TRANSLATION PERFORMANCE OF GOOGLE TRANSLATE AND DEEPL IN TRANSLATING INDONESIAN SHORT STORIES INTO ENGLISH

I Gusti Ayu Mahatma Agung[1], Putu Gede Budiartha[2], Ni Wayan Suryani[3]

Faculty of Foreign Languages
Universitas Mahasaraswati Denpasar
Indonesia
Email address: ayu.mahatma@unmas.ac.id[1], budiartha@unmas.ac.id[2], niwayansuryani1643@gmail.com[3]

Abstract

The prevalence of machine translation systems has increased in recent years due to their accessibility and the high demand for translation services worldwide. While translation tools and systems are useful for several fields and genres, their reliability for literary works continues to be a subject of debate. Critiques are often directed at the inadequate quality of literary texts that have been translated by machines. This research aimed to examine the performance of two neural machine translation systems, Google Translate and DeepL, in translating Indonesian short stories in the book "*Cerita-Cerita Jakarta*". This qualitative research was designed to assess the quality of machine translation in translating literary work. The translation errors category proposed by Koponen (2010) was used as the theoretical framework. The findings revealed several errors in the literary translation by Google Translate and DeepL. The translation errors included untranslated concept, omitted concept, and mistranslated concept. Challenges were encountered by both Google Translate and DeepL in translating cultural terms, onomatopoeia, abbreviations, idiomatic expressions, slang words, and address terms.

Keywords – *DeepL, Google Translate, Translation Performance*

Introduction

In an era defined by the fast spread of machine translation technologies, the global landscape of language interaction has experienced a significant change. The increasing accessibility of machine translation, along with a growing demand for translation services, has led to their extensive integration into various domains (Septarina et al., 2019). Machine translation has demonstrated remarkable effectiveness in translating simple texts, such as product descriptions, travel information, social media posts, and online reviews (Cohen et al., 2022). Additionally, machine translation has proven to be effective in technical domains, including business, science, and medicine (Setiawati et al., 2020; Vieira & Alonso, 2020). However, machine translation still has difficulties when it comes to comprehending creative writings that include cultural references, such as literary works (Guerberos-Arenas & Toral, 2022). Therefore, the application of machine translation to the realm of literature requires a closer examination.

As the utilization of machine translation becomes ubiquitous, the question of its reliability in literary translation has emerged as a topic of intense debate (Omar & Gomaa, 2020). The primary focus of the debate often revolves around the perceived inadequacies of machine translation in conveying the complexity and detail of literary works (Mohar et al., 2020). Literary works function as a vessel of cultural identity and creative expression, using language in a manner that goes beyond simple communication (Rahman & Rahman, 2020). The complexities of storytelling, cultural references, and the subtle nuances of expression present a significant challenge for automated systems aiming to accurately convey the essence of these works (Toral & Way, 2015).

Several prior studies focus on evaluating the performance of machine translation across different language pairs and domains. The literature reveals a growing interest in assessing the accuracy and quality of machine translation outputs, particularly in the context of specialized domains and languages with unique linguistic characteristics. The research conducted by Munkova et al. (2022) identifies common errors in Google Translate's rendition of English newspaper articles into Slovak by using a modified version of Vanko's categorical framework. The research sheds light on the challenges posed by synthetic and inflectional languages, highlighting lexical semantics and syntactic-semantic correlativeness as prominent error categories.

Furthermore, the study conducted by Cambedda et al. (2021) delves into medical translation, specifically examining Russian-Italian translations related to the coronavirus pandemic. The study focuses on two prominent Neural Machine Translation tools, DeepL and Yandex, evaluating their performance on specialized and popular science articles. The research highlights the linguistic implications of the pandemic as it introduces new terms and challenges the capabilities of machine translation systems in handling evolving scientific terminology. The comparative error analysis contributes to a nuanced understanding of the strengths and limitations of DeepL and Yandex in this context.

Moreover, the research conducted by Jabak (2019) addresses the limited research on Arabic-English translation using Google Translate. The study compares Google Translate outputs with model translations, revealing lexical and syntactic errors that impact the overall translation quality. The findings emphasize the significance of human involvement in achieving accurate and acceptable Arabic-English translation, since Google Translate is not reliable enough to be used as the only tool for translating this specific language pair.

While previous studies mostly focus on the performance of machine translation in translating newspaper articles, medical texts, and general material, the current study specifically examines the performance of machine translation in translating literary work. This research aims to examine the performance of Google Translate and DeepL, two neural machine translation systems, in translating Indonesian short stories in the anthology "*Cerita-Cerita Jakarta*" into English. By examining the outputs of Google Translate and DeepL in translating Indonesian short stories, this research offers a deeper understanding of the specific challenges encountered when translating literary works. The findings provide valuable insights for both scholars in translation studies and developers working on refining machine translation algorithms. Moreover, this research aspires to enhance understanding of the complex relationship between technology, language, and culture. Ultimately, it seeks to advance the creation of machine translation systems that are both more accurate and culturally sensitive.

Methods

A qualitative research method was employed in this study to evaluate the machine-generated translations of Indonesian short stories. The primary data source for this study was a compilation of short stories titled "*Cerita-Cerita Jakarta*," edited by Maesy Ang and Teddy W. Kusuma and published by Post Press in 2021. This specific anthology was selected because it represented a wide range of linguistic and cultural aspects essential to Indonesian literature. By focusing on a compilation of short stories that represent the linguistic

e-ISSN : 3025-9002

complexities of the Indonesian language, we aim to identify various translation challenges, such as cultural references and idiomatic expressions embedded in the stories.

The data collection involved a systematic process of immersing into the short stories from "*Cerita-Cerita Jakarta*." Each short story was read thoroughly, and the Indonesian text was then inputted into both Google Translate and DeepL to obtain English translations. Following the translation process, a careful comparison was conducted between the outputs from Google Translate and DeepL. Subsequently, the theory of translation error categories by Koponen (2010) was applied to analyze the translation errors. This framework categorized translation errors into distinct categories, providing a structured approach to assess the quality of machine translation systems' performance in translating literary work. Koponen's theory includes six main types of translation errors. "Omitted Concept" occurs when a concept from the source text is not conveyed or translated into the target text. "Added Concept" involves introducing new information in the target text that was not present in the source text, potentially changing the meaning. "Untranslated Concept" entails directly borrowing source language words into the target text, causing potential comprehension difficulties for the target readers. "Mistranslated Concept" occurs when a target text concept has an incorrect meaning within the given context. "Substituted Concept" involves using a concept in the target text that is not a direct equivalent but is contextually valid. Finally, "Explicitated Concept" is when the target text explicitly states information left implicit in the source text without introducing new details.

Finding and Discussion

The analysis revealed that the translation of Indonesian short stories in the anthology "*Cerita-Cerita Jakarta*" contained several types of errors. The identified translation errors include untranslated, omitted, and mistranslated concepts. Both Google Translate and DeepL faced challenges when translating cultural terminology, onomatopoeia, abbreviations, idiomatic expressions, slang words, and address terms. The following section presents a comprehensive analysis of the translation.

The Performance of Google Translate and DeepL in Translating Cultural Term

Data 1

Source Language (SL): *Aku harus naik* ***angkot*** *sekali ke tempat itu.*
Google Translate (GT): I had to take a **public bus** to that place.
DeepL Translate (DT): I had to take an ***angkot*** once to that place.

In data 1, the Indonesian word "*angkot*" was not translated by DeepL. The word "*angkot*" stands for "*angkutan kota*," which means "public transport." It is usually in the form of a minibus and has its own route within the city. DeepL was not successful in finding the equivalent of the term "*angkot*" in English. The term was not translated, and it may confuse the target readers who may not be familiar with the Indonesian transportation terminology. Therefore, the translation of DeepL in this context can be categorized as an untranslated concept. In contrast, Google Translate translated the word "*angkot*" into "public bus," which conveys a similar meaning. The various translations of the term "*angkot*" provide a clear illustration of the intricate difficulties encountered by machine translation in accurately capturing specialized terminology and cultural nuances. The difference in performance among machine translation systems highlights the necessity of accurately expressing context-specific phrases, particularly in domains such as urban transportation, where exact terminology is crucial for comprehension.

The Performance of Google Translate and DeepL in Translating Onomatopoeia

Data 2

SL: ***Gusrak!*** *Motor kita nyusruk ke dalam lubang lumpur.*
GT: ***Gusrak!*** Our bike plunged into the mud pit.
DT: ***Gusrak!*** Our motorbike crashed into a mud hole.

In data 2, the Indonesian word "*gusrak*" is an onomatopoeia that expresses the sound of something falling after hitting an object. Neither Google Translate nor DeepL translated the onomatopoeia. Therefore, the translation of the onomatopoeia can be categorized as an untranslated concept. This classification indicates that the machine translation tools did not successfully convey the onomatopoeic representation of the sound in the target language, potentially leading to a loss of vividness and specificity in the translated content. The absence of a translated equivalent for an onomatopoeia like "*gusrak*" highlights the challenges that machine translation faces in capturing and reproducing nuanced linguistic elements, especially those rooted in sensory experiences like sound representations. It emphasizes the importance of human intervention and linguistic understanding in handling such onomatopoeic expressions for more contextually accurate translations.

The Performance of Google Translate and DeepL in Translating Abbreviations

Data 3

SL: *Di bawah pohon beringin yang tak sejuk itulah tempat mangkal para pedagang kaki lima, segerombolan calo yang merokok, dan para calon* ***TKW*** *yang kelihatan susah.*
GT: Under the banyan tree, which is not cool, is where the street vendors hang out, a group of touts who smoke, and prospective **migrant workers** who look difficult.
DT: Under the cool banyan trees are the street vendors, the hordes of touts who smoke, and **the women** who look like they're in trouble.

In both Google Translate and DeepL translations in data 3, the abbreviation "TKW" (*Tenaga Kerja Wanita*), which specifically refers to female migrant workers, was not translated accurately. According to Koponen's theory, the translation is an example of an omitted concept. The translation by Google Translate uses the generic term "migrant workers" without specifying the gender. This omission reduces the accuracy of the translation, as "TKW" carries a specific cultural and gender-related connotation that is not fully conveyed in the target language. In the translation provided by DeepL, there is a notable omission and subsequent error in rendering the abbreviation "TKW". DeepL fails to provide an equivalent or explanation for "TKW," and instead, it translates it generically as "the women," which can be categorized as an omitted concept. This omission leads to a loss of specificity and cultural nuance in the source text. The omission of the abbreviation represents a failure to convey the specific meaning associated with female migrant workers in the original Indonesian text. This translation error highlights the challenge machine translation systems encounter in recognizing and appropriately rendering specialized acronyms or culturally specific abbreviations.

The Performance of Google Translate and DeepL in Translating Idiomatic Expressions

Data 4

SL: *Hal yang sudah pasti jelas adalah Pak Oleng* ***tidak kelihatan batang hidungnya.***
GT: What is definitely clear is that Mr Oleng **cannot see the bridge of his nose**.
DT: What was definitely clear was that Mr. Oleng was **nowhere to be seen**.

In data 4, the translation of the Indonesian phrase "*tidak kelihatan batang hidungnya*" by Google Translate and DeepL reflects notable differences. Google Translate rendered it as "cannot see the bridge of his nose," which, when analyzed through Koponen's theory, falls into the category of a mistranslated concept. The literal translation fails to capture the idiomatic nature of the expression, resulting in a confusing and out-of-context interpretation. In contrast, DeepL's translation, "nowhere to be seen," demonstrates a more accurate representation of the intended meaning. The source language phrase "*tidak kelihatan batang hidungnya*" implies the person's complete absence, and DeepL successfully conveys this sense in the target language. Therefore, the analysis of data 4 shows that DeepL outperformed Google Translate in translating this idiomatic expression by preserving its meaning effectively.

The Performance of Google Translate and DeepL in Translating Slang Words

Data 5

SL: *Lantaran tidak mau melewatkan kesempatan emas ini, maka aku rela* ***'menembak'*** *paspor. Biro menjamin si* ***penembak*** *tidak perlu tahu proses rumit berlangsung.*
GT: Because I didn't want to miss this golden opportunity, I was willing to **'shoot'** the passport. The bureau ensures the **shooter** does not need to know the complicated process is taking place.
DT: Because I didn't want to miss this golden opportunity, I was willing to **'shoot'** the passport. The bureau guaranteed that the **shooter** didn't need to know the complicated process was taking place.

In the provided translation in data 5, a significant mistranslation is evident in the interpretation of the Indonesian slang words "*menembak*" and "*penembak*." The translations by Google Translate and DeepL use the term "shoot" to convey "*menembak*," which inaccurately suggests a physical action rather than the intended meaning of taking a shortcut or expedited action to obtain the passport. Similarly, the term "shooter" is employed for "*penembak*," deviating from the original context where it refers to an individual taking a shortcut, not someone physically shooting. This introduces a misleading element, potentially leading to a misunderstanding of the character's actions and motivations. The concept of bypassing official procedures, implicit in "*menembak*," is not effectively conveyed. To align with Koponen's theory, the translation should prioritize accuracy in representing the source language concepts, necessitating a revision that captures the essence of "taking a shortcut to obtain the passport" without introducing inappropriate terms.

The Performance of Google Translate and DeepL in Translating Address Term

Data 6

SL: *"Maukah Ibu meminjamkan sarung Ibu?"*
"Boleh, ***Neng****. Tapi sarung ibu ini kotor sekali. Apa Neng tidak malu?"*
GT: "Will you lend me your sarong?"
"Yes, Neng. But this mother's sarong is very dirty. Aren't you embarrassed, ***Neng***?"
DT: "Will you lend me your sarong?"
"Yes, ***Neng***. But your sarong is very dirty. Aren't you embarrassed?"

In data 6, the Indonesian word "*Neng*" poses a challenge for both Google Translate and DeepL, as it functions as an address term rather than a straightforward name. The absence of a translation for "*Neng*" in both outputs can be categorized under the theory of translation error by Koponen as an untranslated concept. This classification arises because neither Google Translate nor DeepL could accurately determine whether "*Neng*" is a person's name or an address term. This leads to a lack of a corresponding translated counterpart in English. Even if both programs were to identify "*Neng*" as an address term, the challenge lies in finding a suitable and culturally appropriate translation in English. The potential translations like "Miss" or "young lady" may not precisely capture the cultural nuances associated with "*Neng*" and could sound awkward or unfamiliar in the target language context. Therefore, the untranslated concept error emerges from the inherent difficulty in translating address terms with cultural specificity.

Conclusion

Based on the analysis, several errors were found in the translation of Indonesian short stories in the anthology "*Cerita-Cerita Jakarta*". The translation errors found were untranslated concept, omitted concept, and mistranslated concept. Both Google Translate and DeepL still face challenges in translating cultural terms, onomatopoeia, abbreviations, idiomatic expressions, slang words, and address terms. This finding promotes a thorough assessment of the advantages and disadvantages of various machine translation systems, emphasizing the need for users to carefully choose the most suitable tool for their individual language and contextual needs. As we explore automatic translation, it is crucial to acknowledge the possibility of differences in the results and make rational choices to ensure the accuracy of meaning across different languages. Machine translation systems often rely on patterns and statistical associations in large datasets, and when faced with domain-specific or culturally nuanced terms, they may struggle to provide accurate equivalents. This limitation highlights the importance of human translators who possess cultural and contextual knowledge and can decipher the nuances of specialized terms. Future researchers should conduct thorough evaluations of machine translation systems, considering both their strengths and limitations. Additionally, research efforts should focus on developing models that address the challenges of handling domain-specific or culturally nuanced terms and explore collaborative approaches that integrate the expertise of human translators for more accurate and culturally sensitive translations.

References

Ang, M., & Kusuma, T. W. (2021). *Cerita-Cerita Jakarta*. Jakarta: Post Press.

Cambedda, G., Di Nunzio, G. M., & Nosilia, V. (2021). A Study on Machine Translation Tools: A Comparative Error Analysis between DeepL and Yandex for Russian-Italian Medical Translation. *Umanistica Digitale*, *10*(1), 139–163. https://doi.org/10.6092/issn.2532-8816/12631

Cohen, F. S., Zhong, Z., & Li, C. (2022). Semantic Graph for Word Disambiguation in Machine Translation. *Multimedia Tools and Applications*, *81*(30), 43485–43502. https://doi.org/10.1007/s11042-022-13242-y

Guerberos-Arenas, A., & Toral, A. (2022). Creativity in Translation: Machine Translation As a Constraint for Literary Texts. *Translation Spaces*, *11*(2), 184–212. https://doi.org/10.1075/ts.21025.gue

Jabak, O. O. (2019). Assessment of Arabic-English Translation Produced by Google Translate. *International Journal of Linguistics, Literature and Translation (IJLLT)*, *2*(4), 238–247. https://doi.org/10.32996/ijllt.2019.2.4.24

Koponen, M. (2010). Assessing Machine Translation Quality with Error Analysis. *Electronic Proceedings of the VIII KäTu Symposium on Translation and Interpreting Studies*, *4*, 1–12. https://sktl-fi.directo.fi/@Bin/40701/Koponen_MikaEL2010.pdf

Mohar, T., Orthaber, S., & Onič, T. (2020). Machine Translated Atwood: Utopia or Dystopia? *ELOPE: English Language Overseas Perspectives and Enquiries*, *17*(1), 125–141. https://doi.org/10.4312/elope.17.1.125-141

Munkova, D., Panisova, L., & Welnitzova, K. (2022). A Human Evaluation of English-Slovak Machine Translation. *Perspectives: Studies in Translation Theory and Practice*, 1–20. https://doi.org/10.1080/0907676X.2022.2116989

Omar, A., & Gomaa, Y. A. (2020). The Machine Translation of Literature: Implications for Translation Pedagogy. *International Journal of Emerging Technologies in Learning*, *15*(11), 228–235. https://doi.org/10.3991/IJET.V15I11.13275

Rahman, F. F., & Rahman, F. (2020). Translation or Intertextuality: A Literature Comparative Analysis of “The Young Dead Soldiers Do Not Speak” by Archibald MacLeish and “Krawang Bekasi” by Chairil Anwar. *Elsya : Journal of English Language Studies*, *1*(3), 110–117. https://doi.org/10.31849/elsya.v1i3.5320

Septarina, A. A., Rahutomo, F., & Sarosa, M. (2019). Machine Translation of Indonesian: A Review. *Communications in Science and Technology*, *4*(1), 12–19. https://doi.org/10.21924/cst.4.1.2019.104

Setiawati, I. A. M. F., Yadnya, I. B. P., & Aryawibawa, I. N. (2020). A Comparison of Translation Readability between Google Translate and Human Translator in the Medical Book Entitled “Medical-Surgical Nursing.” *Lingua Scientia*, *27*(2), 65–76. https://doi.org/10.23887/ls.v27i2.25590

Toral, A., & Way, A. (2015). Machine-assisted Translation of Literary Text. *Translation Spaces*, *4*(2), 240–267. https://doi.org/10.1075/ts.4.2.04tor

Vieira, L. N., & Alonso, E. (2020). Translating Perceptions and Managing Expectations: An Analysis of Management and Production Perspectives on Machine Translation. *Perspectives: Studies in Translation Theory and Practice*, *28*(2), 163–184. https://doi.org/10.1080/0907676X.2019.1646776