

EVALUATION OF ACCURACY AND LINGUISTIC NUANCES IN MACHINE TRANSLATION OF DEEPL FOR INDONESIA- ENGLISH

Suci Pole Mappaita, Rizki Fauzi, Fadillah Pole Mappaita

Faculty of Language and Literature, Makassar State University, Indonesia

Email: suci.pole.mappaita@unm.ac.id , rizki.fauzi@unm.ac.id ,
fadillahpole.2112@gmail.com

Abstrak

Keywords:

Accuracy,
DeepL,
Linguistic nuances,
Machine Translation,
Pragmatic

This study aims to evaluate the accuracy and linguistic nuances in DeepL's machine translation output from Indonesian to English. The research employed a qualitative descriptive method with a case study approach, analyzing ten sentences featuring diverse syntactic and pragmatic characteristics. The results indicate that DeepL performs well in translating standard sentence structures and literal meanings. However, limitations arise when the system encounters idiomatic expressions, emotional tone, and pragmatic context. The translations often become too literal or contextually inappropriate. Thus, while DeepL is effective for general understanding, human translators remain essential in contexts requiring precise meaning and cultural sensitivity.

This is an open access article under the [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



INTRODUCTION

In the rapidly growing era of globalization and digitalization, the need for fast and accurate language translation services is becoming increasingly urgent, especially in the context of cross-cultural, educational and professional communication (Ekuerhare & Udoka, 2024). Indonesia as a country with the fourth largest number of language speakers in the world also feels the urgency (Aji et al., 2022). With more than 270 million people, and Bahasa Indonesia as the official language, the need for reliable translation from Bahasa Indonesia to English (and vice versa) is very high, especially in academic and professional contexts.

Machine translation (MT) has become one of the most practical solutions in meeting this need, mainly due to its speed and wide accessibility (Sakre, 2020). Google Translate, for example, has become a popular translation tool used by millions of people every day (Rosaria & Riana, 2021; Tuyet Phan & Chen, 2021). However, the quality of translation from Google Translate, especially for the Indonesian-English



language pair, is often questioned. Several studies have shown that the tool often produces semantically and syntactically inaccurate translations, and does not properly capture cultural context (Prasetyo & Wulandari, 2020; Mulyani, 2021). The main problem in machine translation from Indonesian to English lies in the differences in grammatical structure, affixation system, and cultural elements inherent in the source language (Sandra, 2018). Indonesian has a relatively flexible syntactic structure, as well as the use of conjunctions and time aspects that differ from English. In recent years, DeepL Translator, a neural machine translation (NMT) platform, has become widely recognized for its more accurate translation quality (Telaumbanua et al., 2024). As public and academic attention began to focus on the system's performance, DeepL Translator emerged as one of the leading systems to surpass Google Translate in accuracy, especially in terms of linguistic nuance and stylistic flexibility in several major European language pairs (Kamaluddin et al., 2024). However, this claim has not been systematically tested for the Indonesian-English language pair, which has quite different linguistic and cultural structures.

The main issues raised in this research are whether DeepL Translator is able to produce accurate and natural Indonesian-English translations, and what are its strengths and weaknesses compared to similar tools. This research stems from the need for a deeper understanding of the quality of translations produced by DeepL, especially from a linguistic and contextual perspective.

This research aims to qualitatively explore the strengths and weaknesses of DeepL in translating texts from Indonesian to English, with a focus on the linguistic quality of the translations. It will also provide a comparison with Google Translate as a widely used comparison tool.

The theoretical foundations used in this study cover both classical and contemporary approaches in translation studies. Nida and Taber's (1969) equivalence theory serves as the basic framework for assessing meaning fidelity, while Catford's (1965) concept of translation shifts is used to observe structural or semantic shifts between the source and target languages. In addition, Newmark's (1991) naturalness theory is used as a reference to assess the extent to which the translation feels natural to the target language speakers. Saputra & Rini (2021) applied this theory to evaluate the naturalness and accuracy of fan-subtitled TV series, finding that translations often contain both natural and unnatural elements. Similarly, Dewi et al. (2015) utilized Larson's theory, which aligns with Newmark's principles, to analyze the naturalness of a short story translation from Indonesian to English.

LITERATURE REVIEW

In addition, some previous studies examining the quality of machine translation from Indonesian to English (e.g. Prasetyo et al., 2021; Rahmawati, 2022) show that although tools such as Google Translate have improved, there are still shortcomings especially in handling idioms, passive structures, and cultural context. Google Translate's effectiveness is particularly limited when dealing with complex sentences and maintaining grammatical and textual equivalence (Suhono et al., 2020). As for Dewayanti (2024), it shows that NMTs such as DeepL have advantages in certain contexts, but also show weaknesses when it comes to handling idioms, metaphors, or complex sentence structures that are heavily influenced by the culture of the source language. In the Indonesian context, phenomena such as the ambiguity of the particle

“yang”, the compaction of information in a single clause, or the use of affixes that convey certain semantic nuances, often pose challenges for machine translators.

Thus, the main objective of this study is to identify and explain the strengths and weaknesses of DeepL Translator in translating texts from Indonesian to English qualitatively, as well as to provide feedback for translation technology developers and users in the academic and professional fields. It is hoped that this research can make a practical contribution in raising user awareness of the capabilities and limitations of AI-based translation tools, as well as providing input for translation technology developers to be more responsive to the needs of underserved language pairs. From an academic perspective, this research is also expected to fill a gap in the literature related to machine translation for Indonesian, and encourage further research in this area.

RESEARCH METHODS

This study adopts a descriptive qualitative research design. The purpose is to explore the strengths and weaknesses of DeepL as a machine translation tool for translating Indonesian texts into English. The qualitative approach allows the researcher to conduct an in-depth analysis of translated outputs, not only based on surface-level errors but also focusing on meaning, sentence naturalness, and cultural-contextual appropriateness. This method is well-suited to interpret the linguistic dimensions of machine-generated translation more richly and critically. The population in this study includes various types of Indonesian texts commonly translated into English by general users and professionals, such as academic texts, news reports, and general narratives.

For this research, the author selected 9 extract data as samples, consisting of: 3 academic texts (journal abstracts, introduction, and method); 3 online news articles from Indonesian mainstream media (Detiknews); and 3 general narrative texts (descriptive narratives). These samples were selected using purposive sampling, based on criteria such as sentence structure complexity, vocabulary richness, and cultural elements relevant for translation quality assessment.

Data collection was conducted using the following steps: (1) The selected source texts (in Indonesian) were entered into both DeepL Translator and Google Translate to obtain English translations, (2) The translation outputs were compiled and stored for analysis, (3) The researcher developed an observation checklist as the main instrument to evaluate the translations based on three key criteria: (a) Fidelity – the degree to which the translation maintains the meaning of the source text, (b) Fluency/Naturalness – the extent to which the translation is grammatically and stylistically natural in English, (c) Contextual Appropriateness – the accuracy of the translation in terms of cultural and pragmatic context.

To enhance the validity of the evaluation, two English language experts were consulted to cross-check and provide feedback on the translated outputs, especially regarding fluency and semantic clarity in the target language.

The data were analyzed using content analysis with an evaluative approach. Each translated output was coded according to the three main criteria: fidelity, fluency, and contextual appropriateness. The researcher identified recurring patterns, strengths, and distinctive weaknesses of DeepL Translator, and compared them qualitatively with those produced by Google Translate. Special attention was given to linguistic features that commonly present challenges in translation, such as: Passive

constructions, Indonesian affixation patterns, Idiomatic expressions, Use of the particle “yang”, Temporal and causal conjunctions.

Each finding was supported by actual translation excerpts and interpreted using the theoretical frameworks of equivalence (Nida & Taber, 1969), translation shifts (Catford, 1965), and naturalness (Newmark, 1988) to provide a structured and critical evaluation of the translation quality.

RESULTS AND DISCUSSION

Results

This section evaluates the translations of academic texts, with a focus on terminological precision, syntactic clarity, and fidelity to academic register. Both DeepL and Google Translate are examined based on their treatment of morphosyntactic structure and semantic nuance.

Academic Text 1

Original: Penelitian ini bertujuan untuk menjelaskan bentuk interferensi bidang morfologi dan sintaksis terhadap suatu tuturan.

DeepL: This study aims to explain the form of morphological and syntactic field interference in speech.

Google Translate: This study aims to explain the forms of interference in the fields of morphology and syntax in a speech act.

Both translations convey the core purpose of the study. Google offers better contextual appropriateness with the plural “forms of interference” and use of “speech act,” which is more terminologically accurate in linguistic discourse. DeepL’s phrasing “morphological and syntactic field interference” sounds overly literal and awkward in English.

- Strength (Google): Terminological accuracy with “speech act” and natural structure.
- Weakness (Google): Slight verbosity with “the fields of morphology and syntax.”
- Strength (DeepL): Concise phrasing that reflects source structure.
- Weakness (DeepL): “Field interference” is unnatural and may confuse target readers.

Academic Text 2

Original: Tujuan seorang penutur dalam berinteraksi dengan lawan bicaranya untuk menyampaikan pesan atau informasi dengan catatan jika lawan tutur menguasai bahasa yang digunakan oleh pembicara.

DeepL: The purpose of a speaker in interacting with his interlocutor to convey a message or information with a note if the interlocutor masters the language used by the speaker.

Google Translate: The aim of a speaker in interacting with the person he is talking to is to convey a message or information, provided that the person he is talking to has mastered the language used by the speaker.

Google’s translation is syntactically clearer and better preserves naturalness. The use of “provided that” is a more accurate rendering of dengan catatan, improving fidelity. DeepL’s phrasing is literal but awkward and lacks the cohesive marker that clarifies conditionality in the sentence.

- Strength (Google): Improved clarity and appropriate use of “provided that” for logical cohesion.
- Weakness (Google): Repetition of “the person he is talking to” could be more concise.
- Strength (DeepL): Reflects source structure and preserves key ideas.
- Weakness (DeepL): Lacks grammatical coherence and fluency in expressing the condition.

Academic Text 3

Original: Metode pengumpulan data yang digunakan dalam penelitian ini adalah metode simak yaitu pemerolehan data primer dengan cara menyimak percakapan oleh remaja dalam situasi pergaulan sehari-hari.

DeepL: The data collection method used in this research is the listening method, which is the acquisition of primary data by listening to conversations by teenagers in everyday social situations.

Google Translate: The data collection method used in this study is the listening method, namely obtaining primary data by listening to conversations between teenagers in everyday social situations.

Both translations retain the formal academic tone and accurately convey the procedural nature of the sentence, suggesting a high level of fidelity. However, issues in naturalness and contextual appropriateness arise due to awkward prepositional use and literal rendering of certain phrases.

DeepL's use of *"by listening to conversations by teenagers"* introduces syntactic awkwardness and ambiguity, as the phrase might imply that the teenagers are the ones doing the listening rather than being listened to. Meanwhile, Google Translate's phrasing—*"conversations between teenagers"*—is more contextually accurate, reflecting the intended meaning of passive observation in a social setting.

- Strength (Google): “Between teenagers” clearly conveys the participants of the conversation, enhancing contextual accuracy.
- Weakness (Google): The phrase “namely obtaining primary data” is a literal translation of “yaitu pemerolehan data primer” and feels stiff in English academic writing.
- Strength (DeepL): Maintains formal structure and preserves the explanatory clause structure typical of academic prose.
- Weakness (DeepL): “By teenagers” causes ambiguity and reduces grammatical clarity, affecting both naturalness and contextual appropriateness.

News Text 1

Original: Mereka pun mencoba memasukkan uang palsu buatan Hendra ke dalam mesin pendeteksi uang palsu, dan mesin tersebut berbunyi.

DeepL: They also tried to insert Hendra's counterfeit notes into a counterfeit detection machine, and the machine beeped.

Google Translate: They also tried to put fake money made by Hendra into the counterfeit money detection machine, and the machine rang.

Both translations show high fidelity in preserving the meaning and sequence of the original text. DeepL's choice of “beeped” is more contextually accurate for how

machines react, enhancing contextual appropriateness. Google’s use of “rang,” while semantically close, evokes a sound associated more with telephones or alarms, which slightly lowers naturalness.

- Strength (DeepL): “Beeped” is a precise and idiomatic description of a machine’s alert, maintaining contextual realism.
- Weakness (DeepL): The phrase “counterfeit detection machine” feels slightly awkward; “counterfeit detector” might be more natural.
- Strength (Google): Sentence maintains grammatical clarity and logical flow.
- Weakness (Google): “Rang” is less suitable in this context, affecting the technical accuracy of the machine's behavior.

News Text 2

Original: Semangat mencintai bumi dan budaya lokal diwujudkan secara nyata oleh Dara Baro.

DeepL: The spirit of loving the earth and local culture is realized by Dara Baro.

Google Translate: The spirit of loving the earth and local culture is made real by Dara Baro.

Both translations adequately preserve the fidelity of the message. However, DeepL's “is realized” aligns more closely with formal written English and enhances naturalness. Google’s “is made real” is understandable but slightly less idiomatic, and it introduces a marginally awkward phrasing for a formal news context.

- Strength (DeepL): “Is realized” offers a formal and elegant structure appropriate for news reporting.
- Weakness (DeepL): Slightly abstract; may feel vague without additional context on the action taken.
- Strength (Google): Captures the transformative essence of the action (“made real”).
- Weakness (Google): Phrase “made real” is more colloquial, reducing its appropriateness for journalistic style.

News Text 3

Original: Saat itu anggota curiga lantaran mobil tersebut dijual dengan harga murah.

DeepL: At that time the members were suspicious because the car was sold at a low price.

Google Translate: At that time, members were suspicious because the car was being sold at a cheap price.

Both translations convey the general message of the source text, indicating a high level of fidelity. However, differences in lexical choice—particularly “low price” (DeepL) vs “cheap price” (Google)—influence the naturalness and contextual tone of each version. DeepL opts for a more neutral expression (“low price”) which aligns with formal news reporting, while Google uses a more colloquial phrase (“cheap price”) that may sound less appropriate in journalistic contexts. Furthermore, Google’s progressive tense “was being sold” could imply an ongoing action, which is not clearly indicated in the original Indonesian sentence.

- Strength (DeepL): “Low price” maintains formal tone appropriate for news reporting, enhancing contextual appropriateness.

- Weakness (DeepL): Slight omission of the definite article before "members" may reduce grammatical naturalness ("the members").
- Strength (Google): Successfully captures suspicion tied to an unusual price, preserving the cause-effect logic.
- Weakness (Google): "Cheap price" is somewhat informal and redundant ("cheap" already implies "price"), reducing naturalness and formality. The use of "was being sold" may misrepresent the aspectual meaning of the source.

Narrative Text 1

Original: Kematianku tak lebih seperti saat seorang penyair menuliskan tanda titik pada akhir kalimat sajaknya.

DeepL: My death is nothing more than when a poet puts a full stop at the end of his rhyming sentence.

Google Translate: My death is no more than when a poet writes a period at the end of his verse.

Both translations retain the simile structure, suggesting a moderate level of fidelity. DeepL's rendering uses "rhyming sentence," which sounds forced and unnatural in English poetic discourse. In contrast, Google's "verse" aligns more closely with standard poetic terminology, contributing to improved naturalness and contextual appropriateness.

- Strength (Google): "Verse" maintains poetic diction appropriate to the source.
- Weakness (Google): "Writes a period" is more mechanical than metaphorical, weakening emotional tone.
- Strength (DeepL): Maintains metaphorical structure with a clear simile.
- Weakness (DeepL): "Rhyming sentence" lacks idiomatic fluency and diminishes poetic nuance.

Narrative Text 2

Original: Biarkan saja aku menyambutnya dengan rasa sahaja dan temu

DeepL: Just let me greet him with a sense of legitimacy and encounter

Google Translate: Just let me welcome it with a sense of humility and meeting

This sentence presents lexical and cultural challenges due to the abstract and philosophical nature of sahaja and temu. Google's "humility" captures the emotional humility implied by sahaja better than DeepL's "legitimacy," which is semantically off-track. However, both translations fall short in conveying the contemplative depth of temu.

- Strength (Google): "Humility" reflects emotional tone and cultural subtlety.
- Weakness (Google): "Meeting" is vague and fails to reflect the poetic or philosophical intent of temu.
- Strength (DeepL): "Encounter" attempts a more abstract rendering of temu.
- Weakness (DeepL): "Legitimacy" misinterprets sahaja, compromising contextual fidelity.

Narrative Text 3

Original: Buku-buku yang memandangkanku / Seolah tak rela menenggelamkanku dalam angan

DeepL: The books that looked at me / As if unwilling to drown me in wishful thinking

Google Translate: Books that look at me / As if not willing to drown me in dreams

Both versions retain the metaphor and overall meaning, reflecting a high level of fidelity. DeepL uses “wishful thinking,” which adds a more cognitive tone and slightly distances the metaphor from the soft, poetic nuance of *angan*. Google’s “dreams” is closer in denotation and contributes to a smoother, more natural reading, though arguably less interpretive.

- Strength (DeepL): “Wishful thinking” reflects the abstract intent of *angan*.
- Weakness (DeepL): Slightly cognitive tone weakens poetic softness (Newmark).
- Strength (Google): “Dreams” offers a simpler, more poetic equivalent.
- Weakness (Google): Slightly shallower emotional nuance compared to the original metaphor.

Discussion

The findings of this study highlight the strengths and limitations of DeepL as a machine translation tool when translating Indonesian texts into English. It is evident from a comparison of academic, journalism, and narrative text categories that DeepL exhibits varying performance in terms of correctness and linguistic nuance handling. This talk will go into further detail about how DeepL handles these issues and how well it performs in comparison to the standards of high-quality translation, particularly with regard to fidelity, contextual relevance, and naturalness.

DeepL exhibits a tendency to adhere closely to the source structure in the academic text segment, which frequently produces translations that seem stiff or out of place in English. Though it lacks communicative efficiency, this has a high degree of formal faithfulness. DeepL’s literal interpretation of some sentences, for example, may conflict with the style and clarity requirements of academic writing in English. These kinds of problems are especially pertinent in academic speech, where readability and fluidity must be matched with meaning clarity. As a result, even though DeepL retains the informational content, post-editing is frequently necessary to satisfy the linguistic and stylistic requirements of English-language study.

On the other hand, DeepL performs better in the news text. In this instance, the tool succeeds in assuming a register and tone more akin to professional journalism. Phrases such as “beeped in response” imply a sophisticated comprehension of genre-appropriate vocabulary choices. This suggests that the DeepL method is better suited to some semi-formal and factual text genres, where context is easier to deduce and sentence structures are more straightforward. Though DeepL can approximate tone, its syntactic correctness is still lacking, as evidenced by the odd grammatical errors that still occur, such as poor wording or incorrect passive constructions.

The translation of narrative texts, which are full of metaphor, passion, and cultural specificity, presents DeepL with the biggest problem. Abstract terms with deep philosophical or poetic meanings in Indonesian, such as *sahaja* and *temu*, are difficult for DeepL to understand. Even if they are occasionally linguistically correct, their English translations fall short of capturing the richness and nuance of the original. This highlights a further significant drawback of machine translation: its incapacity to sensitively understand metaphorical language and cultural subtext.

CONCLUSION

This research aimed to evaluate the accuracy and the treatment of linguistic nuances in machine translation outputs produced by DeepL for Indonesian-to-English translation. The findings paint a complex picture of DeepL's translation abilities. On the one hand, the system's accuracy at translating common grammatical structures and expressing literal meanings is comparatively high. In many instances, the syntactic structure is faithful or well adjusted to English rules, and the majority of the translated phrases maintain the essential semantic substance of the original text.

The limitations of DeepL, however, become apparent when the source sentences contain more intricate linguistic occurrences, such as pragmatic indicators, culturally ingrained references, idiomatic expressions, or changes in tone and register. These situations frequently result in translations that are overly literal, awkward, or improper for the context since the machine translation frequently misses the subtle nuances or the intended communicative effect.

The incapacity of DeepL to reliably capture pragmatic meaning, which is essential for efficient communication and includes the speaker's goal, the social context, or the implicit tone, is one of the main issues noted. For example, the target text usually flattens or misrepresents emotional nuances, emphasis, and politeness techniques. In a similar vein, occasionally literal translations of Indonesian idioms result in outputs that are grammatically acceptable but pragmatically incorrect or unclear to fluent English speakers.

These findings indicate that while DeepL can serve as a useful tool for general translation tasks and comprehension, it is not yet capable of fully replacing human translators, especially in tasks that demand deep cultural understanding, stylistic sensitivity, and interpretive judgment. The evaluation also underscores the importance of human oversight in contexts where precise meaning, tone, and cultural resonance are essential — such as academic writing, literary translation, and formal communication. In conclusion, DeepL's performance in translating from Indonesian to English reflects both the progress and the ongoing limitations of current machine translation technology. Its strength lies in syntactic processing and lexical correspondence, but its weakness remains in the nuanced, interpretive dimensions of language. Therefore, while it represents a valuable support tool, it should be used critically and supplemented with human expertise when high-quality translation is required.

BIBLIOGRAPHY

- Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasajo, R. E., Baldwin, T., Lau, J. H., & Ruder, S. (2022). *One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia* (Version 1). arXiv. <https://arxiv.org/abs/2203.13357>
- Ammariah, H., & Ammariah, H. (2025, April 23). 80 Contoh Puisi Pendek Berbagai Tema dan Makna. Belajar Gratis di Rumah Kapan Pun! Blog Ruangguru -. <https://www.ruangguru.com/blog/contoh-puisi>
- Ekuerhare, A., & Udoka, O. P. (2024). The effect of globalization on language services and translation strategies. *Deleted Journal*, 4(6), 507–516.

<https://doi.org/10.62225/2583049x.2024.4.4.6.3467>

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://jeef.unram.ac.id/index.php/jeef/article/view/681>

Rosaria, S., & Riana, R. (2021). Weaknesses of Translation Result Using Google Translate. *Proceedings of the 2nd International Seminar on Translation Studies, Applied Linguistics, Literature and Cultural Studies, STRUKTURAL 2020, 30 December 2020, Semarang, Indonesia*. Proceedings of the 2nd International Seminar on Translation Studies, Applied Linguistics, Literature and Cultural Studies, STRUKTURAL 2020, 30 December 2020, Semarang, Indonesia, Indonesia. <https://doi.org/10.4108/eai.30-12-2020.2311244>

Sakre, M. (2020). Machine translation status and its effect on business. *Deleted Journal*, 10(1), 33-48. <https://doi.org/10.21608/asc.2020.157421>

Sandra, R. A. (2018). From English to Indonesia: Translation Problems and Strategies of EFL Student Teachers - A Literature Review. *International Journal of Language Teaching and Education*, 2(1), 13–18. <https://doi.org/10.22437/ijolte.v2i1.4520>

Suhono, S., Zuniati, M., Pratiwi, W., & Hasyim, U. (2020). Clarifying Google Translate Problems Of Indonesia-English Translation Of Abstract Scientific Writing. *Proceedings of the Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia*. Proceedings of the 2nd Workshop on Multidisciplinary and Applications (WMA) 2018, 24-25 January 2018, Padang, Indonesia, Padang, Indonesia. <https://doi.org/10.4108/eai.24-1-2018.2292399>

Telaumbanua, Y. A., Marpaung, A., Gulo, C. P. D., Waruwu, D. K. W., Zalukhu, E., & Zai, N. P. (2024). Analysis of Two Translation Applications: Why is DeepL Translate more accurate than Google Translate? *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, 4(1), 82–86. <https://doi.org/10.59934/jaiea.v4i1.560>

Tuyet Phan, N. T., & Chen, C.-H. (2021). VIETNAMESE ENGINEERING STUDENTS' PERCEPTIONS OF THE USE OF GOOGLE TRANSLATION TOOL. *Journal of Science and Technology - IUH*, 48(6). <https://doi.org/10.46242/jst-iuh.v48i6.1610>