salloum et al. (2024) applied an Extended Technology Acceptance Model (TAM) to identify factors affecting ChatGPT's adoption in translation practices, revealing that perceived ease of use and usefulness are

critical factors in its acceptance.

This study examines the strengths and limitations of ChatGPT in generating grammatically accurate and structurally coherent English-Indonesian academic translations, particularly noting shortcomings in cultural nuances, idioms, and discipline-specific terminology. By analyzing the post-editing efforts of a professional translator through keystroke logging, screen recording, and structured feedback, this research seeks to ascertain the extent of human intervention needed for ChatGPT-generated translations to align with academic quality standards.

To accomplish these aims, the study employs a descriptive qualitative case study approach (Yin, 2003), featuring a professional Indonesian-English translator who post-edits ChatGPT-generated translations of academic abstracts. The research integrates comparative data analysis methods, including keystroke logging, screen recording, and participant feedback, to evaluate ChatGPT’s performance in English-Indonesian academic post-editing workflows. By exploring these elements, the study aspires to provide an in-depth assessment of ChatGPT’s role in English-Indonesian academic post-editing, enhancing our understanding of AI-assisted translation within academic contexts and illuminating the capabilities and limitations of AI-driven translation tools in scholarly communication.

2. Literature Review

Translation Studies is a comprehensive academic discipline combining theoretical perspectives with practical applications. Holmes’s Map of Translation Studies, expanded by Toury and analyzed by Munday (2008), provides a foundational framework for understanding the discipline.

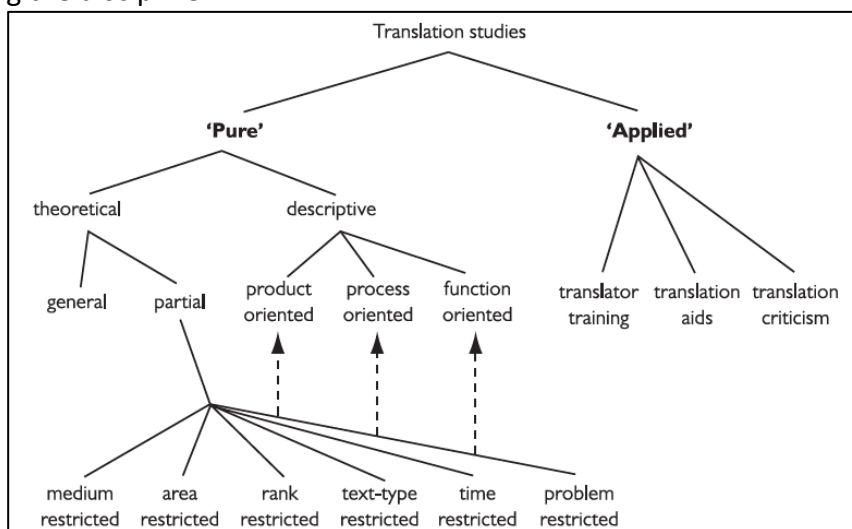


Figure 2.1. "Holmes's Map" of Translation Studies (from Toury, 1995, p. 10) cited from (Munday, 2008, p. 12)

The Pure branch of Translation Studies examines theoretical concepts like equivalence dynamics, translation universals, and the socio-cultural context of texts. These topics aim to identify universal principles that govern translation practices. Conversely, the Applied branch deals with practical translation elements, such as creating translation tools, improving training methods, and assessing translation quality. These branches emphasize that translation is both an intellectual endeavor and a professional skill, highlighting its significance in facilitating cross-cultural communication and knowledge sharing.

Post-editing has become essential to contemporary translation workflows, bridging the divide between machine-generated translations and high-quality human outputs. Vieira, Alonso, & Bywood (2019) emphasize that post-editing is vital for ensuring accuracy and aligning translations with cultural and stylistic needs. There are two main types of post-editing: light post-editing, which aims to make machine translations understandable, and full post-editing, which seeks to enhance machine outputs to match human translation quality. The rise of AI-assisted translation tools, like ChatGPT, has notably changed post-editing practices, providing coherent drafts that allow human translators to focus on higher-level improvements, such as cultural adaptation, fluency, and specific domain accuracy.

The emergence of machine translation (MT) tools has dramatically affected translation practices by enhancing speed and accessibility. However, both Google Translate and DeepL still encounter issues related to accuracy, fluency, and adaptation of domain-specific terminology (Brahmana, Sofyan, & Putri, 2020; Kamaluddin et al., 2024). A study by Habeeb (2020) explored Google Translate's effectiveness among Iraqi students, revealing that while widely accessible and cost-efficient, the tool often falls short in grammatical precision and lacks proofreading features, making it less suitable for critical academic translation tasks. Students generally viewed Google Translate positively but recognized its limitations in dealing with complex academic content.

Similarly, Brahmana, Sofyan, & Putri (2020) identified weaknesses in Google Translate's performance in academic contexts, noting that 31% of translation errors stemmed from misinterpretations, resulting in incorrect word choices. Furthermore, 30% of errors were associated with grammatical mistakes and structural issues, further questioning Google Translate's reliability for academic work. These insights illustrate that while Google Translate can serve everyday translation needs, it is inadequate for scholarly and professional tasks where precision, coherence, and contextual consideration are paramount.

While Google Translate is celebrated for its convenience and speed, DeepL is praised for its superior accuracy in interpreting complex sentence structures, idiomatic expressions, and specialized terminology (Kamaluddin et al., 2024). DeepL's neural network framework and data augmentation strategies enable it to produce more natural and context-sensitive translations than Google Translate. Kamaluddin et al. (2024) observe that DeepL consistently surpasses other MT tools in BLEU scores and human assessments, particularly in academic and professional settings. Despite its improvements, DeepL still necessitates post-editing to ensure that translations meet the specific requirements of scholarly discourse.

ChatGPT presents a more sophisticated translation approach than traditional MT tools, employing a transformer-based model and interactive features. Rather than translating text piecemeal, ChatGPT analyzes entire contexts, yielding more coherent, fluent, and contextually refined academic translations (Gill & Kaur, 2023). ChatGPT's flexibility, which allows real-time refinement of translations, marks a significant advancement in post-editing workflows—a capacity that Google Translate and DeepL lack. Nonetheless, ChatGPT struggles with idiomatic phrases, cultural subtleties, and discipline-specific terminology, emphasizing the need for human intervention in academic post-editing to maintain linguistic and contextual integrity.

The integration of AI tools into academic translation processes has undergone extensive examination. Hidayati & Nihayah (2024) found that ChatGPT outperformed traditional MT tools in terms of clarity and contextual understanding, making it a favorable option for translating complex academic texts. Similarly, Ye (2024) acknowledged ChatGPT's sophisticated linguistic skills while emphasizing further improvements to enhance its

effectiveness in academic translation endeavors. Additionally, Gao, Lin, Zhao, & Cai (2024) demonstrated that ChatGPT preserves stylistic consistency and fidelity in translating literature and academic texts, often outdoing conventional MT systems in nuanced scenarios.

Beyond performance evaluations, ethical considerations are also critical in AI-assisted academic translation. Ye (2024) raises concerns about potential plagiarism and authorship ambiguity, particularly when AI-generated translations are used without proper attribution. Furthermore, biases in AI-driven translations remain a significant issue, as these models are trained on extensive but often skewed datasets, which can compromise translation accuracy and cultural representation (Hidayati & Nihayah, 2024). Ensuring transparency in AI applications and adhering to ethical standards is crucial for maintaining academic integrity in AI-assisted translation.

As AI-driven translation tools evolve, the collaboration between human translators and AI becomes increasingly vital. While ChatGPT improves efficiency and delivers context-aware translations, human input remains crucial for fine-tuning stylistic, cultural, and disciplinary subtleties. The ongoing interaction between AI and human expertise emphasizes the importance of a hybrid translation model that leverages both strengths to achieve outstanding quality in academic post-editing workflows.

3. Research Method

The research adopts an intrinsic case study approach, as described by Berg & Lune (2017), to analyze the strengths and limitations of ChatGPT-generated translations in academic post-editing workflows, specifically in the context of English-Indonesian academic texts. This approach enables a detailed exploration of how ChatGPT supports human translators in achieving linguistic accuracy, contextual relevance, and overall translation quality. A single professional translator was selected as the participant to conduct an in-depth qualitative analysis. While a larger sample could provide more generalizable findings, this study focuses on a highly experienced translator, ensuring a thorough and expert-level assessment of ChatGPT's capabilities in English-Indonesian academic post-editing workflows.

This approach aligns with Stake (1995), who asserts that intrinsic case studies are particularly valuable when the goal is to gain deep insights into a specific case, rather than pursuing broad statistical generalizations. The chosen participant's extensive experience in academic translation and post-editing, combined with their educational background in Artificial Intelligence (AI), further strengthens the credibility of this study. Their dual expertise in AI and translation allows them to critically evaluate ChatGPT's impact on English-Indonesian academic texts from a linguistic perspective and through an AI-driven analytical lens. This ensures that their assessments are not merely subjective opinions but are based on scientific reasoning and technological understanding. Consequently, the findings of this study contribute meaningful insights into the strengths and limitations of ChatGPT-assisted post-editing, emphasizing its role in improving translation workflows while addressing its existing challenges in handling academic texts.

3.1 Methodological Design

The research adopts an intrinsic case study approach described by Berg & Lune (2017). This method emphasizes the distinct attributes of a particular case—in this case, the role of ChatGPT in academic post-editing workflows. Such a design facilitates an in-depth examination of the translation process, highlighting how ChatGPT aids human

translators in ensuring linguistic precision and contextual significance.

A professional English-Indonesian translator was chosen as the participant to ensure a thorough analysis. This translator worked as the post-editor, refining translations generated by ChatGPT. By concentrating exclusively on the post-editors' engagement with ChatGPT, the study seeks to deliver detailed insights into the translator's strategies, cognitive processes, and the strengths and weaknesses of ChatGPT as an academic post-editing tool.

3.2 Participant

This study involved a skilled professional translator who is a native speaker of Indonesian, specializing in English-Indonesian academic translation and post-editing. The participant is actively engaged with the Association of Indonesian Translators (HPI), demonstrating their commitment to upholding professional standards and ethical practices in translation.

The post-editor has earned a master's degree in Data Science and Artificial Intelligence, equipping them to evaluate ChatGPT-generated translations from both linguistic and technological perspectives. Their blend of AI and translation expertise enables assessments based on scientific reasoning rather than personal opinions. Additionally, the participant scored 108 on the TOEFL iBT, reflecting a strong command of English. With four to six years of experience in academic translation and post-editing, they have developed advanced abilities in refining AI-assisted translations, making them an excellent fit for this study.

As part of the research, the post-editor worked on improving an Indonesian translation of an English academic abstract produced by ChatGPT. This effort aimed to assess ChatGPT's ability to maintain accuracy, fluency, and contextual relevance in academic post-editing tasks. The participant's insights were vital for identifying both the strengths and weaknesses of translations generated by ChatGPT, providing a comprehensive view of AI-assisted post-editing in line with academic and professional standards.

To enhance the methodological rigor of this study, ethical considerations were thoroughly addressed. The participant gave informed consent for using keystroke logging and screen recording for research purposes. To protect confidentiality, all personally identifiable information was anonymized, and the study followed ethical guidelines outlined by Creswell (2011) and the Association of Indonesian Translators (HPI) concerning the ethical employment of AI in translation research.

3.3 Text Selection

In this study, the text selection process aimed to ensure that the materials chosen for translation were relevant and representative of academic writing. This meticulous selection was vital for maintaining the validity of the results when assessing ChatGPT's effectiveness in the post-editing phase. A range of criteria and analytical tools were utilized to identify and choose appropriate texts. Following Sofyan (2016), who highlighted the need for analytical tools to find suitable examples for translation, this study employed various text analysis methods. Among these were the Simple Measure of Gobbledygook (SMOG) and the Flesch-Kincaid readability tests. The SMOG test estimates the educational level needed to fully comprehend a text, while the Flesch-Kincaid test evaluates text readability based on word frequency and sentence length. These tools offered valuable insights into the complexity and

appropriateness of the selected text study.

To ensure consistency and relevance, the research focused on selecting abstracts from scholarly articles written by native English speakers across various fields. These abstracts were sourced from ScienceDirect, a respected database of peer-reviewed journals. Initially, five abstracts were selected and evaluated using the readability tools discussed earlier, as shown in the table below.

Readability Test	Text 1 (no abstract)	Text 2	Text 3	Text 4 (no abstract)	Text 5
Number of Words	241	251	201	344	99
Flesch Reading Ease Score	45.47	22.42	-	10.34	-
SMOG Readability	14.12	15.50	20.80	20.17	21.79

Table 3.3.1 Readability test level from Text 1 until Text 5

Score	School Level	Notes
100.00-90.00	5th grade	Very easy to read
90.0-80.0	6th grade	Easy to read
80.0-70.0	7th grade	Fairly easy to read
70.0-60.0	8th & 9th grade	Plain English
60.0-50.0	10th to 12th grade	Fairly difficult to read
50.0-30.0	College	Difficult to read
30.0-10.0	College Graduate	Very difficult to read
10.0-0	Professional	Extremely difficult to read

Table 3.3.2 Flesch-Kincaid Reading Ease Scores table
(Source from: <https://charactercalculator.com/flesch-reading-ease/>)

Score	Education Level
4.9 or lower	Elementary school
5 - 8.9	Middle school
9 - 12.9	High school
13 - 16.9	Undergraduate
17 or higher	Graduate

Table 3.3.3 SMOG Readability Scores table
(Source from: <https://charactercalculator.com/smog-readability/>)

Out of the five samples, Text 2 was selected for this study based on its readability test results. This choice met three crucial criteria to ensure both academic relevance and methodological integrity:

1. The text needed to be an abstract from a scholarly article, ensuring alignment with

- academic writing style and structure.
2. The text had to obtain valid scores in both the Flesch Reading Ease and SMOG Readability tests, ensuring suitable complexity for translation analysis.
 3. Abstracts lacking complete readability scores or containing missing sections were excluded to uphold data consistency and reliability.

By meeting these rigorous selection criteria, Sample Text 2 was identified as the most appropriate for translation and post-editing. This decision enabled the study to effectively examine ChatGPT's strengths and weaknesses in academic translation processes, laying a solid foundation for evaluating the performance of AI-assisted post-editing.

3.4 Research Instruments

This study thoroughly assessed the advantages and drawbacks of using ChatGPT-generated translations in academic post-editing workflows, with a specific focus on English-Indonesian academic texts, utilizing various research tools. The selected instruments included OBS Studio for screen recording, Translog-II for keystroke logging, and a structured feedback questionnaire, chosen to gather both objective data from behavioral tracking and subjective insights through participant feedback regarding ChatGPT's role in English-Indonesian academic post-editing.

OBS Studio enables the real-time recording of participants' screen activities, facilitating the analysis of editing behaviors, usage of external resources, and correction trends. This tool provided a visual overview of the post-editing process, highlighting key areas that required significant adjustments to the translations generated by ChatGPT.

Translog-II, a well-known keystroke logging software, tracked typing activities, pauses, and editing behaviors, yielding insights into cognitive processing, decision-making approaches, and challenges faced during post-editing. This tool helped researchers assess the effort required to enhance the accuracy, fluency, and contextual relevance of ChatGPT-generated translations in academic texts.

Additionally, a structured feedback questionnaire was distributed after the post-editing session to gather qualitative insights into participants' views on ChatGPT's translation precision, contextual relevance, and workflow efficiency. This questionnaire, developed based on previous AI-assisted translation studies, was designed to reflect user experiences and challenges faced during academic post-editing effectively.

To ensure the reliability and validity of the research instruments, this study referred to prior empirical applications of both Translog-II and OBS Studio in translation process research. Translog-II has been widely utilized in translation studies to log user activity data, offering precise insights into translators' cognitive behaviors through keystroke and pause analysis (Carl, 2012). Its reliability in capturing nuanced translation processes—such as drafting, revision, and editing—has been demonstrated across various studies of translation processes. Complementing this, OBS Studio has also proven to be a valuable tool for screen-based observation. Sinulingga, Sofyan, & Mono. (2023) employed OBS Studio alongside Translog-II to record the translation behaviors of professional translators, illustrating its effectiveness in capturing real-time activities such as information searches and decision-making pauses. By combining these validated tools, the present study strengthens its

methodological rigor in assessing ChatGPT's role in academic translation workflows, particularly in identifying post-editing strengths and limitations across multiple linguistic dimensions.

3.5 Data Collection Procedures

The data collection process aimed to evaluate the strengths and weaknesses of ChatGPT-generated translations in post-editing workflows for English-Indonesian academic texts thoroughly. This involved three main phases: the post-editing session, the feedback reflection and questionnaire, and the data analysis phase.

In the post-editing session, the participant edited an Indonesian translation of an English academic abstract produced by ChatGPT. This step assessed ChatGPT's ability to maintain linguistic accuracy, contextual relevance, and adherence to academic writing conventions. Throughout this process, Translog-II recorded keystrokes, capturing real-time decision-making, revision strategies, and cognitive effort as the translations were refined. Additionally, OBS Studio documented screen interactions, including the use of external resources, correction strategies, and areas necessitating significant human intervention. These recordings offered valuable insights into the participant's workflow efficiency and challenges faced during post-editing, analyzing the degree to which AI-generated translations require human refinement.

After the post-editing session, the participant completed a structured feedback questionnaire reflecting on their experience. This phase gathered qualitative insights on ChatGPT's translation accuracy, fluency, and contextual appropriateness in academic texts. The responses helped identify common strengths and weaknesses in ChatGPT's translations, highlighting specific linguistic or contextual areas needing human input. This phase provided a critical assessment of ChatGPT's role in English-Indonesian academic post-editing workflows, aligning with the study's goal to evaluate AI's impact on professional translation practices.

During the data analysis phase, a thematic analysis was conducted on the qualitative responses, revealing patterns in editing behavior and interactions between humans and AI. Observational data from Translog-II and OBS Studio were also reviewed to identify key post-editing challenges, workflow efficiency, and AI-related translation errors. By integrating keystroke logging, screen recordings, and participant feedback, this study provided a multi-faceted and methodologically robust analysis of ChatGPT's strengths and limitations in post-editing English-Indonesian academic texts. The findings contribute to the broader discussion on AI-assisted translation, underscoring the crucial role of human expertise in refining AI-generated translations to meet academic and professional standards.

4. Results and Discussion

This study provides in-depth insights into the strengths and limitations of utilizing ChatGPT-generated translations in academic post-editing tasks. The results suggest that ChatGPT is a valuable tool for improving translation speed and enhancing grammatical accuracy. Its ability to produce well-structured and polished initial drafts significantly reduces the time and effort required for minor corrections. However, the findings also highlight that ChatGPT faces challenges in addressing cultural nuances and preserving semantic integrity, both of which are crucial in academic translations. These limitations highlight the crucial role of human expertise in refining outputs to maintain quality

standards.

The data collected from questionnaire responses and feedback from a single post-editor participant offers a comprehensive overview of ChatGPT's strengths. While it excels in grammar correction and maintaining an academic tone, it struggles to navigate complex cultural contexts and idiomatic expressions, particularly in less-represented languages like Indonesian.

4.1 Strengths of ChatGPT-Generated Translations in Post-Editing

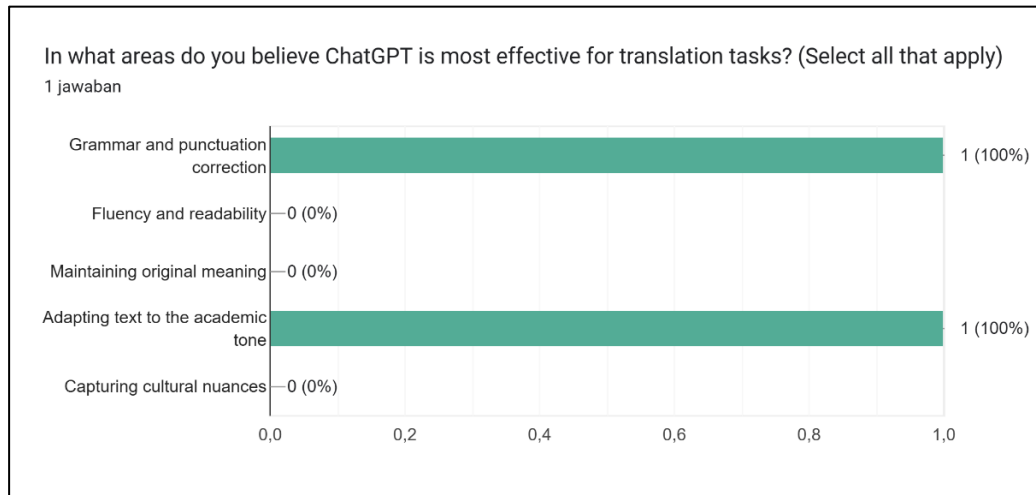


Figure 4.1.1. The Key Areas of ChatGPT Capabilities

Figure 4.1.1 shows that the post-editor identified grammar and punctuation correction and adapting to academic tone as ChatGPT's most effective capabilities (100%). This aligns with previous findings by Gill & Kaur (2023) and Hidayati & Nihayah (2024), who reported that ChatGPT can produce syntactically sound drafts suitable for formal contexts. However, the absence of effectiveness in other areas—such as capturing cultural nuances—reinforces the need for human oversight in sensitive translation tasks.

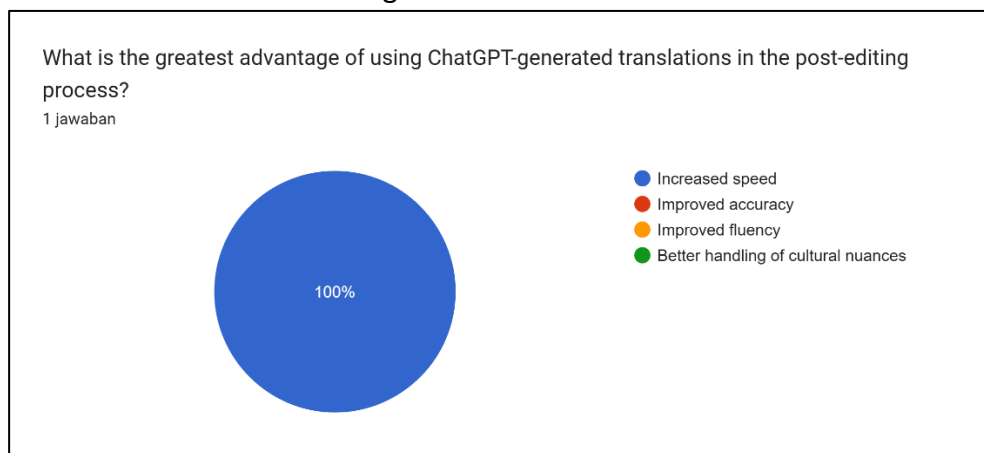


Figure 4.1.2. Primary Benefits of ChatGPT in Post-Editing

Figure 4.1.2 above shows that speed is identified as the primary benefit of ChatGPT in academic post-editing. One participant reported that employing ChatGPT for post-editing enhanced efficiency by four times compared to traditional methods. Nonetheless, this assertion is qualitative and should be supported by future studies that incorporate time-

tracking data or statistical standards for validation. This finding supports Mukti et al. (2024), who observed that AI tools like ChatGPT reduce repetitive correction efforts, allowing post-editors to concentrate on higher-order revisions, which ultimately accelerates translation turnaround time.

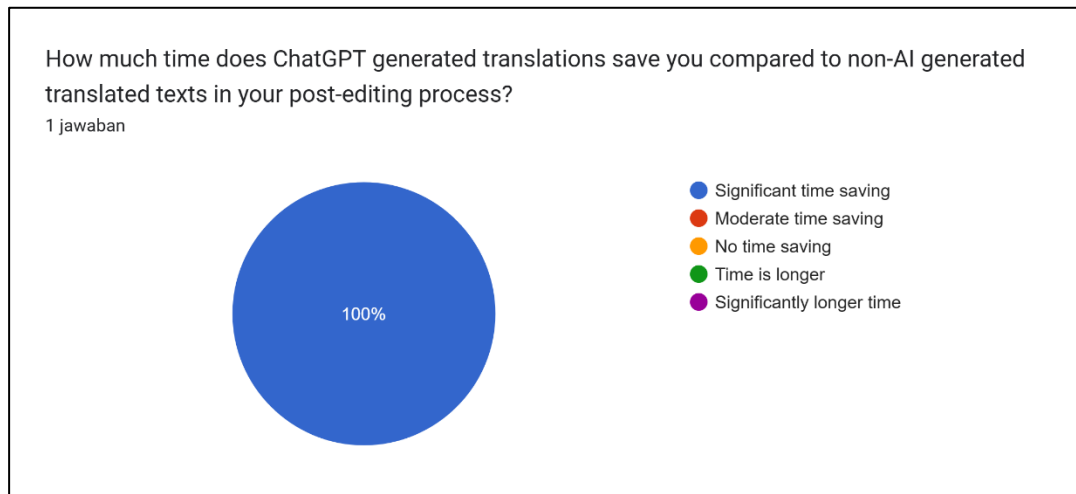


Figure 4.1.3. The Perceived Impact of ChatGPT on Time Savings in Post-Editing

Figure 4.1.3 illustrates a unanimous perception of significant time savings when using ChatGPT-generated translations compared to non-AI outputs. The participant confirmed that ChatGPT-enabled workflows reduced the time spent on revisions, particularly in academic contexts that require fast delivery. While this response supports the efficiency claim, the data would be more compelling if it included quantitative comparisons, such as average minutes saved per task. This observation is reinforced by Gao Lin, Zhao, & Cai (2024), who emphasized that AI-driven tools enhance productivity in translation workflows but still require human refinement for contextual accuracy and fluency.

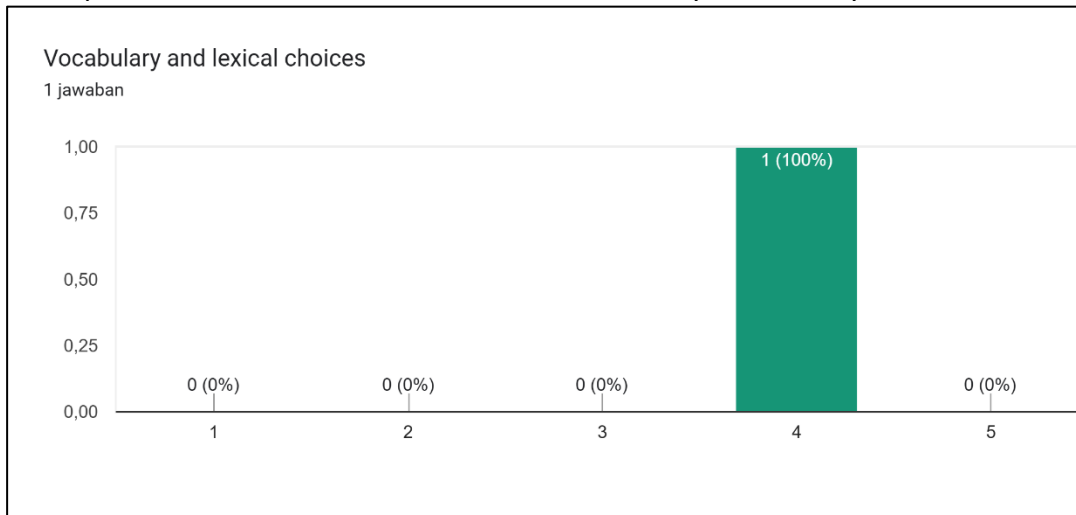


Figure 4.1.4: Evaluation of Vocabulary and Lexical Choices in ChatGPT-Generated Translations

Figure 4.1.4 illustrates the post-editor's evaluation of ChatGPT's ability to select vocabulary and lexical expressions suitable for academic settings. The participant scored this feature a 4 out of 5, indicating its strong yet imperfect performance. This high rating suggests that ChatGPT effectively understands general academic vocabulary and produces

contextually appropriate word choices. However, the rating also highlights occasional discrepancies or less-than-ideal word choices, particularly with complex, subject-specific material. This outcome emphasizes that although ChatGPT provides a robust base in lexical selection, human oversight is crucial for achieving accuracy in specialized domain translations.

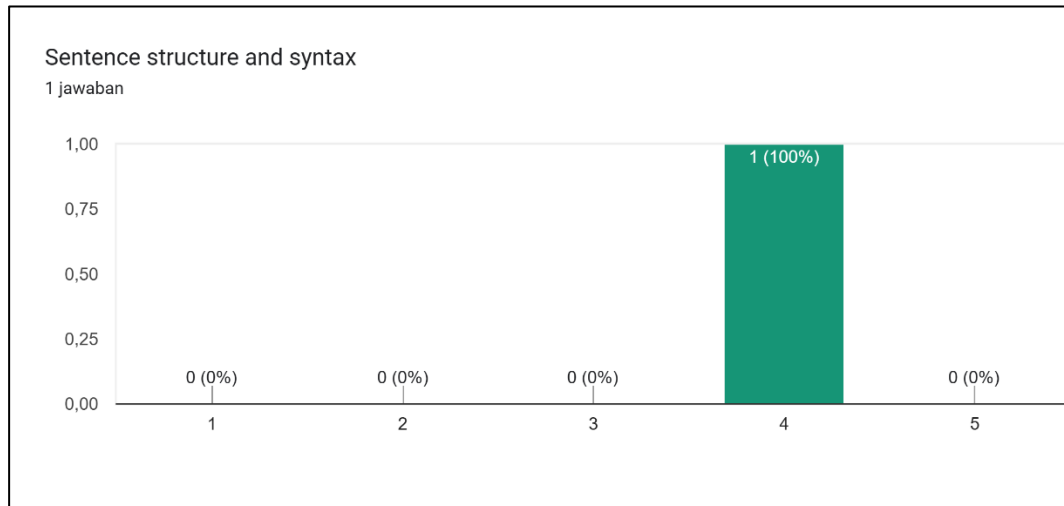


Figure 4.1.5: Evaluation of Sentence Structure and Syntax

In Figure 4.1.5, the post-editor evaluated ChatGPT's ability to produce well-structured sentences and accurate syntax, giving it a score of 4 out of 5. This rating highlights the model's proficiency in creating grammatically correct and logically ordered sentences, which is crucial in academic writing. The tool consistently applies sentence patterns and syntactic flow that conform to scholarly standards. Nonetheless, some awkward phrasing or overly literal interpretations could account for the point deduction. These results emphasize the necessity of post-editing to improve fluency and clarity in complex sentence structures.

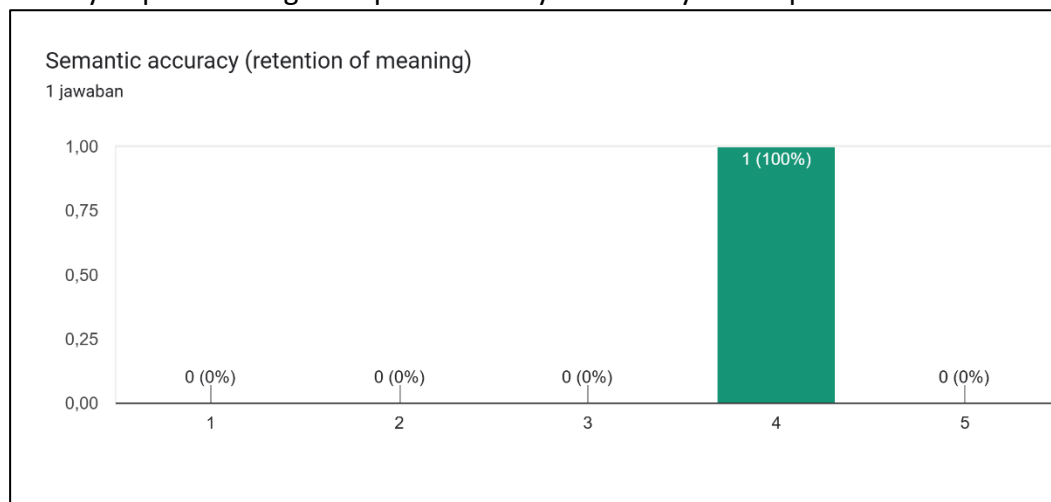


Figure 4.1.6: Evaluation of Semantic Accuracy (Retention of Meaning)

Figure 4.1.6 examines semantic accuracy, which refers to ChatGPT's ability to maintain the meaning of the source text during translation. The participant scored this aspect a 4 out of 5, indicating that ChatGPT effectively conveys intended meanings overall. This score suggests that the tool is dependable when translating straightforward content but may encounter difficulties with nuanced or ambiguous phrases, potentially diminishing the

depth or subtlety of the original message. This conclusion corroborates earlier research, including Gao Lin, Zhao, & Cai (2024), which indicates that although AI tools like ChatGPT generally excel in semantic fidelity, complex academic concepts still benefit from human input interpretation.

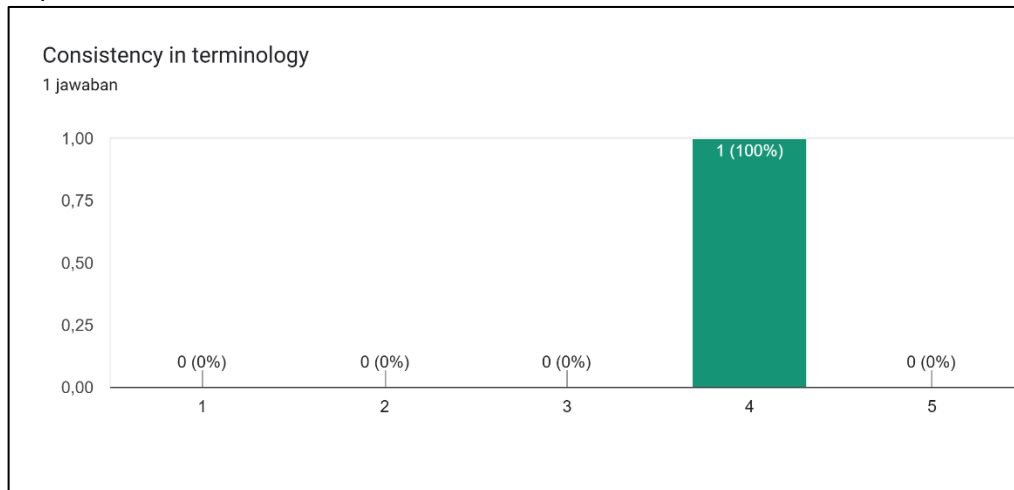


Figure 4.1.7: Evaluation of Consistency in Terminology

Figure 4.1.7 illustrates that the post-editor rated ChatGPT’s consistency in terminology usage at 4 out of 5. Consistent application of terminology in translations is vital in academic writing to prevent confusion or misinterpretation. Although the high score reflects ChatGPT’s competency, occasional discrepancies in terminology—particularly with technical or field-specific terms—could contribute to a less-than-perfect score. These findings suggest that while ChatGPT effectively manages repetitive language, human oversight may still be necessary to ensure uniformity in specialized documents.

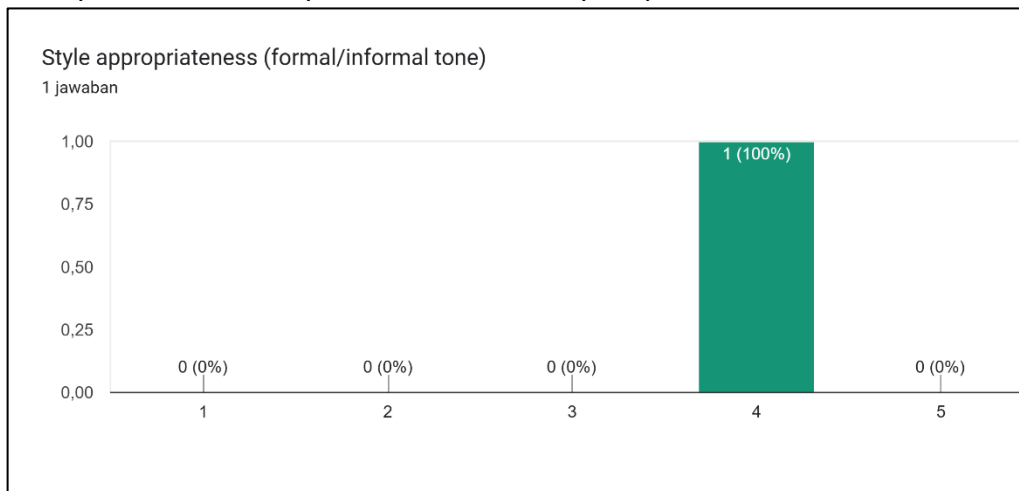


Figure 4.1.8: Evaluation of Style Appropriateness (Formal/Informal Tone)

Figure 4.1.8 assesses ChatGPT’s capability to maintain an appropriate tone, particularly its alignment with the formal style expected in academic writing. The post-editor rated this criterion a 4 out of 5, indicating that ChatGPT usually follows academic tone conventions but may sometimes resort to conversational language or misinterpret tone expectations based on the context. This result underscores the necessity of tone-checking during post-editing to guarantee the translated output aligns with academic standards

expectations.

4.2 Limitations of ChatGPT-Generated Translations in Post-Editing

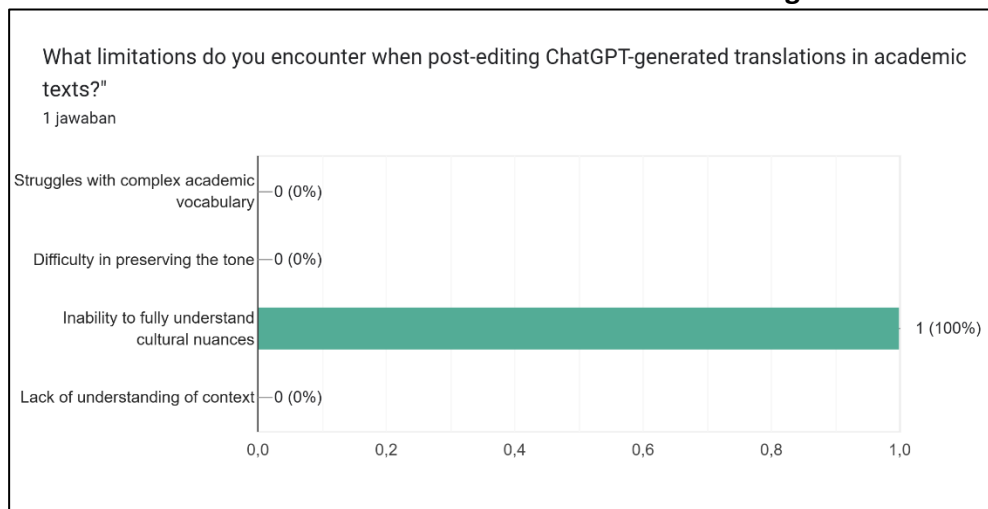


Figure 4.2.1: The Limitations of ChatGPT-Generated Translations

Figure 4.2.1 highlights the major limitation of not fully grasping cultural nuances. The post-editor noted that this issue commonly arose when translating English-Indonesian academic texts. It became particularly challenging with phrases that contained local cultural references or nuanced meanings, which ChatGPT often oversimplified or inaccurately translated. This limitation highlights the challenges associated with translating into languages that are underrepresented in the training datasets of large language models (LLMs).

This observation aligns with Ye (2024) findings, which suggest that ChatGPT often lacks the necessary contextual depth for culturally rich or underrepresented languages. Ye noted that while AI-generated outputs may exhibit fluency and structural correctness, they often fail to accurately convey subtle cultural or idiomatic meanings—crucial elements in academic translation. Therefore, despite ChatGPT's ability to minimize surface-level errors, human intervention is vital for achieving deeper semantic understanding, particularly in scholarly work.

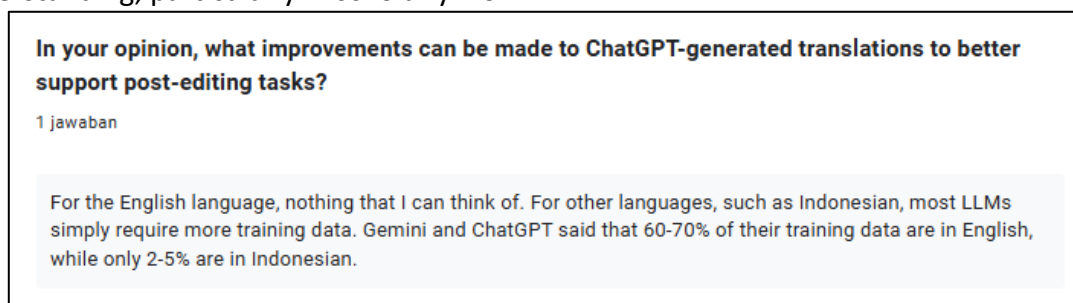


Figure 4.2.2: Suggested Improvements for ChatGPT-Generated Translations in Post-Editing Tasks

Figure 4.2.2 elaborates on the previous limitation by highlighting the root issue: an

imbalance in training data. The participant mentioned that although ChatGPT excels in English, its performance in Indonesian is limited due to a lack of sufficient training data. Expressly, it was noted that 60–70% of the training data is in English, while Indonesian accounts for only 2–5%, as reported by other LLM developers, such as Gemini.

This disparity hampers ChatGPT's ability to deliver accurate and contextually relevant translations for less-represented languages. These observations reinforce the concerns raised by Gao Lin, Zhao, & Cai (2024), who noted that AI tools typically perform worse in low-resource language environments. The participant emphasized that enhancing the diversity of the training data is crucial for improving ChatGPT's effectiveness in multilingual academic settings.

5. Conclusion

This study concludes that ChatGPT presents significant benefits for academic post-editing processes, particularly in enhancing efficiency, correcting grammar and punctuation, and aligning text with an academic tone. These advantages make it a valuable resource for professional translators who face tight deadlines or manage substantial amounts of material. A participant in this case study noted that ChatGPT's organized and coherent outputs allowed for more attention to be devoted to higher-level edits, such as enhancing contextual relevance and coherence.

Nevertheless, the findings uncover considerable limitations. The primary issue is ChatGPT's struggle with cultural nuances and idiomatic phrases—particularly in languages that are underrepresented in its training dataset, like Indonesian. This limitation suggests that while ChatGPT excels in essential linguistic accuracy, human intervention remains necessary to ensure that translations meet academic standards in terms of meaning, tone, and cultural sensitivity.

The study's limitations stem from its reliance on a single participant, which constrains the generalizability of the results. Although this participant's knowledge in both translation and Artificial Intelligence provided valuable perspectives, future studies should involve a larger pool of translators from various backgrounds to yield more comprehensive and representative findings.

To enhance ChatGPT's effectiveness in academic translation, particularly for underrepresented languages, AI developers should take the following concrete actions:

- Expand the diversity and quantity of training data for non-English languages, such as Indonesian.
- Integrate academic domain-specific corpora into model fine-tuning to better capture the stylistic and terminological standards of scholarly writing.
- Improve context retention by refining how ChatGPT processes sequential or referential data across longer texts.
- Create customizable modules that let translators modify tone, terminology, or citation style according to discipline-specific requirements.

In conclusion, while ChatGPT shows significant promise as an auxiliary tool in academic translation workflows, its best application hinges on continuous collaboration between humans and AI. Human oversight is not only suggested but also vital to maintaining the linguistic accuracy, cultural richness, and academic standards essential in scholarly communication. Further research and technological advancements are necessary to expand

ChatGPT's inclusivity and adaptability across a broader range of academic and linguistic contexts.

References

- Atarchi, K., Elamari, A., & Marouane, M. (2024). International journal of translation and interpretation studies the role of artificial intelligence translation tools in academic translation: Faculties of pure sciences as a case study. *International Journal of Translation and Interpretation Studies*, 4(3), 7. <https://doi.org/10.32996/ijtis>
- Berg, B. L., & Lune, H. (2017). *Qualitative research methods for the social sciences* (9th ed.). London, England: Pearson Education Limited.
- Brahmana, C. R. P. S., Sofyan, R., & Putri, D. M. (2020). Problems in the application of google translate as a learning media in translation. *Language Literacy: Journal of Linguistics, Literature, and Language Teaching*, 4(2), 384–389. <https://doi.org/10.30743/ll.v4i2.2893>
- Carl, M. (2012). Translog II – a program for recording user activity data for empirical translation process research. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Mehmet Uğur Doğan, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, & Stelios Piperidis (Eds.), *The 8th International Conference on Language Resources and Evaluation* (pp. 4108–4112). European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/614_Paper.pdf
- Gao, R., Lin, Y., Zhao, N., & Cai, Z. G. (2024). Machine translation of Chinese classical poetry: a comparison among chatgpt, google translate, and deepl translator. *Humanities and Social Sciences Communications*, 11(1). <https://doi.org/10.1057/s41599-024-03363-0>
- Gill, S. S., & Kaur, R. (2023). ChatGPT: Vision and challenges. *Internet of Things and Cyber-Physical Systems*, 3, 262–271. <https://doi.org/https://doi.org/10.1016/j.iotcps.2023.05.004>
- Habeeb, L. S. (2020). Investigating the effectiveness of google translate among Iraqi students. *International Journal of Innovation, Creativity and Change*, 12(12).
- Hidayati, N. N., & Nihayah, D. H. (2024). Google translate, chatgpt or google Bard AI: A study toward non-English department college students' preference and translation comparison. *Inspiring: English Education Journal*, 7(1), 14–33. <https://doi.org/10.35905/inspiring.v7i1.8821>
- Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). *Is chatgpt a good translator? yes with gpt-4 as the engine*. Retrieved from <http://arxiv.org/abs/2301.08745>
- Kamaluddin, M. I., Rasyid, Moch. W. K., Abqoriyyah, F. H., & Saehu, A. (2024). Accuracy analysis of deepl: Breakthroughs in machine translation technology. *Journal of English Education Forum (JEEF)*, 4(2), 122–126. <https://doi.org/10.29303/jeeef.v4i2.681>
- Liu, Y., Han, T., Ma, S., Zhang, J., Yang, Y., Tian, J., He, H., Li, A., He, M., Liu, Z., Wu, Z., Zhao, L., Zhu, D., Li, X., Qiang, N., Shen, D., Liu, T., & Ge, B. (2023). Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology*, 1(2). <https://doi.org/10.1016/j.metrad.2023.100017>
- Mukti, M. A. P., Maulana, M. T., Rohmah, K. N., Abqoriyyah, F. H., & Saehu, A. (2024). Effectiveness of artificial intelligence usage as translation medium among English literature student of UIN Sunan Gunung Djati Bandung. *Journal of English Education Forum (JEEF)*, 4(2), 107–112. <https://doi.org/10.29303/jeeef.v4i2.683>
- Munday, J. (2008). *Introducing translation studies: Theories and applications* (2nd ed.).

London, UK: Routledge.

- Rebeca, L., & Stiegelbauer, P. (2013). Automatic translations versus human translations in nowadays world. *Procedia - Social and Behavioral Sciences*, 70, 1768–1777. <https://doi.org/10.1016/j.sbspro.2013.01.252>
- Salloum, S. A., Aljanada, R. A., Alfaisal, A. M., Al Saidat, M. R., & Alfaisal, R. (2024). Exploring the acceptance of chatgpt for translation: An extended TAM model approach. In *Studies in Big Data* (Vol. 144, pp. 527–542). springer science and business media deutschland gmbh. https://doi.org/10.1007/978-3-031-52280-2_33
- Sinulingga, A. W. K., Sofyan, R., & Mono, U. (2023). Time management and translation method in translating a scientific article: A case study on a professional translator. *JOALL (Journal of Applied Linguistics and Literature)*, 8(1), 213–226. <https://doi.org/10.33369/joall.v8i1.22250>
- Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage Publications.
- Vieira, L. N., Alonso, E., & Bywood, L. (2019). Introduction - Post-editing in practice_Process, product. *JoSTrans: The Journal of Specialised Translation*, 2(31), 1–13. https://jostrans.org/issue31/art_introduction.pdf
- Wang, L., Xu, S., & Liu, K. (2024). *Understanding students' acceptance of chatgpt as a translation tool: A UTAUT model analysis*. Retrieved from <https://www.researchgate.net/publication/381307164>
- Ye, L. (2024). Frontiers in computing and intelligent systems the feasibility study of artificial intelligence chatgpt in translation field. *Frontiers in Computing and Intelligent Systems*, 8(1), 52-57. <http://dx.doi.org/10.54097/5vp4mn42>
- Yin, R. K. (2003). *Case study research design and methods* (3rd ed.). Thousand Oaks, CA: Sage Publications.