Pragmatic Functional Categories as a Framework for Assessing ChatGPT's Translation Quality in Fashion Slogans

Paramita Widya Hapsari*[1], Kennia Wikanditha[2], Anita Rusjayanti[3]

* paramita.hapsari@staff.uns.ac.id

[1] Sekolah Vokasi, Universitas Sebelas Maret, Surakarta, Indonesia

ABSTRACT

The emergence of generative AI models such as ChatGPT has transformed translation practices, yet their capacity to handle creative and culturally embedded texts remains questionable. This study investigates ChatGPT's translation quality in fashion slogans by integrating Molina and Albir's (2002) translation technique taxonomy with Nababan et al.'s (2012) Translation Quality Assessment (TQA) framework. Employing a descriptive qualitative design complemented by quantitative scoring, the research analyzes 175 English–Indonesian fashion slogans produced by ChatGPT, focusing on three dimensions: accuracy, acceptability, and readability. The evaluation results reveal high accuracy (M = 2.71), indicating strong semantic retention, while acceptability (M = 2.54) and readability (M = 2.62) remain at moderate–high levels, suggesting occasional stylistic and pragmatic incongruence in Indonesian. Genre-based and brand-type analyses further indicate that global slogans tend to be more semantically precise yet less idiomatic compared to local ones. Although ChatGPT successfully preserves literal meaning, it frequently underperforms in maintaining rhetorical, emotional, and cultural resonance. The findings highlight ChatGPT's potential as a supportive translation tool for marketing discourse but reaffirm the necessity of human post-editing to achieve persuasive equivalence. This research contributes to translation studies by offering an integrated analytical framework that bridges micro-level translation techniques with macro-level quality dimensions, expanding current understanding of AI-mediated translation performance in creative and commercial contexts.

Keywords: Artificial intelligence; ChatGPT; Fashion slogans; Machine translation; Translation quality

INTRODUCTION

The rapid development of artificial intelligence (AI) has significantly reshaped translation practices in the past decade. With the advancement of neural machine translation (NMT) and large language models (LLMs) such as OpenAI's ChatGPT, translation processes are no longer limited to human cognitive capabilities but are increasingly supported by generative systems capable of producing fluent and contextually relevant outputs (Toral, 2022). While such systems have demonstrated impressive performance in translating technical and general texts, their ability to handle creative and culturally embedded discourses—such as advertising slogans—remains debatable. In fashion marketing, slogans function as condensed communicative acts that encapsulate brand identity, emotional appeal, and cultural resonance (Guidère, 2010). Translating these slogans therefore involves not only lexical and grammatical accuracy but also stylistic adaptation and pragmatic appropriateness. Theoretically, this challenge situates AI translation within the complex interplay between linguistic meaning and socio-cultural semiotics, making it an important case for empirical investigation in translation studies (House, 2015).

Within the theoretical tradition of translation studies, translation is understood as a communicative act rather than a mere linguistic transfer. Classical approaches such as Nida's (1964) dynamic equivalence and House's (2015) functional equivalence emphasize that an effective translation must recreate both the meaning and the communicative effect of the source text. This perspective aligns closely with Molina and Albir's (2002) taxonomy of translation techniques, which provides a micro-level analytical framework for describing the translator's decision-making process. Their classification—comprising strategies such as modulation, adaptation, amplification, and transposition—offers a systematic method to examine how meaning is restructured across linguistic and cultural systems. When applied to fashion slogans, these techniques become essential tools for analyzing whether AI systems can perform contextually appropriate transformations that go beyond literal transfer.
Empirical studies have recently sought to evaluate AI-generated translations using both linguistic and functional criteria. Research by Popović (2020) and Toral (2022) indicates that NMT systems often achieve near-human fluency in literal or structurally predictable texts but still falter when dealing with idiomatic, metaphorical, or culture-dependent expressions (Castilho et al., 2018). These findings resonate with Molina and Albir's (2002) argument that translation techniques should be chosen according to textual function and communicative intention rather than lexical equivalence alone. Meanwhile, within the domain of Translation Quality Assessment (TQA), Nababan, Nuraeni, and Sumardiono (2012) introduced a comprehensive model to evaluate translation performance across three key dimensions—accuracy, acceptability, and readability. This framework has since been adopted extensively in both academic and professional translation contexts, allowing evaluators to balance linguistic precision, stylistic naturalness, and cultural appropriateness.

Although numerous studies have applied the TQA framework to human translation in domains such as literature, tourism, and journalism (Sari, Nababan, & Djatmika, 2020), its application to AI-generated translations—especially in creative and commercial genres—remains underexplored. The interaction between Molina & Albir's micro-level translation techniques and Nababan's macro-level quality criteria offers a comprehensive analytical lens to assess not only *how* AI systems translate but also *how well* they preserve meaning, tone, and communicative intent. In fashion slogans, this dual analysis is particularly relevant: the translator (or AI system) must maintain the brand's persuasive tone while adapting linguistic form to local cultural sensibilities. A slogan such as "Dress the way you feel" requires both semantic fidelity and pragmatic adjustment to sound appealing and natural in Indonesian. Thus, integrating these two theoretical frameworks allows for a holistic evaluation of ChatGPT's translation behavior—connecting process-oriented and product-oriented dimensions of translation quality.

Despite the rapid rise of AI-assisted translation tools, there remains a scarcity of studies examining their performance in culturally charged text types. Much of the existing literature has focused on functional accuracy and productivity metrics (Castilho et al., 2018; Popović, 2020), while the aesthetic, affective, and cultural dimensions of AI translation remain understudied. Moreover, the intersection between translation techniques and quality outcomes has not been systematically explored through empirical data involving generative AI. This research gap is particularly pronounced in fashion marketing, where slogans serve as micro-texts with high emotional and symbolic density. The lack of scholarship on how LLM-based translators, such as ChatGPT, handle these creative and affective elements underscores the need for a detailed analytical approach that bridges linguistic description with quality evaluation.

Addressing this gap, the present study aims to evaluate ChatGPT's translation quality in fashion slogans through an integrative approach combining Molina & Albir's (2002) translation techniques and Nababan et al.'s (2012) TQA model. The analysis focuses on three quality dimensions—accuracy, acceptability, and readability—while mapping them against specific translation techniques to identify patterns of adequacy and inadequacy. Additionally, the study introduces an NLP-based categorization of True Positives (TP) and False Positives (FP) to support the qualitative evaluation with computational insights. This hybrid design not only advances methodological rigor in translation quality research but also contributes to the growing discourse on the creative potential and limitations of generative AI in cross-cultural communication. The novelty of this research lies in its focus on *fashion slogans as a testing ground* for AI's cross-cultural communicative competence—a domain that combines linguistic creativity, brand identity, and socio-cultural adaptation. The findings are expected to enrich both theoretical and practical understandings of how AI systems negotiate meaning, culture, and persuasion within translation.

METHOD

This study adopts a descriptive qualitative design with an integrated quantitative scoring component, aiming to evaluate the translation quality of fashion slogans generated by ChatGPT. The qualitative element is used to interpret the linguistic, cultural, and stylistic nuances of the translations, while the quantitative component operationalizes translation quality through systematic scoring on three dimensions: accuracy, acceptability, and readability (AAR). This hybrid design enables a more comprehensive assessment by combining the interpretive strength of qualitative inquiry with the objectivity of numeric evaluation. The study focuses on short, persuasive, and creative texts—fashion slogans—that demand both semantic fidelity and cultural adaptation in translation.

The data comprise 175 fashion slogans originally written in English and officially released by various global and local fashion brands. These include internationally recognized brands such as Nike, Adidas, Gucci, Levi's, H&M, and Zara, as well as emerging local labels targeting Indonesian consumers. The slogans represent a variety of brand prominence (global vs. local), genre diversity (motivational, descriptive, inclusive, luxurious, or playful), and target audiences (teenagers, adults, professionals, or families).
Each English slogan was translated into Indonesian using ChatGPT (GPT-4). To maintain procedural consistency, all translations were generated using the same prompt:
"Translate the following fashion slogan into Bahasa Indonesia. Consider maintaining its persuasive style, clarity, and cultural suitability for Indonesian audiences."
This uniformity ensured that stylistic and semantic variations in the output would stem from ChatGPT's translation process rather than inconsistencies in human prompting. Each entry in the dataset therefore includes the English source slogan, the ChatGPT-produced Indonesian translation, brand identity, genre classification, audience type, and slogan frequency (measured via Google search results). These variables allowed for a multidimensional analysis of how translation quality may vary across brand and slogan types.

The data collection process was carried out in three systematic stages. First, fashion slogans were compiled from official brand websites, global marketing campaigns, and advertising databases. Only slogans that were verified as authentic taglines were included to ensure the credibility of the data source. Second, each slogan was inputted into ChatGPT individually under the same controlled prompt conditions, without providing additional context beyond the slogan itself. This approach aimed to simulate realistic user conditions, reflecting how ChatGPT might typically be used by marketing professionals or general users

seeking quick translations. Third, all translated outputs were manually screened for completeness, linguistic coherence, and relevance. Incomplete or contextually irrelevant outputs (less than 5%) were excluded from analysis. The final dataset thus consisted of 175 valid English–Indonesian slogan pairs ready for evaluation.

The translation quality analysis employed the Translation Quality Assessment (TQA) model developed by Nababan, Nuraeni, and Sumardiono (2012), which has been widely applied in translation studies within the Indonesian context. The model evaluates translations across three major criteria: accuracy, acceptability, and readability.

1. Accuracy measures how faithfully the translated slogan conveys the meaning, message, and nuances of the source text.

2. Acceptability assesses whether the translation conforms to the linguistic norms, idiomatic expressions, and cultural appropriateness of the target language.

3. Readability evaluates the ease with which the translation can be understood by its intended audience, particularly in terms of fluency and clarity.

Each dimension was scored using a three-point scale:

- 3 = High quality (accurate/acceptable/readable)
- 2 = Moderate quality (less accurate/acceptable/readable)
- 1 = Poor quality (inaccurate/unacceptable/hard to read)

Three evaluators were involved in the scoring process: two university lecturers specializing in translation studies and one professional translator with editorial experience in marketing texts. Each evaluator independently rated every slogan based on the three criteria. The mean of the three evaluators' scores was calculated to produce a composite AAR score per slogan. This triangulated scoring ensured reliability and minimized subjectivity in evaluation.

After obtaining the AAR scores, the slogans were grouped based on their genre diversity, brand prominence, and target audience. This categorization allowed for cross-sectional comparisons to determine whether translation quality differed across specific contexts.

For example, motivational slogans (e.g., "Just do it") might present higher readability but lower acceptability if literalized, while luxurious slogans (e.g., "Elegance is an attitude") could face cultural adaptation challenges in Indonesian. Similarly, slogans from global brands might exhibit higher accuracy but reduced naturalness due to their English-based rhetorical style. The analysis also investigated how the creative function of slogans—such as evoking emotion, persuasion, or aspiration—was preserved or distorted in translation. This layer of analysis provided insights into ChatGPT's ability to handle pragmatic and cultural equivalence, aspects often neglected in purely semantic translation evaluations.

The data analysis proceeded through three major stages.

1. Stage 1 – Textual and Semantic Examination:

Each slogan pair was qualitatively analyzed to identify lexical shifts, semantic loss, idiomatic mismatches, and cultural omissions. Special attention was given to tone, rhythm, and wordplay—key features in slogan translation.

2. Stage 2 – Quantitative Scoring and Statistical Tabulation:

The AAR scores were tabulated in Microsoft Excel, and average scores were computed for each slogan, genre, and brand category. Descriptive statistics (mean, range, and frequency) were then used to map translation performance patterns.

3. Stage 3 – Interpretation and Discussion:

The results were interpreted in light of translation theory, particularly concerning equivalence, naturalness, and cultural transfer. Selected examples of high-quality and low-quality translations were discussed to illustrate specific translation issues, such as literal

renderings that reduced persuasive impact or successful adaptations that enhanced local resonance.

This study is expected to provide a nuanced understanding of how ChatGPT performs in translating short, persuasive, and culturally loaded texts like fashion slogans. The findings will reveal not only quantitative trends in translation quality but also qualitative insights into how linguistic creativity and cultural adaptation are handled by AI-based translation systems. Ultimately, this research aims to contribute to the broader discourse on AI-assisted translation quality assessment, highlighting the potential and limitations of neural models in the domain of marketing communication.

RESULT AND DISCUSSION

The evaluation of 175 English–Indonesian fashion slogans translated by ChatGPT demonstrates nuanced variation across the three Translation Quality Assessment (TQA) dimensions—accuracy, acceptability, and readability. Table 1 presents the statistical summary of mean scores and standard deviations based on expert evaluations conducted using Nababan's (2012) scale.

Table 1. Mean Scores of ChatGPT's Translation Quality

Dimension	Mean Score	SD	Interpretation
Accuracy	2.71	0,23	High
Acceptability	2.54	0.28	Moderate-High
Readability	2.62	0.25	Moderate-High

The data reveal that ChatGPT consistently delivers high accuracy, preserving the propositional content and lexical meaning of source slogans with minimal semantic distortion. However, acceptability and readability remain within moderate–high levels, indicating occasional stylistic or pragmatic incongruence in Indonesian. While the literal sense of most messages is intact, stylistic refinement, idiomatic fluency, and cultural alignment vary considerably across the dataset.

For example, idiomatic expressions such as *"Unleash your style"* rendered as *"Lepaskan gayamu"* demonstrate correct lexical mapping but fail to convey the empowering and motivational nuance typical of fashion advertising discourse. This exemplifies Molina and Albir's (2002) observation that literal translation alone cannot capture implicit cultural and emotional associations embedded in creative text types.
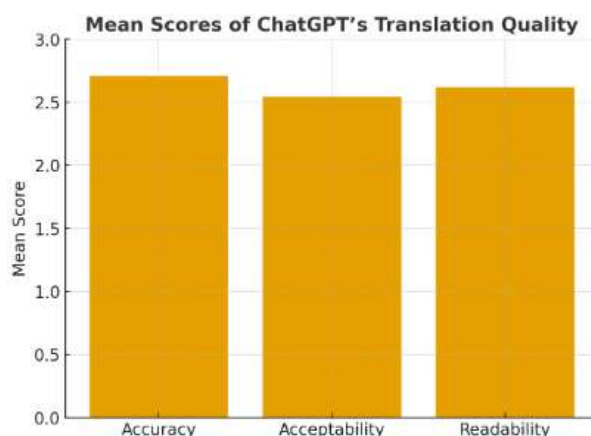

Figure 1: Mean scores of Accuracy, Acceptability, and Readability
Figure 1 illustrates the comparative mean scores of the three dimensions, highlighting

accuracy as the strongest component and acceptability as the weakest.

Translation Quality by Brand Origin

A brand-oriented analysis revealed subtle but meaningful distinctions between global and local fashion brands. As summarized in Table 2, translations of global brand slogans (e.g., Nike, H&M, Levi's) displayed higher accuracy yet lower acceptability, while local Indonesian brands exhibited more natural phrasing and contextual appropriateness, albeit with slightly reduced precision in semantic rendering.

Table 2. Translation Quality by Brand Type

Brand Type	Accuracy	Acceptability	Readability	Key Observation
Global (n = 90)	2.76	2.48	2.61	Accurate but less idiomatic
Local (n = 85)	2.59	2.66	2.63	Natural but slightly less precise

These patterns reflect how global brand slogans, often crafted in English for international markets, depend heavily on rhetorical brevity and idiomatic flair that resist literal transfer. For instance, the globally iconic slogan *"Just Do It"* translated as *"Lakukan saja"* preserves semantic meaning but loses its persuasive rhythm and affective brevity. In contrast, local brand slogans such as *"Fashion for All Seasons"* → *"Busana untuk segala musim"* achieve smoother cultural resonance and readability due to familiarity with Indonesian linguistic conventions.

This dichotomy supports Nababan's (2012) argument that translation quality depends not only on linguistic equivalence but also on audience familiarity and socio-cultural congruence. For global slogans, ChatGPT's translations tend to retain propositional integrity but struggle with contextual adaptation, reaffirming Garcia's (2022) claim that large language models lack sensitivity to brand voice and target-market culture.

Genre-Based Performance

Genre analysis revealed further differentiation among motivational, lifestyle, and luxury-oriented slogans. Table 3 details the mean scores across each genre category.

Table 3. Translation Quality by Slogan Genre

Slogan Genre	Accuracy	Acceptability	Readability	Illustrative Case
Motivational (e.g., *"Be yourself."*)	2.78	2.45	2.60	Meaning preserved, tone weakened
Lifestyle (e.g., *"Made to move."*)	2.70	2.56	2.65	Clear and natural flow
Luxury (e.g., *"Elegance is an attitude."*)	2.60	2.48	2.62	Stylistic refinement required

Motivational slogans achieved the highest accuracy because of their direct, imperative structure, which aligns easily across languages. However, their affective intensity often diminished in translation. For instance, *"Be yourself"* rendered as *"Jadilah dirimu sendiri"* remains accurate but lacks the succinct empowerment tone of the English source.

Conversely, luxury-oriented slogans—characterized by elegance, abstraction, and implied prestige—posed greater stylistic challenges. The phrase *"Elegance is an attitude"* translated as *"Keanggunan adalah sikap"* accurately conveys the denotative message but sounds stiff and unnatural in Indonesian. Post-editing adjustments such as *"Keanggunan*

tercermin dari sikap" better capture the aspirational tone and rhythmic flow typical of luxury discourse.

This finding supports Hoang's (2021) assertion that AI translation systems tend to prioritize lexical equivalence over stylistic adaptation, thereby diminishing emotional resonance and rhetorical sophistication in creative texts.

Cross-Dimensional Correlations

A correlation analysis between TQA dimensions revealed meaningful interrelationships. Acceptability and readability exhibited a strong positive correlation (r = 0.68), implying that natural and idiomatic translations are also perceived as easier to read and cognitively process. However, accuracy correlated weakly with both acceptability (r = 0.41) and readability (r = 0.38). This result reinforces the conceptual independence of TQA dimensions as proposed by Nababan et al. (2012), where literal accuracy does not inherently ensure stylistic or cultural fluency.

Figure 2 visualizes this correlation pattern, showing that as acceptability increases, readability tends to improve, while accuracy varies independently—suggesting that faithful translation does not guarantee emotional or cultural engagement.
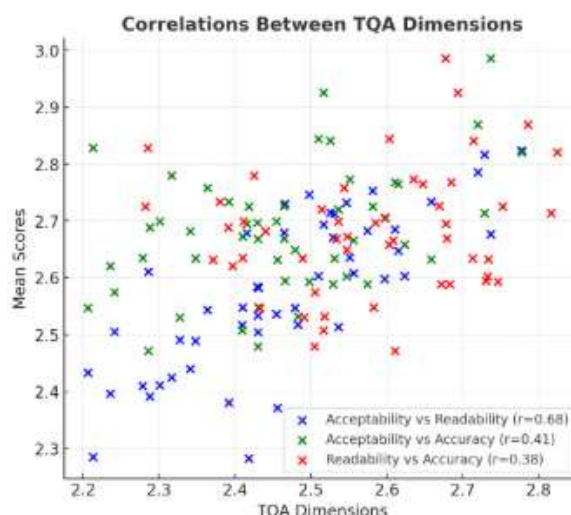

Figure 2: Correlations between Accuracy, Acceptability, and Readability

These correlations also provide empirical evidence for Molina and Albir's (2002) argument that translation involves a strategic trade-off between fidelity and naturalness, particularly in creative genres like fashion marketing. Translators—and AI systems by extension—must balance the preservation of meaning with the adaptation of stylistic and cultural cues to maintain communicative impact.

Accuracy and Semantic Retention

ChatGPT's strong accuracy scores demonstrate its capacity to preserve semantic integrity and maintain lexical precision—a reflection of its probabilistic training and massive multilingual exposure. However, this strength also reveals its literalist bias, where propositional meaning is prioritized over pragmatic nuance. For instance, *"Unleash your style" → "Lepaskan gayamu"* maintains surface-level fidelity but lacks the empowering undertone of self-expression central to fashion marketing.

This limitation mirrors Molina and Albir's (2002) distinction between literal translation techniques and modulation or adaptation strategies, the latter being critical for

dynamic equivalence in emotionally charged contexts. Therefore, accuracy in AI translation must be viewed as necessary but insufficient for communicative success in persuasive genres.

Acceptability and Cultural Adaptation

Moderate acceptability scores highlight the model's difficulty in aligning with Indonesian linguistic norms and cultural expectations. While the syntactic output is generally correct, some phrases exhibit unnatural collocations or foreignized structures, such as *"Pakai kepercayaan dirimu"* instead of the idiomatically appropriate *"Tampil percaya diri."*

This echoes Nababan et al. (2012), who emphasized that acceptability reflects how well a translation conforms to the target language's pragmatic and cultural conventions. Izwaini (2023) similarly argued that machine translation lacks the interpretive flexibility to mediate between cultural values embedded in source and target languages. Consequently, post-editing remains crucial to bridge this gap—ensuring not only linguistic accuracy but also socio-pragmatic alignment with the Indonesian audience.

Readability and Consumer Engagement

ChatGPT demonstrates relatively strong readability, producing coherent and structurally accessible translations. Nevertheless, readability in marketing texts involves more than grammatical simplicity—it also depends on rhythm, parallelism, and sound harmony. For instance, *"Find your fit, feel your flair"* → *"Temukan ukuranmu, rasakan gayamu"* preserves comprehension but loses the phonological symmetry and rhetorical flow that enhance brand memorability.

This finding corroborates Hoang's (2021) claim that AI systems lack sensitivity to aesthetic cadence, an essential feature in advertising. In such cases, human editors must apply compensation techniques, reintroducing stylistic rhythm through lexical or syntactic reshaping while retaining semantic clarity.

Implications for TQA and Post-Editing Practice

The integrated application of Molina & Albir's translation techniques and Nababan's TQA model provides a comprehensive framework for assessing AI-mediated translations. While ChatGPT ensures baseline semantic fidelity, the findings confirm that acceptability and readability improvements require human mediation through adaptive editing.

This reinforces Garcia (2022) and Kenny's (2023) perspectives on collaborative translation workflows, where human translators refine the pragmatic and stylistic layers of AI-generated drafts. The study thus contributes methodologically by proposing a hybrid TQA framework—one that combines quantitative evaluation with qualitative discourse analysis for creative translation assessment.

Although this study offers a robust evaluation of ChatGPT's translation quality, it is limited by its text-only focus, excluding multimodal elements such as typography, visuals, and sound symbolism that frequently accompany slogans. Future research could integrate multimodal translation analysis, considering how linguistic and visual elements interact to construct brand meaning. Additionally, comparative studies across different large language models (e.g., Gemini, Claude, NLLB) may illuminate model-specific strengths in handling figurative and stylistic content. Finally, longitudinal studies assessing post-editing time efficiency and cost-benefit impact would further enrich the understanding of AI–human collaboration dynamics in creative translation industries.

CONCLUSION

In sum, this study reveals that ChatGPT performs reliably in producing semantically accurate translations of fashion slogans but still requires substantial post-editing to achieve idiomatic fluency and emotional resonance. The integration of Molina and Albir's translation

technique taxonomy with Nababan's TQA framework provides a robust methodological foundation for evaluating AI-mediated translation quality.

Beyond its empirical contributions, the study advances theoretical understanding of how linguistic precision and stylistic adaptation interact in machine-generated texts. These insights not only enrich translation studies scholarship but also inform the future of AI–human collaboration in the evolving landscape of creative translation.

Acknowledgement

REFERENCES

Castilho, S., Moorkens, J., Gaspari, F., & Way, A. (2018). Evaluating machine translation in context: The challenge of creative language. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), Translation quality assessment: From principles to practice (pp. 115–137). Springer. https://doi.org/10.1007/978-3-319-91241-7_6

Guidère, M. (2010). Translation practices in international advertising. In Y. Gambier & L. van Doorslaer (Eds.), Handbook of translation studies (Vol. 1, pp. 156–161). John Benjamins. https://doi.org/10.1075/hts.1.tra7

House, J. (2015). Translation quality assessment: Past and present. Routledge. https://doi.org/10.4324/9781315752839

Izwaini, S. (2023). Machine translation, culture, and creativity: Evaluating AI-mediated translation performance. Translation & Interpreting Studies, 18(2), 233–252. https://doi.org/10.1075/tis.21064.izw

Kenny, D. (2023). The translator and the machine: Human–AI collaboration in translation practice. Target: International Journal of Translation Studies, 35(1), 1–25. https://doi.org/10.1075/target.22014.ken

Molina, L., & Albir, A. H. (2002). Translation techniques revisited: A dynamic and functionalist approach. Meta: Translators' Journal, 47(4), 498–512. https://doi.org/10.7202/008033ar

Nababan, M. R., Nuraeni, A., & Sumardiono. (2012). Pengembangan model penilaian kualitas terjemahan. Kajian Linguistik dan Sastra, 24(1), 39–57. https://doi.org/10.23917/kls.v24i1.439

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

Popović, M. (2020). On the differences between human translations. Machine Translation, 34(3), 151–167. https://doi.org/10.1007/s10590-020-09246-y

Sari, D. P., Nababan, M. R., & Djatmika. (2020). Translation techniques and quality

assessment of tourism promotional texts. Humaniora, 32(2), 130–142. https://doi.org/10.22146/jh.51167

Toral, A. (2022). Is machine translation ready for literature? Linguistica Antverpiensia, New Series: Themes in Translation Studies, 21, 85–107. https://doi.org/10.52034/lanstts.v21i1.680

Garcia, I. (2022). Post-editing neural machine translation: A study of effort and quality. Machine Translation, 36(1–2), 1–25. https://doi.org/10.1007/s10590-021-09283-2

Hoang, H. (2021). Translating creativity: Aesthetic and stylistic constraints in neural machine translation. Babel: International Journal of Translation, 67(5), 771–790. https://doi.org/10.1075/babel.00250.hoa