Wanastra: Jurnal Bahasa dan Sastra
Volume 18 No. 1 March 2026
P-ISSN 2086-6151 E-ISSN 2579-3438
https://doi.org/10.31294/wanastra.v18i2

Evaluating Naturalness in Machine Translation: A Case Study of DeepL on Philosophical Texts

Noer Hayati[1], Sri Wahyuni[2]

Universitas Gajayana Malang [1,2]
email: nur086835@gmail.com [1], sri.wahyuni@unigamalang.ac.id [2]
e-mail corresponding author: nur086835@gmail.com

Received : 28-02-2026
Revised : 13-03-2026
Accepted : 31-03-2026

Abstrak *– Perkembangan pesat Neural Machine Translation (NMT) telah secara signifikan mengubah praktik penerjemahan, khususnya dalam hal kecepatan dan aksesibilitas. Di antara berbagai sistem tersebut, DeepL Translator dikenal luas karena kemampuannya menghasilkan terjemahan yang lancar dan terdengar alami. Namun, kinerjanya dalam menerjemahkan wacana akademik yang kompleks masih belum banyak dikaji, terutama dari aspek kealamian (naturalness). Penelitian ini bertujuan untuk menganalisis tingkat kealamian terjemahan bahasa Indonesia yang dihasilkan oleh DeepL dalam menerjemahkan "What is a Speech Act?" karya John Searle. Penelitian ini menggunakan pendekatan kualitatif dengan desain deskriptif. Data berupa kata, frasa, klausa, dan kalimat yang diambil dari sepuluh paragraf yang dipilih secara acak dari teks sumber beserta hasil terjemahannya dalam bahasa Indonesia. Analisis dilakukan dengan mengacu pada konsep kealamian serta didukung oleh kerangka Translation Quality Assessment (TQA). Penilaian dilakukan menggunakan skala tiga tingkat, yaitu alami, kurang alami, dan tidak alami, berdasarkan aspek akurasi makna, pilihan leksikal, dan struktur sintaksis. Hasil penelitian menunjukkan bahwa sebagian besar terjemahan termasuk dalam kategori kurang alami, yang mengindikasikan adanya kesepadanan parsial. Terjemahan yang alami umumnya ditandai oleh akurasi makna, penggunaan terminologi yang tepat, serta struktur kalimat yang fleksibel. Sebaliknya, terjemahan yang kurang alami dan tidak alami ditandai oleh kecenderungan terjemahan literal, pilihan kata yang kurang tepat, struktur yang kaku, serta dalam beberapa kasus terjadi distorsi makna. Permasalahan ini terutama muncul pada bagian teks yang mengandung konsep abstrak dan struktur kalimat yang kompleks. Penelitian ini menyimpulkan bahwa meskipun DeepL efektif dalam menerjemahkan bagian teks yang sederhana secara struktural dan stabil secara terminologis, sistem ini masih memiliki keterbatasan dalam menangani makna yang bergantung pada konteks dan kompleksitas konseptual dalam teks akademik. Oleh karena itu, penyuntingan oleh manusia tetap diperlukan untuk memastikan akurasi dan kealamian terjemahan. Penelitian ini memberikan kontribusi terhadap kajian Translation Quality Assessment dengan menyajikan temuan empiris mengenai kinerja penerjemahan mesin dalam wacana akademik serta menegaskan pentingnya keseimbangan antara akurasi makna, ketepatan leksikal, dan adaptasi sintaksis.*

Kata Kunci: kealamian terjemahan, terjemahan yang dihasilkan mesin, Deep L, teks informatif

Abstract - The rapid development of Neural Machine Translation (NMT) has significantly transformed translation practices, particularly in terms of speed and accessibility. Among these systems, DeepL Translator is widely recognized for producing fluent and natural-sounding translations. However, its performance in translating complex academic discourse remains underexplored, particularly in terms of naturalness. This study aims to analyze the naturalness of Indonesian translations generated by DeepL in translating "What is a Speech Act?" by John Searle. This research employs a qualitative descriptive design. The data consist of words, phrases, clauses, and sentences selected from ten randomly sampled paragraphs of the source text and their corresponding Indonesian translations. The analysis is guided by the concept of naturalness and supported by Translation Quality Assessment frameworks. A three-point scale—natural, less natural, and unnatural—was used to evaluate the data based on semantic accuracy, lexical choice, and syntactic structure. The findings reveal that the majority of translations fall into the less natural category, indicating partial equivalence. Natural translations occur when semantic accuracy, appropriate terminology, and flexible sentence structure are achieved. In contrast, less natural and unnatural translations are characterized by literal rendering, inappropriate lexical choices, structural rigidity, and, in some cases, semantic distortion. These issues are particularly evident in segments involving abstract concepts and complex sentence structures. The study concludes that while DeepL demonstrates effectiveness in translating structurally simple and terminologically stable segments, it remains limited in handling context-

dependent meaning and conceptual complexity in academic texts. Therefore, human post-editing is necessary to ensure both accuracy and naturalness. This study contributes to Translation Quality Assessment by providing empirical insights into the performance of machine translation in academic discourse and highlighting the importance of balancing semantic accuracy, lexical appropriateness, and syntactic adaptation.

Key words: translation naturalness, machine-generated translation, Deep L, informative text

INTRODUCTION

Translating from one language to another was once a challenging process. In earlier periods, translation practices were predominantly carried out by human translators, who were required to carefully read, interpret, and reproduce texts to ensure accurate transfer of meaning into the target language (TL). To identify contextually appropriate lexical choices, translators often relied on manual consultation of dictionaries, which was a time-consuming process. This process required not only linguistic knowledge but also cultural understanding and contextual awareness (Larson, 1984). The rapid advancement of machine translation (MT) technology has significantly transformed translation practices, making the process faster, more efficient, and increasingly accessible (Fiederer & O'Brien, 2009). This technology has revolutionized how people communicate across languages and access global information. Among these tools, one platform that has attracted much attention is DeepL Translator. DeepL is often praised by users for producing translations that sound more fluent, natural, and contextually accurate compared to other MT tools.

Nevertheless, the performance of DeepL remains a subject of debate, particularly regarding the quality of its translation output. Although it employs advanced artificial intelligence and neural network systems to analyze meaning and structure, it still faces fundamental limitations, such as the absence of human intuition, cultural nuance, and background knowledge of the subject being translated. Moreover, DeepL operates in a relatively fixed input and output system, where users simply copy and paste text from the source language (SL) to the target language (TL). Unlike generative AI models that allow interactive prompting or contextual adjustment, it does not accommodate user guidance or clarification during the translation process. It is important to note that translation is not merely a process of replacing words from one language to another but a complex act of transferring meaning from the SL to the TL while preserving the message, tone, and intention of the original text (Munday, 2001). In doing so, translators must follow a series of translation processes. Human translation typically involves a series of interrelated stages, including comprehension of the source text, selection of appropriate translation strategies, and restructuring to ensure accuracy, readability, and naturalness in the target language.

Unlike human translators who consciously follow stages such as comprehension, drafting, and revision, DeepL operates through automated computational processes based on neural network models, in which linguistic patterns are processed implicitly without explicit user-controlled stages. As a result, the output may exhibit linguistic features such as rigid syntactic structures, less idiomatic expressions, or limited contextual adaptation, which can indicate its machine-generated nature. A good translation, according to Catford (1965), should show equivalents that align with both meaning and cultural context. Meaning should always take precedence over literal equivalence to ensure that the message is conveyed naturally and appropriately for the target audience. This complexity arises because each language embodies its own system, structure, and worldview. Words or expressions in one language may not always have a direct equivalent in another. Therefore, translators should interpret not only the literal meaning but also the context in which a word or phrase is used. Idioms, figurative expressions, and discipline-specific terms, for instance, demand deeper comprehension and contextual sensitivity (Baker, 2018). Hence, professional translators require adequate background knowledge of the subject matter to ensure accurate and contextually appropriate translations.

Given this complexity, conducting an in-depth Translation Quality Assessment (TQA) of DeepL output is essential, especially when translating informative academic texts. Such an assessment is crucial to determine whether the translation successfully conveys the intended message or important information of the source text while maintaining naturalness. In this study, 'naturalness' refers to how well a translation reads as if it were originally written in the TL. According to Nida (1964), a good translation "aims at complete naturalness of expression," meaning that the target text should sound natural and culturally appropriate to the receptor language. Similarly, Wisudawanto & Susylowati (2024) defines naturalness as the extent to which a translation reads fluently and idiomatically without sounding like a translated text. A natural translation not only conveys accurate meaning but also flows smoothly and sounds idiomatic to native readers. Conversely, an unnatural translation tends to sound awkward, mechanical, or slightly off in tone and style, making it evident that the text has been translated.

Several studies have examined the use of machine translation in academic and educational contexts, particularly focusing on translation quality and user perception. Laksana & Komara (2024) investigated Indonesian EFL students' perceptions of DeepL and found that the tool is widely used due to its fluency, efficiency, and ease of use. However, their study also emphasized that DeepL still requires human evaluation, especially when translating complex academic texts, as users often encounter problems related to context and meaning accuracy. From the perspective of translation naturalness, Aisyah & Harjanti (2020) analyzed the Indonesian translation of "Harry Potter" and "The Half-Blood Prince" and reported that although many translated sentences were natural, several instances of unnatural translation occurred due to literal rendering and inappropriate lexical choices. Their findings indicate that naturalness remains a crucial aspect of translation quality, even in professionally translated literary works.

Furthermore, Wisudawanto & Susylowati (2024) emphasized that naturalness is an essential parameter in translation quality assessment from a pragmatic perspective. Their study demonstrated that a translation may be semantically accurate but still considered unnatural if it fails to reflect appropriate linguistic and communicative norms of the target language. This highlights the importance of evaluating translation beyond accuracy alone. However, existing studies have predominantly focused on user perception and literary texts, leaving machine-generated translations of academic and philosophical discourse—characterized by abstract concepts, technical terminology, and complex structures—significantly underexplored.

When combined, these findings indicate that although DeepL frequently performs better than other MT systems in terms of perceived naturalness and fluency, there are still substantial limits, especially when it comes to rendering context-specific meanings, idiomatic expressions, and abstract concepts. Most previous studies have focused on either human-translated works or non-academic text genres, as well as user perceptions of DeepL, rather than examining its performance in highly abstract academic texts. Consequently, limited attention has been given to how DeepL processes academic texts that require accuracy, coherence, and naturalness. In order to close this gap, the present study evaluates the level of naturalness in DeepL translation of "What is a Speech Act?" by Searle and pinpoints the factors influencing differences in naturalness. By addressing this gap, the present study contributes to the field of Translation Quality Assessment by providing a systematic analysis of naturalness in machine-generated academic translation, as well as offering practical insights for translation practice and the development of machine translation systems. Therefore, this study specifically examines the naturalness of DeepL's Indonesian translation of John Searle's 'What is a Speech Act?' by analyzing lexical choice, sentence structure, and semantic accuracy, in order to identify its strengths and limitations in rendering academic discourse.

1. Machine-Generated Translation

Machine-generated translation refers to the automated process of transferring meaning from a source language (SL) to a target language (TL) using computational systems without direct human intervention. The emergence of Neural Machine Translation (NMT) has significantly improved translation quality compared to earlier rule-based and statistical models, particularly in terms of fluency and contextual coherence (Ashraf, 2024). By processing entire sentences rather than isolated lexical units, NMT systems are able to capture broader contextual relationships and generate more natural-sounding outputs.

However, despite these advancements, the effectiveness of NMT remains contested. While it performs well in producing grammatically acceptable sentences, several scholars argue that it still lacks deep semantic and pragmatic understanding (Naveen & Trojovský, 2024). This limitation stems from its reliance on probabilistic pattern recognition rather than genuine conceptual interpretation (Ahmed et al., 2023). As a result, machine-generated translations may appear fluent on the surface but fail to accurately convey implicit meaning, logical relations, or authorial intent—especially in complex academic discourse.

Furthermore, studies such as Chen & Lin (2025) highlight that NMT systems often resort to literal translation when encountering unfamiliar structures or domain-specific expressions. This tendency reflects what can be interpreted, in Venuti (1995) terms, as a form of foreignization, where traces of the source language remain visible in the target text, resulting in reduced naturalness. In contrast, human translators tend to adopt domestication strategies to produce more natural and culturally appropriate expressions. This distinction indicates that the issue of naturalness in machine translation is not merely technical but also theoretical, involving translation strategies and ideological choices.

2. Informative Text in Translation

Informative texts are primarily designed to convey factual information, knowledge, or conceptual explanations in a clear and objective manner (Newmark, 1988). Unlike expressive or persuasive texts, informative texts prioritize semantic accuracy, clarity, and logical organization. This characteristic makes them particularly demanding in translation, as even minor deviations can lead to distortion of meaning.

From a translation perspective, Newmark (1988) emphasizes that informative texts require a semantic translation approach, where accuracy and terminological consistency take precedence over stylistic freedom. However, this perspective can be contrasted with functionalist and descriptive approaches, such as Toury's (1980) Descriptive Translation Studies, which argue that translations should also be analyzed based on their acceptability within the target culture. This suggests that achieving accuracy alone is insufficient; the translation must also conform to the linguistic norms and expectations of the target audience.

In academic and philosophical texts, these challenges are intensified by the presence of abstract concepts, dense argumentation, and discipline-specific terminology (Taye & Mengesha, 2024). Meaning is often constructed through logical relationships across sentences rather than isolated units. Consequently, a translation that fails to preserve these relationships may disrupt coherence and reduce comprehensibility. This makes informative texts a critical domain for evaluating machine translation performance, particularly in balancing accuracy and naturalness. Given these complexities, a systematic framework is required to evaluate how effectively such texts are translated. This leads to the relevance of Translation Quality Assessment (TQA) as an analytical tool.

3. Translation Quality Assessment and Naturalness

Translation Quality Assessment (TQA) provides a systematic framework for evaluating translation performance. One influential model is proposed by House (2015), who distinguishes between overt and covert translation and emphasizes functional equivalence between the source and target texts. In this framework, naturalness is closely related to the concept of a covert translation, where the target text reads as if it were originally produced in the target language.

Naturalness itself is widely recognized as a key parameter in translation quality. Nida (1964) conceptualizes this through the notion of dynamic equivalence, where the translation should produce a similar effect on the target audience. This idea is further operationalized by Nababan et al. (2012), who define naturalness as the degree to which a translation conforms to the grammatical, lexical, and stylistic norms of the target language.

However, naturalness should not be viewed in isolation. It must be understood in relation to other dimensions of translation quality, particularly accuracy and acceptability. A translation may achieve high accuracy but still be considered inadequate if it sounds unnatural or overly influenced by the source language. Conversely, excessive focus on naturalness may lead to loss of semantic precision. This tension reflects an ongoing debate in translation studies regarding the balance between fidelity and readability, where prioritizing one dimension may inevitably compromise the other. This issue becomes particularly critical in machine translation, where such decisions are not made consciously but emerge from algorithmic processing.

In the context of machine translation, this issue becomes more complex. Unlike human translators, machine systems do not consciously apply translation strategies or consider audience reception. Instead, they generate output based on learned statistical patterns. From the perspective of Toury’s descriptive framework, this raises questions about whether machine-generated translations can truly conform to target-language norms or merely approximate them. Therefore, evaluating naturalness in machine-generated translation, particularly in complex academic texts, is essential to determine the extent to which such systems can achieve functional and communicative equivalence in academic discourse. This perspective provides the analytical foundation for the present study.

RESEARCH METHODOLOGY

This study is a qualitative one with a descriptive design because it aims to describe the naturalness of machine-generated translation as it is, rather than to examine it statistically. A qualitative approach was considered appropriate because the focus of the study is a social phenomenon in the form of linguistic matters, particularly the quality of translation in terms of meaning, word choice, and sentence structure. The qualitative approach allows the researchers to analyze the data and interpret the findings in depth when exploring how meaning is constructed in the TL text, which leads to the naturalness of the text.

The data of this study were SL words, phrases, clauses, and sentences and their corresponding rendering results in the TL, which indicated naturalness of the translation. They were taken from Searle's Speech Acts, a well-known philosophical and linguistic text that discusses the theory of speech acts. This text was selected because it represents an informative and theoretical academic text, characterized by abstract concepts, technical terminology, and complex sentence structures. The ten paragraphs were selected using a simple random sampling technique, in which each paragraph in the source text had an equal probability of being chosen. The selection of ten paragraphs was considered sufficient to represent the linguistic characteristics of the text while allowing for in-depth qualitative analysis. They were translated into Indonesian using DeepL machine translation (document translation feature).

The data were collected through systematic steps. The source text was obtained from a published academic source. It was read attentively. Ten paragraphs were selected as the representative sample of the SL data, and they were translated into Indonesian using DeepL's document translation feature. The Indonesian TL text produced by DeepL was carefully read to obtain the corresponding data. Thus, there were two sets of data, namely SL data and their corresponding TL ones.

The collected data were analyzed based on themes and categories that reflect the questions of the study. The analysis of naturalness was guided by the concept of naturalness proposed by Nida (1964), which emphasizes the use of natural and acceptable expressions in the target language. In addition, the evaluation was supported by the Translation Quality Assessment model proposed by (House, 2015). A three-point scale was used to assess the level of naturalness: (3) natural, (2) less natural, and (1) unnatural. This scale was applied to each unit of analysis based on its fluency, readability, and acceptability in the target language. The SL and TL texts were aligned paragraph by paragraph. The TL text was carefully read and compared with the original text to identify linguistic features related to naturalness. The data analysis was conducted through several steps: (1) aligning the SL and TL texts, (2) identifying units of analysis in the form of words, phrases, clauses, and sentences, (3) evaluating their naturalness using the established scale, (4) categorizing the findings based on the level of naturalness, and (5) interpreting the results in relation to the theoretical framework. To ensure the reliability of the analysis, a second rater with a background in translation studies was involved in assessing the data. The results from both raters were compared and discussed to reach agreement, thereby minimizing subjectivity.

RESULTS AND DISCUSSION

1. Naturalness of the translation

The analysis reveals three levels of translation naturalness: natural, less natural, and unnatural. Rather than presenting each paragraph as isolated cases, the findings are synthesized to highlight recurring linguistic patterns influencing translation quality.

Table 1. Summary of Naturalness Classification

Category	Paragraphs	Key Issues Identified
Natural	5, 6, 8	Accurate meaning, appropriate terminology, fluent structure
Less Natural	1, 2, 3, 4	Minor lexical issues, awkward phrasing, partial fluency disruption
Unnatural	7, 9, 10	Meaning distortion, incorrect terminology, structural rigidity

As shown in Table 1, the distribution indicates that natural translations are relatively limited, while most instances fall into the less natural category. This suggests that the translation system demonstrates partial competence, performing effectively in structurally simple segments but encountering difficulties when processing more abstract and conceptually dense material. The occurrence of unnatural translations in later paragraphs further reinforces the impact of textual complexity on translation quality.

A. Natural translation

The translations in paragraphs 5, 6, and 8 demonstrate a high level of naturalness, characterized by accurate meaning transfer and acceptable linguistic form. These findings align with dynamic equivalence, which emphasizes the reproduction of meaning and communicative effect in the target language (Nida, 1964). A central feature of this category is the appropriate use of academic terminology. Expressions such as "aturan semantic", "aturan konstitutif", and "tindak ilokusi" reflect strong alignment with Indonesian scholarly conventions. This indicates that the translation system performs effectively when dealing with standardized or widely recognized terminology, supporting the notion that terminological stability enhances translation quality (Baker, 2018).

In paragraph 5, the logical structure of the philosophical argument is preserved, particularly in the discussion of semantic rules. The translation maintains the progression of ideas, demonstrating strong semantic equivalence. Minor issues, such as redundancy ("aturan penggunaan atau penggunaannya") and slightly cumbersome phrasing, indicate limitations at the stylistic level. However, these do not significantly affect comprehension, suggesting that naturalness is primarily determined by meaning clarity rather than stylistic refinement.

A similar pattern is observed in paragraph 6, where the distinction between regulative and constitutive rules is conveyed accurately. Although expressions such as "hubungan ini ada secara independen" reflect mild structural interference, the translation remains clear and coherent. This supports the argument that semantic accuracy can compensate for minor syntactic rigidity.

Paragraph 8 further illustrates the system's ability to preserve both argumentation and exemplification. The formulation "X counts as Y" is rendered understandably, although the literal translation "X dihitung sebagai Y" could be improved to "X dianggap sebagai Y". This example highlights the importance of context-sensitive lexical choice in achieving full naturalness (Newmark, 1988). Overall, the natural category demonstrates that high translation quality emerges when semantic accuracy, terminological adequacy, and acceptable fluency operate simultaneously. Minor stylistic imperfections do not significantly reduce naturalness as long as communicative effectiveness is maintained.

B. Less natural translation

As indicated in Table 1, the less natural category represents the largest portion of the data, suggesting that machine translation frequently achieves only partial equivalence. In these cases, meaning is generally preserved, but naturalness is reduced due to lexical inaccuracies and structural interference. A dominant pattern in this category is source-language interference, where English structures are transferred directly into Indonesian. This is evident in expressions such as "mengatur panggung" (for set the stage) and "produksi tanda" (for production of the token). These translations reflect a reliance on formal equivalence rather than communicative adaptation, resulting in texts that are grammatically acceptable but stylistically unnatural.

This tendency aligns with Toury's concept of adequacy-oriented translation, where translations remain closer to the source text at the expense of target-language norms (Toury, 1980). As a result, although the propositional meaning is accessible, the reading process becomes less fluent and requires greater cognitive effort. Lexical choice also emerges as a critical issue. For example, translating knight as "ksatria" instead of "kuda" demonstrates a lack of contextual awareness, while "menjengkelkan" introduces an informal tone inconsistent with academic discourse. These examples highlight the importance of pragmatic appropriateness, which requires sensitivity to both context and register (Baker, 2018).

Syntactically, many translations in this category retain English sentence patterns, resulting in rigidity. Phrases such as "pembicara secara khas akan bergerak rahang dan lidahnya" illustrate insufficient restructuring. This supports the argument that naturalness depends on syntactic adaptation, not merely grammatical correctness (Newmark, 1988). Despite these issues, the overall message remains understandable. This indicates that less natural translations occupy an intermediate position, where semantic equivalence is achieved but communicative naturalness is only partially realized.

C. Unnatural translation

The unnatural category represents the most severe level of translation deficiency, where errors significantly affect both meaning and readability. As shown in Table 1, these translations are characterized by semantic distortion, incorrect terminology, and lack of coherence. In paragraph 7, the omission of touchdown and its replacement with "skakmat" disrupt the intended comparison, leading to conceptual confusion. Similarly, in paragraph 10, the mistranslation of predicate as "memprediksikan" reflects a misunderstanding of key linguistic concepts. These errors indicate a breakdown of semantic equivalence, where the core meaning of the source text is not preserved.

From the perspective of House's TQA model, such errors represent failures in both overt and covert translation, as the text neither accurately conveys meaning nor conforms to target-language norms (House, 2015). Additionally, expressions such as "pertunjukan tindakan ilokusi" introduce misleading connotations, reflecting inappropriate lexical choices influenced by literal translation strategies. Another critical issue is textual inconsistency, including repetition and omission. The duplication of "skakmat" and omission of touchdown reduce coherence and disrupt logical relationships within the text. This aligns with Baker's (2018) emphasis on cohesion and coherence as essential components of translation quality. These findings suggest that machine

translation systems struggle significantly when dealing with abstract, theory-driven discourse, where meaning depends heavily on conceptual precision and contextual interpretation.

Discussion

The findings indicate that the naturalness of machine-generated translation is primarily determined by the interaction of three key linguistic aspects examined in this study, namely semantic accuracy, lexical choice, and sentence structure. These aspects do not function independently; rather, they collectively influence how effectively the translated text conveys meaning while adhering to the linguistic norms of the target language. This confirms that naturalness is a multidimensional construct, closely related to communicative effectiveness and contextual appropriateness (Baker, 2018).

The results show that high levels of naturalness are achieved when semantic accuracy is maintained alongside appropriate lexical selection and flexible syntactic restructuring. In such cases, the translation is able to preserve the intended meaning of the source text while producing expressions that sound natural and acceptable in Indonesian academic discourse. This finding supports the principle of dynamic equivalence, which emphasizes the importance of conveying meaning and communicative effect rather than maintaining formal similarity (Nida, 1964). It also reinforces the argument that accurate and context-sensitive terminology is essential in translating informative and academic texts (Newmark, 1988).

However, the findings also reveal that naturalness decreases when the translation system encounters idiomatic expressions and structurally complex sentences. In these situations, the system tends to transfer source-language patterns directly into the target language, resulting in literal translations and rigid sentence constructions. Although the general meaning is often still accessible, the lack of syntactic adaptation and inappropriate lexical choices reduce fluency and readability. This pattern reflects a tendency toward adequacy-oriented translation, where the translation remains closer to the source text at the expense of target-language naturalness (Toury, 1980).

More importantly, the study identifies cases were inaccuracies in lexical choice led to semantic distortion, particularly in the translation of key theoretical terms. These errors demonstrate that the system struggles to interpret context-dependent meaning, especially in abstract and philosophical discourse. This finding highlights a fundamental limitation of machine translation, which relies on pattern recognition rather than deep conceptual understanding. As a result, naturalness cannot be separated from semantic accuracy, as failure to convey precise meaning directly affects the clarity and coherence of the translation (House, 2015).

In addition, sentence structure plays a significant role in determining naturalness. Translations that retain source-language syntactic patterns tend to appear rigid and less fluent, even when the meaning is generally preserved. This supports the view that naturalness depends not only on grammatical correctness but also on the ability to restructure sentences according to target-language conventions (Newmark, 1988). Therefore, effective syntactic adaptation is necessary to ensure that the translation reads smoothly and naturally.

In relation to previous studies, these findings partially support earlier research highlighting the fluency and usability of machine translation tools in academic contexts (Laksana & Komara, 2024). However, the present study demonstrates that such performance is not consistent across different types of texts. In contrast to studies reporting generally high levels of naturalness in literary or less complex texts Aisyah & Harjanti (2020), this research shows that academic and philosophical texts pose greater challenges due to their conceptual density and abstract nature. This suggests that text type and complexity are critical factors influencing machine translation quality, as also emphasized by (Wisudawanto & Susylowati, 2024).

These findings have both theoretical and practical implications. Theoretically, this study reinforces the view that translation naturalness is closely linked to semantic and functional equivalence, as proposed in Translation Quality Assessment frameworks. Practically, the results suggest that while DeepL can be effectively used for translating general or moderately complex texts, human revision remains essential when dealing with specialized academic discourse. Particular attention should be given to lexical accuracy and sentence restructuring to ensure that the translated text maintains both clarity and naturalness.

Overall, this study confirms that achieving naturalness in machine-generated translation involves balancing meaning accuracy, appropriate lexical choice, and flexible sentence structure. When these elements are successfully integrated, the translation approaches target-language norms. However, when one or more of these aspects are compromised, the resulting translation tends to become less natural or even difficult to understand, especially in texts that require high levels of conceptual precision.

CONCLUSION

This study examined the naturalness of DeepL's Indonesian translation of John Searle's "What is a Speech Act?" by focusing on semantic accuracy, lexical choice, and sentence structure. The findings indicate that DeepL produces natural translations when handling structurally simple texts with clear meanings, as reflected in accurate semantic transfer, appropriate terminology, and acceptable sentence structure. However, its performance declines when processing idiomatic expressions, complex structures, and context-dependent meanings, resulting in literal rendering, lexical inaccuracies, and reduced fluency.

These results demonstrate that translation naturalness is shaped by the interaction of semantic accuracy, lexical choice, and syntactic adaptation. From a theoretical perspective, this study reinforces the importance of integrating functional and semantic equivalence in Translation Quality Assessment. From a practical perspective, it suggests that while DeepL can assist in translating general texts, human post-editing remains necessary to ensure accuracy and naturalness in academic discourse. Future research is recommended to examine larger datasets and different text types to further explore the limitations of machine-generated translation.

REFERENCE

Ahmed, M., Ouda, A., Abusharkh, M., Kohli, S., & Rai, K. (2023). An Optimized Approach to Translate Technical Patents from English to Japanese Using Machine Translation Models. *Applied Sciences*, *13*(12), 7126. https://doi.org/10.3390/app13127126

Aisyah, S., & Harjanti, D. (2020). A Naturalness of Translation in J. K. Rowling's Novel Harry Potter and the Half Blood Prince into Indonesian. *Al-Lisan: Jurnal Bahasa*, *5*(2), 200–214. https://doi.org/10.30603/al.v6i2.1331

Ashraf, M. (2024). Innovations and challenges in neural machine translation: A review. *International Journal of Science and Research (IJSR)*, *13*(10). https://doi.org/10.21275/SR241008113336

Baker, M. (2018). *In Other Words A Coursebook on Translation* (3rd Edition (ed.)). Routledge.

Catford, J. C. (1965). *A Linguistic Theory of Translation*. Oxford University Press. https://ia601602.us.archive.org/13/items/J.C.CatfordALinguisticTheoryOfTranslationOxfordUniv.Press1965/j.c. catford-a linguistic theory of translation-oxford univ. press %281965%29.pdf

Chen, S., & Lin, Y. (2025). A multidimensional comparison of ChatGPT, Google Translate, and DeepL in Chinese tourism texts translation: fidelity, fluency, cultural sensitivity, and persuasiveness. *Frontiers in Artificial Intelligence*. https://doi.org/10.3389/frai.2025.1619489

Fiederer, R., & O'Brien, S. (2009). Quality and Machine Translation: A realistic objective? *Journal of Specialised Translation*, *11*, 52–74. https://doi.org/10.26034/cm.jostrans.2009.639

House, J. (2015). *Translation Quality Assessment Past and Present* (1st ed.). Routledge.

Laksana, K. N., & Komara, C. (2024). Indonesian EFL Students' Perceptions of DeepL Machine Translation Tool: Utilization, Advantages, and Disadvantages. *JOLLS: Journal of Language and Literature Studies*, *4*(2), 256–276. https://doi.org/10.36312/jolls.v4i2.1931

Larson, M. L. (1984). *Meaning-based Translation: A Guide to Cross-language Equivalence*. University Press of America.

Munday, J. (2001). *Introducing translation studies: Theories and applications*. Routledge.

Nababan, M., Nuraeni, A., & Sumardiono. (2012). Pengembangan model penilaian kualitas terjemahan. *Kajian Linguistik Dan Sastra*, *24*(1), 39–57.

Naveen, P., & Trojovský, P. (2024). Overview and challenges of machine translation for contextually appropriate translations. *IScience*, *27*(10), 110878. https://doi.org/10.1016/j.isci.2024.110878

Newmark, P. (1988). *A textbook of translation*. Peter Newmark.

Nida, E. A. (1964). *Toward a science of translating: With special reference to principles and procedures involved in Bible translating*. Brill.

Taye, T., & Mengesha, M. (2024). Identifying and analyzing common English writing challenges among regular undergraduate students. *Heliyon*, *10*(17). https://doi.org/10.1016/j.heliyon.2024.e36876

Toury, G. (1980). *In Search of a Theory of Translation*. Tel Aviv, Israel: Porter Institute.

Venuti, L. (1995). *The Translator's Invisibility A History of Translation* (1st ed.). Routledge.

Wisudawanto, R., & Susylowati, E. (2024). Naturalness translation: A parameter of translation quality assessment in a pragmatic perspective. *Journal of English Language Teaching and Learning (JETLE)*, *5*(2), 198–204. https://doi.org/10.18860/jetle.v5i2.26483