

# A TRANSLATION PRODUCT ANALYSIS OF CHATGPT BASED ON THE ERROR TAXONOMY BY VILAR ET AL. (2006): PROBLEMS AND SOLUTIONS

DEWI ROSNITA HARDIANY\*<sup>1</sup>, MOCHAMAD RIZQI ADHI PRATAMA<sup>2</sup>

<sup>1,2</sup>Universitas Ngudi Waluyo

First received: April 12, 2025

Final proof received: May 2, 2025

\*Corresponding Author: dewirosnita@unw.ac.id

## Abstract

This study aims to analyze the translation product of ChatGPT by identifying the errors based on the error taxonomy proposed by Vilar et al. (2006). As AI-powered machine translation (MT) becomes increasingly prevalent, evaluating the quality of the output and identifying specific weaknesses is crucial. This research used a descriptive qualitative approach to examine 30 English-to-Indonesian translation outputs generated by ChatGPT. The analysis was guided by Vilar et al.'s error categories: missing words, word order, incorrect words, unknown words, and punctuation. Results show that while ChatGPT performs well in terms of fluency, some errors still occur in the form of incorrect words (33,3%), word orders (23,3%), missing words (20%), and punctuation (13,3%) and unknown words (10%). It can be concluded that the most frequent errors are incorrect words, word orders, and missing words. Proposed solution involves post-editing strategies by adapting translation techniques by Molina and Albir in order to make the translation result more accurate, natural, and aligned with the norms of the target language.

Keywords: Translation error, Vilar et al. taxonomy, ChatGPT

## INTRODUCTION

AI (Artificial Intelligence) is one of the most beneficial technologies for facilitating various human needs, including in the field of translation. The term *machine translation* (MT) has existed since the mid-20th century (Koerner & Asher, 1995). This term refers to a modern method of translation carried out with the help of computers. Machine translation combines the field of translation with computer science, involving the process of translati-

on performed by computers (Lin & Chien, 2009, p. 134). Machine translation involves both computers and translation software. During the translation process, the machine operates automatically. The system converts a source language text into a target language text. Machine translation has transformed the dynamics of multilingual communication in recent decades. The surge in globalization, coupled with advances in artificial intelligence (AI), has led to the creation of powerful translation tools ca-

pable of generating near-human quality outputs (Hutchins, 2005; Koehn, 2010). ChatGPT, a large language model developed by OpenAI, is one such tool that has received significant attention for its natural language understanding and generation capabilities (OpenAI, 2023). Its adaptability across languages and genres positions it as a versatile resource for translation and communication (Brown et al., 2020). One of ChatGPT's advantages is its ability to facilitate the translation process from the source language to the target language quickly and efficiently. In addition, ChatGPT also helps users identify synonyms and aspects of grammar. Despite these advances, the fidelity and accuracy of ChatGPT's translations remain under scrutiny. Research in computational linguistics emphasizes the distinction between fluency and adequacy in translation outputs (Papineni et al., 2002; Bahdanau et al., 2015). Particularly for under-resourced language pairs like Indonesian-English, translation quality is often compromised by lack of parallel corpora, syntactic divergence, and cultural-linguistic nuances (Suryawinata et al., 2021; Purwarianti & Pribadi, 2020).

To determine the accuracy of ChatGPT in translating texts from Indonesian to English, an analysis of its translation output is necessary. One way to assess translation accuracy is by analyzing the translated texts using the error taxonomy proposed by Vilar et al. (2006), which classifies various types of translation errors.

Translation error analysis serves as a crucial diagnostic tool in evaluating MT outputs (Vilar et al., 2006; Lommel et al., 2014). By classifying translation errors into specific categories, scholars and developers can pinpoint the systemic limitations of MT systems and improve future iterations (Bentivogli et al., 2016; Castilho et al., 2017). This study employs the error

taxonomy of Vilar et al. (2006) to systematically evaluate the output of ChatGPT in English-Indonesian translation tasks. The objective is twofold: (1) to identify common patterns of translation errors, and (2) to provide solutions to translation inaccuracies in ChatGPT outputs by adapting translation techniques proposed by Molina and Albir. By using both theoretical frameworks, it is expected that the translation results will become more accurate and acceptable.

Machine translation (MT) refers to the automatic conversion of text from one language to another using computational algorithms. Recent AI-based systems such as Google Translate, DeepL, and ChatGPT operate on neural network architectures, leveraging massive datasets and context-sensitive algorithms to achieve near-human translation fluency (Bahdanau et al., 2015; Vaswani et al., 2017).

ChatGPT is an AI model trained using reinforcement learning from human feedback (RLHF) and large language corpora. Though not exclusively built as a translation tool, ChatGPT has demonstrated strong capabilities in cross-linguistic tasks (OpenAI, 2023). However, studies have pointed out inconsistencies, especially in complex or idiomatic expressions (Wu et al., 2023).

Vilar et al. (2006) proposed a comprehensive classification of machine translation errors to evaluate the output of statistical and neural MT systems. Their taxonomy includes six main categories, namely: (1) Missing Words: A word or phrase from the source text is omitted in the translation, resulting in an incomplete or inaccurate transfer of meaning. (2) Incorrect Words: A word in the translation is semantically or contextually incorrect, although it may be a valid translation in other contexts. (3) Word Order Errors: Words are placed

in an incorrect order, violating the syntactic rules of the target language or making the sentence sound unnatural. (4) Unknown Words: One or more words from the source language remain untranslated, either due to system limitations or lack of corresponding entries in the model's database. (5) Punctuation error occurs when punctuation marks (like commas, periods, quotation marks, question marks, etc.) are incorrectly transferred, omitted, or misplaced in the translation process. These errors can affect clarity, readability and accuracy.

These categories help researchers and practitioners evaluate both surface-level errors (such as missing or untranslated words) and deeper linguistic inaccuracies (such as incorrect semantics or morphology). Applying this taxonomy in translation analysis allows for more targeted improvement strategies for machine translation systems like ChatGPT.

Indonesian and English differ significantly in terms of word order, tense marking, and idiomatic expressions. Studies show that MT tools often struggle with Indonesian passive constructions, pronoun drop, and reduplication (Purnamasari et al., 2022; Purwarianti & Pribadi, 2020). Such challenges highlight the importance of targeted error analysis when evaluating AI-generated translations in this language pair.

Translation techniques are specific procedures or methods that translators use to solve problems when translating a unit of text (words, phrases, or sentences) from one language (source language or SL) into another (target language or TL). They operate at a micro-level, meaning they focus on individual segments of a text (not the whole text) and help ensure the meaning, style, tone, and function of the source text is conveyed appropriately in the target text. "Translation techniques are procedures to analyze and clas-

sify how translation equivalence works." (Molina & Hurtado Albir, 2002). They propose 18 translation techniques and emphasize that techniques are contextual (not always right or wrong). Moreover, the choice of technique depends on factors such as the type of translation, purpose, audience, and cultural differences

## METHOD

This research used a descriptive qualitative method with content analysis to assess translation outputs generated by ChatGPT. The analysis focused on error categorization based on the taxonomy of Vilar et al. (2006). The source data consisted of 30 English source texts of various genres (e.g., academic abstracts, news excerpts, and informal dialogues), translated into Indonesian using ChatGPT (GPT-4 model, accessed May-June 2025). The source texts were selected to represent various linguistic structures and complexities, including idioms, passive constructions, phrasal verbs, and technical terminology.

Each sentence produced by ChatGPT was compared to a reference translation created by human translators. The differences were analyzed and classified into one or more of Vilar's error categories. In addition, to provide solutions to translation inaccuracies in ChatGPT outputs by adapting translation techniques proposed by Molina and Albir.

## FINDINGS AND DISCUSSION

The researchers conducted data analysis on 30 sentences from English source texts of various genres (e.g., academic abstracts, news excerpts, and informal dialogues). Those sentences were translated into Indonesian using ChatGPT (GPT-4 model, accessed May-June 2025). The last step was analyzing the data using error classification from Vilar et al. The results

can be seen on this following table:

No	Types of Error	Occurrence (out of 30 texts)	Percentage
1	Incorrect words	10	33,3%
2	Word order	7	23,3%
3	Missing words	6	20%
4	Punctuation	4	13,3%
5	Unknown words	3	10%

The most frequent error type was **incorrect word choice**, where ChatGPT selected semantically or contextually inappropriate words. An example includes translating “*critical thinking skills*” into “*ke-mampuan berpikir kritis*”, but in an academic context, a better translation would be “*keterampilan berpikir kritis*”. Followed by **Word Order Errors** were mainly found in sentences with passive constructions, where the Indonesian structure significantly differs from English. Meanwhile **Missing Words** occurred when some grammatical or contextual elements were omitted, affecting sentence completeness and meaning. For instance, ChatGPT often dropped prepositions or auxiliaries in longer compound sentences. Here are some examples of the errors and the analysis:

#### **Incorrect Words (Lexical Choice)** **Sentence a**

Source: “He filed a motion to dismiss.”

ChatGPT: “Dia mengajukan gerakan untuk menolak.”

Error Analysis: “Motion” was literally translated to “**gerakan**”, ignoring the **legal meaning**. In **legal contexts**, “motion” is a **formal legal request to the court**, not a physical act.

Suggested fix: use correct technique established equivalent to translate the word “motion”. Established Equivalent = “Use

of a term recognized as a standard equivalent in the target language.”

Better Translation: “Dia mengajukan permohonan untuk membatalkan gugatan.”

#### Sentence b

Source: “The book is hard to understand.”

ChatGPT: “Buku ini keras untuk dipahami.”

Error Analysis: chose the wrong meaning by applying literal translation without considering context. The word “hard” was literally translated as “*keras*” (physically hard), rather than the correct contextual meaning “*sulit*” (difficult).

Suggested Fix: Shift from literal translation to modulation. It is used when there is a change in point of view, semantic category, or cognitive perspective

Better Translation: “Buku ini sulit dipahami”

#### **Word Order**

Sentence a:

Source: “Only after the rain stopped did he leave.”

ChatGPT: “Hanya setelah hujan berhenti dia pergi.”

Error Analysis: While grammatically acceptable, the sentence **sounds unnatural or awkward** in formal or idiomatic Indonesian.

Suggested Fix: to fix the problem, we can

use two translation techniques from Molina Albir, namely (1) Transposition by reordering the sentence to fit Indonesian structure and flow and (2) Modulation by changing the point of view or cultural expression: “only after” → “baru”

Better Translation: “Dia baru pergi setelah hujan berhenti.”

Sentence b:

**Source:** “She quickly ran to the store.”

**ChatGPT:** “Dia ke toko cepat-cepat lari.”

**Error Analysis:** The MT used literal translation for all units. It failed to restructure the sentence to follow Indonesian SVO + adverb order. The structure “**ke toko cepat-cepat lari**” is **not idiomatic** in Bahasa Indonesia. The adverb should be placed in a way that reflects **natural Indonesian syntax**. The result is disfluent and awkward for native speakers.

**Suggested Fix:** to fix the problem, we can use transposition techniques from Molina Albir to change of grammatical structure without changing the meaning.

English: “quickly ran” (Adv + V) → Indonesian: “berlari cepat-cepat” (V + Adv)

Better translation: “*Dia berlari cepat-cepat ke toko.*”

Missing Words

**Source:** “Due to his illness, he couldn’t attend.”

**ChatGPT:** “Karena sakit, dia tidak hadir”

**Error Analysis:** The possessive pronoun “his” is omitted. While the meaning is still generally understood in casual Indonesian, certain contexts (e.g., medical reports, formal documents) require explicit reference to the subject’s ownership of the illness. The subject “his” is important in some contexts.

**Suggested fix:** Use Literal Translation to preserve clarity and cohesion between cause (illness) and subject (he).

The word “penyakitnya” = “his illness” → Possessive suffix “-nya” restores the original meaning.

Better translation: “Karena penyakitnya, dia tidak bisa hadir.”

**Unknown Words**

Sentence a:

**Source:** “He works as a barista.”

**ChatGPT:** “Dia bekerja sebagai barista.”

**Error Analysis:** The word “**barista**” is a **loanword** that might **not be fully understood** by all Indonesian audiences, particularly those outside urban or coffeehouse culture. In some contexts, “barista” may need cultural adaptation or clarification (“*peracik kopi*”) if the target audience is unfamiliar with the term.

**Suggested fix:** We can choose between these two options:

Use the combination of borrowing and amplification techniques to keep the term but add an explanation.

Use Adaptation techniques to replace culturally specific term with one more familiar to the target audience

Better translation:

“Dia bekerja sebagai barista (peracik kopi).”  
“Dia bekerja sebagai peracik kopi.”

**Punctuation Errors**

**Source (Indonesian):** Dia berkata, “Saya akan datang besok.”

**ChatGPT:** He said “I will come tomorrow”.

**Error Analysis:** ChatGPT occasionally omits commas in complex sentences or overuses periods in enumeration, affecting readability. In the above sentence quotation marks are not correctly closed or punctuated.

**Suggested fix:** In English, quotation punctuation rules differ from Indonesian. The comma should be placed inside the quotation marks, and a period is required

at the end. The absence of the comma before the quote and missing period at the end constitutes a punctuation error.

**Better Translation:** He said, “I will come tomorrow.”

The results show that while ChatGPT can produce relatively fluent translations, semantic accuracy and syntactic fidelity still pose challenges. These findings align with Wu et al. (2023), who found that AI-based MT systems still lack consistent control over register and collocation in bilingual transfer. To address these issues, post-editing by human translators using suitable translation techniques is recommended, particularly for specialized texts.

## CONCLUSION

This study examined the translation quality of ChatGPT using the error taxonomy of Vilar et al. (2006). The results reveal that ChatGPT is a promising translation tool, yet still prone to key errors. The results of the research show that while ChatGPT performs well in terms of fluency, some errors still occur in the form of incorrect words (33,3%), word orders (23,3%), missing words (20%), and punctuation (13,3%) and unknown words (10%). It can be concluded that the most frequent errors are incorrect words, word orders, and missing words. To overcome those problems, proposed solution involves post-editing strategies by adapting translation techniques by Molina and Albir. It is expected that by conducting post-editing using correct techniques, the result of the translation will be more accurate, natural, and aligned with the norms of the target language, thereby improving both the linguistic quality and communicative effectiveness of the translated text

## REFERENCES

- Bahdanau, D., Cho, K., & Bengio, Y. (2015). *Neural machine translation by Jointly learning to align and translate*. arXiv preprint arXiv:1409.0473.
- Bentivogli, L., Cettolo, M., Girardi, C., & Federico, M. (2016). *Neural versus phrase-based machine translation quality: A case study*. EMNLP.
- Brown, T. et al. (2020). *Language Models are Few-Shot Learners*. NeurIPS.
- Dwyer, T. (2017). *Translation and the machine: Technology, meaning, ethics*. Routledge.
- Hutchins, W. J. (2005). *The history of machine translation in a nutshell*. Retrieved from <http://www.hutchinsweb.me.uk/Nutshell-2005.pdf>
- Koerner, E. F. K., & Asher, R. E. (Eds.). (1995). *Concise History of the Language Sciences: From the Sumerians to the Cognitivists*. Pergamon
- Lin, L.-C., & Chien, C.-H. (2009). *An exploration of translation students' learning styles and strategies*. *Asian EFL Journal*, 11(1), 134–154
- Molina, L., & Albir, A. (2002). *Translation techniques revisited: A dynamic and functionalist approach*. *Meta: Translators' Journal*, 47(4), 498–512. <https://doi.org/10.7202/008033ar>
- OpenAI. (2023). *GPT-4 technical report*. <https://openai.com/research/gpt-4>
- Purwarianti, A., & Pribadi, F. (2020). *Neural Machine Translation for Indonesian-English using Transformer*. IEEE ICCED.
- Purnamasari, I., et al. (2022). *Assessing Indonesian-English Machine Translation Performance*. *Lingua*, 18(1), 45–56
- Suryawinata, A., et al. (2021). *Exploring Indonesian-English Neural Machine Translation*. arXiv preprint arXiv:2106.00129.
- Vilar, D., Xu, J., D'Haro, L. F., & Ney, H. (2006). *Error analysis of statistical machine translation output*. In Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC).
- Vaswani, A., Shazeer, N., Parmar, N., Usz-

- koreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is all you need*. In Advances in neural information processing systems (pp. 5998-6008).
- Wu, X., Zhang, Y., & Yang, H. (2023). *Evaluating the fluency and fidelity of ChatGPT translations*. Journal of Computational Linguistics and Translation Studies, 5(2), 66–81.