Lingua: Jurnal Ilmiah, *Vol.22 No.2 October, 2026*
e-ISSN: 2715-6478

The accuracy of cultural words translation by DeepL in the novel *Tarian Bumi*

Firman Asfari*, Neneng Sri Wahyuningsih

Universitas LIA, Jakarta, Indonesia
**Corresponding author: firmanasfari@hotmail.com*

ABSTRACT

These days, the usage of machine translation, such as DeepL, has increased in number. Furthermore, this study aims to assess the accuracy of DeepL in handling culturally nuanced words, in conjunction with Yildiz's (2021) theory for evaluating translation accuracy. This study employs a descriptive-qualitative method to analyze the source data. Three independent professional translators assessed DeepL's accuracy. Of the 303 rich points identified, 95 of them are raters' consensus, followed by raters' dissent for 76 rich points. According to the assessment, DeepL can identify terms that have been administered globally. However, some shortcomings are apparent regarding words with culturally dense nuances. DeepL's inconsistency was also highlighted in translating a word that appeared 44 times but produced six different translations. For the accuracy of the translation, DeepL is given a final score of 0.54, within the range of -8 (Critical Error) to +4 (Acceptable). This score is considered partially acceptable and sufficient for initial understanding. Then, it concludes that DeepL's translation still requires human involvement in preserving cultural nuances. This research recommends enhancing MT systems with more diverse cultural words and phrases datasets. For future studies, it is suggested to use multiple MTs and quantitative metrics like BLEU scores for broader, comparative evaluations.

Keywords: cultural words; DeepL; translation accuracy; machine translation; novel

Received: 6 May 2026 *Accepted: 31 August 2026* *Published online: 6 October 2026*

INTRODUCTION

People see translation as altering text from one language to another; in fact, it embraces a broader scope. Translation, as described by experts, encompasses various approaches and emphases. Catford (1965/1978, p. 20) demonstrates translation as "the replacement of textual material in one language (source language) by equivalent textual material in another language (target language), " focusing on the linguistic features and structural connection between the source languages (SL) and target languages (TL). Furthermore, Nida and Taber (1969/2003, p. 12) depict translation as "the process of reproducing in the receptor language the closest natural equivalent of the SL message, first in terms of meaning and secondly in terms of style." In other words, these scholars focus on meaning rather than style, which includes not only the accuracy in grammar rules but also comprehension of the target language. Then, Newmark (1988, p. 8) explains that translation involves "rendering the meaning of a text into another language in the

way that the author intended the text," pointing up the translator's part in keeping the writer's purpose and the text's meaning.

Furthermore, Larson (1998) extends an applied methodology that operationalizes these theories into an organized translation process. The translation process is a systematic approach intended to ensure the accuracy of meaning from the source text (ST) to the target text (TT) (Larson, 1998). A meaning-based approach is underlined by Larson (1998) to transfer the meaning to a literal concept. The analysis stage is initiated, where the ST is explored to comprehend its semantic structures, syntactical, and grammatical. For accurate translation, a comprehensive understanding is essential. The transfer stage is then carried on, where equivalent expressions are in the TL, conveying the meaning of the ST. At this stage, the understanding of both languages must be in place to ensure the translation stays on the original message. Restructuring is the final stage, requiring verifying the result to have the natural texts when read. Cultural relevance, fluency, and accuracy are included to rewrite the text at this stage. This process ensures both accuracy and comprehension with the readers. The importance of comprehension of both SL and TL, and cultures has been accentuated in Larson's (1998) process, so it creates both precise and relevant translations.

Typically, translation is conducted by humans who can ensure accurate and culturally sensitive translation results (Bowker & Ciro, 2019). Moreover, the current translation process is often carried out with the help of a translation machine, called Neural Machine Translation (NMT). NMT is a machine translation utilizing neural network techniques to foresee a set of texts in order (Wiggins, 2023). NMT has advancements in the products in the global market, such as Google Translate (GT), Microsoft Translator (MT), and DeepL. As described by Rysz (n.d.), a free tool using advanced natural language processing has been applied to GT, including machine learning and neural networks. Moreover, Caswell (2024) reported that Artificial Intelligence is applied to enlarge the support of language variety in GT. As Google's competitor, Microsoft also offers NMT, known as MT, a real-time language translation service that helps users break down language barriers (Microsoft, 2023). Lastly, DeepL is a free web translation service that has gained a foothold in the market, and more than a billion people use it (Wyndham, 2021). However, the advance of those NMTs has shortcomings, such as the inconsistency in translation (Naveen & Trojovský, 2024).

DeepL has been popular lately because of its features. Wyndham (2021) reported that since its release in 2017, DeepL has continued expanding its language offerings. Thus, DeepL is not only popular in NMT but also has some features that others do not. The DeepL free version provides file translation ability, permitting users to upload up to three files per month, with each file being up to 5 MB in size to be translated. The paid version offers more features and additional benefits, such as the ability to use one account for unlimited users, bulk file translation, glossary management, and many more (DeepL, n.d.). DeepL also provides various

translation choices, equipping users with more flexibility and control over the final writing. Furthermore, it was also mentioned that DeepL is recognized for its precision (Egwuogu, 2024; Hidalgo-Ternero, 2021) and natural-sound translations (Egwuogu, 2024; Macketanz et al., 2018).

However, one of the challenges that DeepL faces is rendering cultural words, which must be proven because of the unique aspect of cultural words as described by Newmark (1988). Cultural words, embedded in the culture's traditions, customs, and social practices, are terms that are specific to a particular culture and often do not have a direct equivalent in other languages (Newmark, 1988). Newmark classifies cultural categories into five parts: (1) Ecology consists of plants, animals, geographical elements, and climate-related terms particular to a specific culture; (2) Material culture is about entities or artifacts associated with a certain tradition, including clothing, food, housing, transport, and equipment; (3) Social culture includes social norms, practices, and social customs unique to a certain culture; (4) Organizations, customs, activities, procedures, and concepts encompass political, social, legal, religious, and artistic aspects; (5) Gestures and habits are about culturally specific gestures or habits.

In addition, Olk (2013) introduced cultural references when discussing cultural words. Cultural references, according to Olk (2013), are the lexical items in the ST that do not exist in the target culture or contradict in both meaning and connotation. The cultural references include proper names, customs, historical events, food, and geographical elements, as well as linguistic features such as idioms and proverbs. These are crucial when delivering culture-specific terms because they create challenges in rendering the result of translation due to its cultural elements.

Building on the discussion of cultural words by Newmark and cultural references by Olk (2013) in translation studies, it is essential to examine the quality of translations that drive these cultural complexities. In this regard, the Translation Quality Assessment (TQA) framework established by Yıldız (2021) offers an assessment tool to measure the accuracy of the translation. This is a process for scoring translations, itemizing them into five categories based on the error tier, namely Critical Error, Major Error, Minor Error, Partially Acceptable, and Acceptable. These categories assist in demonstrating a grading level to rich points referred to as a specific component or portion of the source text, likely to have an issue due to linguistic, cultural, or contextual complexity. Furthermore, it manages raters in their assessment process.

To implement this grading system effectively, Yıldız (2021) provides a breakdown of how each error level translates into specific point assignments, reflecting the severity and impact of the errors on translation accuracy. The grading table provides an outline of the evaluation criteria for translation accuracy. Each level is scored based on either the severity of the error or the quality of the solution. -8 points deduct a Critical Error, while -4 points deduct a Major Error, and a Minor

Error is deducted by -2 points. On the other hand, partially acceptable solution parts are donated +2 points, and those that are considered acceptable gain +4 points.

In practical terms, Yıldız (2021) explains that those grading points are described as follows:

- Critical Error Degree (-8): The error does not make sense and is not related to either the text or the context.
- Major Error Degree (-4): The error does not make sense in the text or context.
- Minor Error Degree (-2): The equivalent of the meaning is not exact, but it still makes sense in the context of the text.
- Partially Acceptable Degree (+2): It makes sense in the context but has slight semantic confusion.
- Acceptable Degree (+4): It is natural and easy for the reader to understand.

In addition to frameworks such as Yıldız (2021), there are quantitative mechanisms for assessing machine translation (MT) accuracy, such as BLEU (Bilingual Evaluation Understudy) score, introduced by Papineni et al. (2002). BLEU is an automatic evaluation method that compares MT output with one or more human reference translations, producing a score from 0 to 1 (or 0–100%). This metric is particularly useful for handling large datasets and comparing the performance of different MT systems.

As the continuity of capabilities of MT, Karabayeva and Kalizhanova (2024) examine the effectiveness of DeepL in literary texts. A single theory is not used in their research paper. Instead, it refers to various previous studies on MT and NMT conducted by Moorkens (2017, 2018) and Castilho et al. (2017, 2018). These theories emphasize that although NMT produced smoother translations, it still struggles to maintain the creative, rhetorical, and cultural aspects of literary texts. Furthermore, Gaspari et al. (2015) and Chesterman (2021) also provided theoretical foundations regarding MT competence and the challenges of AI in addressing creativity and subjectivity. This research uses a qualitative comparative method. Various research data, such as poems, books, and dialogues, are deployed in these studies to examine the MT's performance. From that research, it was found that uncomplicated texts could be translated by tools like ChatGPT and DeepL. However, they had some issues when it came to complex literature containing idioms, metaphors, and cultural nuances.

Then, Munthe et al. (2023) assessed TQA with a variety of texts, such as narrative texts, procedural texts, and literary works. Using the theories of Nababan et al. (2012) in accuracy parameters, the studies applied a descriptive-qualitative method. Findings were noted that even though MT tools like GT were faster and free-cost web, translating in complicated phrases, idioms, and cultural contexts often failed when compared to human translators. Then, their assessments underline the failure of MTs in maintaining the cultural contexts in the STs.

The next study is from Hasyim et al. (2021) focusing on the MT's accuracy on GT in rendering culinary texts from French to Indonesian. The culinary recipes

collected from online sources were compromised, and then descriptive qualitative analysis was used. The author uses a semiotic framework of (Danesi, 2010) and (Jakobson, 2012) to explain how MT, like GT, works in transferring meaning by referring to the concept of signifier (form) and signified (concept) to analyze translation accuracy. This research found that GT can translate from words to sentences. However, it fell short when translating literal translations and cultural equivalence. The GT's limitation in conveying and recognizing the cultural context was pointed out.

This research has the same alignment as the studies conducted by Karabayeva and Kalizhanova (2024), Munthe et al. (2023), and Hasyim et al. (2021), which is to assess the translation accuracy by employing MT. Similarly, the importance of examining MT in handling the translation's accuracy and literary texts was emphasized. They highlight the differences in translating books, poems, and dialogues compared to translations done by humans. Thus, Munthe et al. (2023) compare the accuracy of GT and a human translation, which was assessed by the rater. Hasyim et al. (2021) examine the accuracy of MT by translating culinary texts. Overall, these studies help to have a better understanding of the strengths and weaknesses of MTs.

Meanwhile, there exist gaps that have to be addressed. The research gaps indicated in recent studies emphasize the ongoing challenges and limitations of MT systems in accurately managing literary texts and cultural nuances. Translation of cultural words presents a significant challenge in translation studies, especially when using MT tools like DeepL. Cultural words are attached to the SL's cultural and social elements, and their accurate translation is critical for maintaining the original definition and effect of the text.

The novel *Tarian Bumi*, written by Oka Rusmini, demonstrates some cultural terminologies that are deeply embedded in Balinese culture, for example, a social hierarchy, caste relations, religious practices, rituals, and traditional customs. Moreover, they have not only lexical but also social, religious, and cultural meanings, making it challenging to translate accurately for DeepL. Exploring cultural words in this novel is challenging because of defining cultural word categories, translating accurately, maintaining originality, and delivering awareness of the strengths and restrictions of current translation tools. Then, the performance of DeepL in translating the novel with cultural nuances into another language has not been comprehensively researched. The purpose of this study is to assess the accuracy of DeepL in translating cultural words and references from *Tarian Bumi*, in conjunction with Yıldız's (2021) theory for evaluating translation accuracy.

METHOD

This research uses a descriptive-qualitative method that aims to provide an accurate and comprehensive description of circumstances or a phenomenon.

Descriptive qualitative research employs descriptions to investigate and comprehend the phenomenon, proposing an extensive portrayal without large reliance on theoretical frameworks (Patton, 2015). Merriam and Tisdell (2015) add that descriptive qualitative research provides an analysis of an event or phenomenon within its natural context. Moreover, the data is processed by content analysis. Systematically exploring communication data to analyze patterns, themes, and definitions by coding textual data into manageable classifications is required in this method, as explained Krippendorff (2022). This procedure aids in the interpretation of the communication material's substance and context, introducing significant concepts into the subject matter.

Throughout the novel, nuanced cultural words and expressions are identified as the research data based on specific criteria, which are Organizations, Customs, Activities, Procedures, and Concepts, as well as Material Culture (Artefacts). These criteria's meanings are closely associated with Balinese cultural knowledge and could not be adequately interpreted solely through their literal lexical meaning. Furthermore, the novel has existed for more than two decades in the market, but it still has high accessibility. The readers can get the novel both through physical bookstores and online platforms, including its availability in global markets such as Amazon (Amazon.com, n.d.). This novel accommodates a thorough narrative of Balinese customs that are not widely familiar, and it comes up with a concept of the caste system in Bali (Marlina, 2024). Furthermore, it is relevant as a study of cultural words because it raises the traditions, values, and dynamics of Balinese culture. Due to the market demands, the novel was republished in 2007 by Gramedia, while the first publisher was Indonesia Tera in 2000. Thus, in 2011, the Lontar Foundation published Tarian Bumi's English version, titled Earth Dance, which was translated by Rani Amboyo and Thomas M. Hunter. Besides its popularity, it has also been an award winner; the Language Center of the Ministry of National Education of Indonesia presented the author with the 2003 Literary Award (Gramedia, n.d.).

Considering limitations on the time frame and its rich cultural nuances, this study only focuses on two of the five categories. Carrying out all five categories will require more time to analyze the data. Thoroughly examining all the categories would be a challenge within the time frame, particularly those that are not the novel's main character. The categories in this research are Organizations, Customs, Activities, Procedures, and Concepts, as well as Material Culture (Artefacts). The three categories that are not included in this research are Ecology, Social culture, and Gestures and habits. The category of ecology that consists of plants, animals, geographical elements, and climate is not the main storyline of the novel. Then, the exclusion of Social Culture (social norms, practices, customs) because of its partial overlap with Organizations, Customs, Activities, Procedures, and Concepts. For instance, the term *patiwangi* can be categorized under Social Culture because it represents regular practices in Balinese society. However, it was also observed

under ritual in Balinese society, which is part of the category Organizations, Customs, Activities, Procedures, and Concepts. Another example is the term *pemangku*, a priest in Hinduism, embodying social practices, but it is under religious procedures in Hinduism. By having those reasons, aspects of social culture are partially brought up in the categories already included, and to prevent redundancy, Social Culture is not included as a sole category in this research. Gestures and habits are often delivered in non-verbal forms such as body language, intonation, and physical expression. Because of that, accurately capturing those elements cannot be analyzed meaningfully in a text-based environment where DeepL applies.

The study implicates a structured procedure to assess the accuracy of DeepL in translating cultural words or references from the Indonesian version into the English version. Initially, sentences with cultural nuances in the original Indonesian novel are marked and put into the data card. After arranging all cultural words from the source language, the sentences in the data card are input into DeepL for the next step, which is to be translated into English. Once DeepL has accomplished the translations, they are handed over to three professional translators to rate the accuracy. The requirements for translators are that they have more than five years of experience in the translation industry. Next, the translations are assessed and set scores based on Yıldız's (2021) TQA theory. To acquire the results, the scores are then averaged, adding an assessment of DeepL's performance when translating culturally specific elements.

RESULTS AND DISCUSSION

Findings

To calculate the average score given by the three assessments to the overall DeepL translation quality, the following formula is used:

$$X = \frac{n_1 + n_2 + n_3}{Y}$$

$$\bar{S} = \frac{X}{R}$$

Notes:

- $n_1 + n_2$ and n_3 mean the total scores from Rater 1, Rater 2, and Rater 3.
- Y means the number of the raters, three raters.
- X means the average scores from all the raters.
- R means the number of rich points.
- $\bar{S}$ means the average score per rich point.

Based on the Translation Quality Assessment (TQA) framework developed by Yıldız (2021), 303 rich points compiled from the novel Tarian Bumi reached an average of 0.54, as illustrated as follows:

$$X = \frac{(-628) + 710 + 412}{3} = 164.67$$

$$\bar{S} = \frac{164.67}{303} = 0.54$$

The final scores from the three independent raters for the TQA were -628, 710, and 412, respectively, as shown in the above formula, with a total of 494. Then, it is divided into three, resulting in 164.67. Furthermore, 164.67 is divided into 303 rich points, yielding 0.54 points in the total score of DeepL accuracy.

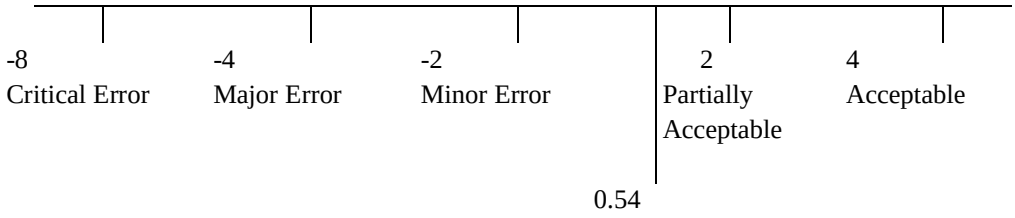


Figure 1. DeepL's Average Score in Relation to the Yıldız TQA Rubric

The significant differences in scores between raters reflect subjectivity in the translation evaluation process, especially when the source text is heavily culturally and contextually loaded. The average final score across the three raters was 164.7, or about 0.54 per rich point, indicating that the translation quality was quite good, although not entirely consistent. This average score of 0.54, in the context of the TQA scale ranging from -2 (Minor Errors) to +2 (Partially Acceptable), indicates that the translation reflects a mix of acceptable solutions and minor to moderate errors. DeepL translation results in this context can be considered partially acceptable and sufficient for initial understanding or internal purposes; however, professional revision is still required to meet the quality standards required for publication, especially in literary works, culturally meaningful texts, or related to ceremonies.

Discussion

Raters' Consensus

A total of 95 rich points were given the same score by the three independent raters, both those that received a perfect score, Acceptable grade (+4), and those that received the lowest score, Critical Error grade (-8). The rich points mostly consist of terms and phrases that have strong cultural meanings and are inherent in the local context of the novel, such as *dewa* and *sesaji*, which have standard cultural nuances. In addition, words with strong cultural nuances, such as *pemangku* and *canang sari*, were successfully maintained by DeepL along with explanations of their meanings from the source texts, so that they can still be understood by foreign readers. Proper nouns such as *Legong Keraton* and *Joged Bumbung* were also maintained. By borrowing the proper nouns, the uniqueness of the terms is kept so

that the cultural meanings are retained (Bassnett, 2014). Meanwhile, a globally known term, such as *kebaya*, was also successfully maintained in the translation. Lastly, all the raters gave -8 (Critical Error) to the inaccurate translation of *ngaben* because DeepL translated *ngaben* as 'buried' without accommodating the deep meaning of customs and cultural context. According to Partami et al. (2016), *ngaben* is *upacara pembakaran mayat dalam agama Hindu* (back translated: a cremation ceremony in the Hindu religion).

The grading table for categories Major Error, Minor Error, and Partially Acceptable with -4, -2, and 2, respectively, is not found in the consensus. This shows that consensus between raters only occurs at the two ends of the rating scale. When the translation quality is satisfactory or faulty, the raters can get a clear picture of the evaluative cues, as shown in Table 1.

Table 1. Examples of Raters' Consensus on DeepL Translation

No.	Rich Point	Source Text	DeepL	Rater 1	Rater 2	Rater 3
46.	Dewa	Dia tersenyum seperti menantang para dewa.	She smiled like she was challenging the gods.	4	4	4
57.	Ngaben	"Perempuan itu tidak boleh diabenkan. Dia harus…"	"The woman should not be buried. She must..."	-8	-8	-8
85.	Kebaya	Kebaya Telaga robek.	Telaga's kebaya was torn.	4	4	4
94.	Sesaji	Sesaji sudah berada di hadapan Telaga.	The offerings were already in front of Telaga.	4	4	4
95.	Pemangku	Seorang pemangku mengucapkan mantra-mantra.	A pemangku chanted mantras.	4	4	4

In line with Munthe et al.'s (2023) study, this study employs professional translators to evaluate the translation. However, Munthe et al. (2023) compare the accuracy of target texts between Google Translate and a professional translator, and then the translations were assessed by three independent translators. This study measures the accuracy of DeepL translations and evaluates it using the translations provided by three independent translators, resulting in a total of 95 rich points.

Raters' Dissent

In the rater's dissent, where raters gave different scores to the data distributed, 76 rich points had significant differences in scoring patterns. The differences in scoring were seen in assessing the quality of the same translation. This finding emphasizes the complexity of assessing cultural translations, especially when the translations are neither completely accurate nor completely wrong. The disagreements all occurred in scores other than Critical Error and Acceptable, indicating that variability in interpretation affects scoring in ambiguous or nuanced cases. These results strengthen the view that although the TQA rubric provides a

scoring structure, rater interpretation remains a major factor in evaluating machine translations of culturally rich texts.

While this study notes that DeepL faces the challenges in handling culturally dense nuances, Karabayeva and Kalizhanova (2024) point out that DeepL, as well as ChatGPT, show their failure. When crafting creative language, such as nuance, fluidity, and aesthetic mastery, both DeepL and ChatGPT struggle to convey the message as effectively as professional human translators do. Karabayeva and Kalizhanova (2024) also bring up DeepL, which tends to translate literally instead of interpreting the underlying meaning. In other words, they concluded that DeepL generally produces translations without awkward phrasing.

Examples of rich points affecting different interpretations by the raters are seen in Table 2. According to Rater 1, the translation is common and natural, so the highest score is given. However, others have lower scores. While Rater 2 scored the translation as Partially Acceptable because additional information is required, Rater 3 emphasized the lack of conveying of cultural content. Cultural content in this sentence refers to the spiritual concept in the Hindu Bali context, which features various gods of dance, such as Shiva Nataraja. Furthermore, Rater 3 proposes the translation as 'A dancer who has all the beauty of Shiva Nataraja.'

Table 2. Examples of Raters' Dissent on *dewa tari* and 'the god of dance'

No.	Rich Point	Source Text	DeepL	Rater 1	Rater 2	Rater 3
2.	Dewa tari	Penari yang memiliki seluruh kecantikan **dewa tari**.	A dancer who has all the beauty of **the god of dance**.	4	2	-2

The rich point above, *dewa tari*, received different points from three raters. Rater 1 sees the translation as clear and common English, and it fits general use. Rater 2 has a different opinion in translating this phrase; additional information is required to bridge the cultural concept. In contrast, instead of translating *dewa tari* as 'the god of dance', Rater 3 opts for 'Shiva Nataraja' because 'Shiva Nataraja' refers to specific figures and spiritual aesthetic concepts in local cultures.

Table 3 is another example of rich points that have variant perceptions by the assessors. Rater 1 pointed out that *arak* should be translated into 'arrack' because 'arrack' is an Asian alcoholic beverage like rum, distilled from a fermented mash of malted rice with toddy or molasses (Merriam-Webster, n.d.-a). Because of that, Rater 1 gave a score of -8 (Critical Error). Rater 2 has a different opinion, causing point 2 (Partially Acceptable). *Arak* is almost like 'wine,' even though the raw materials are different, but they both smell and are intoxicating. Thus, Rater 3 gave a score of Minor Error (-2). The definition of *arak* that Rater 3 emphasized is a traditional Indonesian liquor, usually made from rice or coconut, while wine is a Western fermented grape drink with different cultural connotations.

Table 3. Examples of Raters' Dissent on *arak* and 'wine'

No.	Rich Point	Source Text	DeepL	Rater 1	Rater 2	Rater 3
38.	Arak	Mulutnya berbau **arak**.	His mouth smelled of **wine**.	-8	2	-2

Those results from Tables 2 and 3 have different approaches and interpretations. Rater 1 emphasizes the details in an accurate translation to preserve both linguistic and cultural identity, causing the Major Error (-8). Rater 2 has a different perspective, which is because of dissimilarities. They both smell and are intoxicating, so it is considered partially acceptable. The cultural gap, by generalizing *arak* as 'wine,' is pointed out by Rater 3, resulting in a scoring of Minor Error (-2).

Translation of Repetitive Rich Points

It is observed that there are repetitive rich points in the data. The top five repetitive rich points are as follows: *griya*, *pura*, *Hyang Widhi*, *dewa*, and *sesaji*. *Griya* is the most frequently appearing term, with a frequency of 44 times. Then, it is followed by *pura*, *Hyang Widhi*, *dewa*, and *sesaji*, with 30, 24, 23, and 18, respectively. While *griya* has six distinct translations showing inconsistency, others have only one, as seen in Table 4.

Table 4. Translation Results for Repetitive Rich Points

Rank	Rich Ponts	Frequency of Appearance	Of Distinct Translations
1.	Griya	44	6
2.	Pura	30	1
3.	Hyang Widi	24	1
4.	Dewa	23	1
5.	Sesaji	18	1

The term *griya* appears 44 times, making it the most frequent cultural word appearing in the novel. However, DeepL cannot maintain consistency in the translation. According to KBBI (n.d.-a), *griya* is (1) *bangunan tempat tinggal, rumah*; (2) *kompleks perumahan, permukiman* (back translated: (1) a residential building, house; (2) a housing complex, residential area). Furthermore, the glossary in the novel defines *griya* as *rumah tempat tinggal kasta Brahmana* (back translated: a house where the Brahmana caste lives). By annotating a glossary from the novel, it can maintain the cultural nuances and preservation to enrich the reader's understanding approaches (Venuti, 2008).

Of the 44 appearances, there are six differences in translation, namely *griya* (28 times), house (seven times), palace (five times), village (twice), church (once), and temple (once). It indicates inconsistencies in how DeepL translates *griya* across contexts in the novel. Its inconsistency is likely due to the diverse narrative roles of *griya*. This finding underscores a major challenge in machine translation, namely, maintaining consistency and cultural fidelity in translating frequently occurring

cultural terms. The consistency by MTs, like DeepL, is also noted by Naveen and Trojovský (2024). The research mentioned that MTs often strive to maintain consistency in repetitive words in a cultural context and accurately translate idiomatic expressions and domain-specific terms. This causes problems when translating texts that are rich in local or cultural elements.

Table 5. Translation Influenced by Context

No.	Source Text	DeepL
5.	Ayah Nenek seorang **pendeta** yang memiliki banyak sisia, orang-orang yang setia dan hormat pada **griya**.	Grandmother's father was **a priest** who had many sisia, people who were loyal and respectful to his **church**.
8.	Kata orang-orang griya, dulu Nenek adalah perempuan tercantik di **desa**… Nenek lembut dan penuh penghargaan pada sesama. Dia tidak sombong. Banyak laki-laki di **griya** yang diam-diam mencintai Nenek.	The people of the village said that Grandmother used to be the most beautiful woman in **the village**… She was not arrogant. Many men in **the village** secretly loved Grandmother.
9.	Sayang, Nenek tidak tertarik dengan laki-laki yang masih kerabatnya. Kata Nenek, waktu itu semua laki-laki dalam **griya** sudah dianggap saudara sendiri.	Unfortunately, Grandmother was not interested in men who were her relatives. Grandmother said that at that time, all the men in **the village** were considered her own brothers.
35.	Sepuluh tahun lebih Telaga tidak pernah datang ke **rumahnya**. Telaga tidak tahu, kenapa begitu menginjakkan kakinya untuk yang pertama kali di pintu masuk **griya** dia merasakan nyeri.	For more than ten years Telaga had never come to her **house**. Telaga didn't know why she felt pain as soon as she set foot for the first time at the entrance of her **house**.
43.	"Sari tunggu Meme di **pura**. Meme ingin bicara dengan orang itu." Telaga berbisik pada Sari agar keluar dari **griya**.	"Sari wait for Meme at **the temple**. Meme wants to talk to that person." Telaga whispered to Sari to come out of **the temple**.

The various translations occur in contexts that refer to physical buildings, social institutions, or spiritual entities, nobility, caste hierarchy, or high social class environments. Those references have different cultural connotations, as seen in Table 5 above. For instance, the word church in datum five might come from the presence of the word priest in the sentence, relating to the term in a Christian context. In datum eight and datum nine, DeepL renders *griya* as 'village,' although *griya* in the novel specifically refers to a Brahmana residence. This rendering appears to be influenced by the word 'village' in the preceding narrative context, where one of the characters is described as having been the most beautiful woman in the village. In datum 35, the context mentions 'house' in the first sentence. This result translates *griya* as house in the next sentence. Thus, in datum 43, *pura*, which is 'temple' in English (Echols et al., 1997), is part of the conversational context when DeepL translates *griya* in the next sentence. This clue seems to lead DeepL to translate temple as the translation of *griya*. It can be concluded that DeepL seems to be sensitive to context, resulting in the loss of some cultural meanings in literary translation. Moreover, DeepL also adjusts the translations based on contextual clues.

Proper Nouns

Proper nouns, as explained by Xamrokulovna (2023), refer to entities and carry cultural and linguistic meanings. DeepL can identify the proper nouns, but still lacks in recognizing all of them. One of the factors that makes this success possible is DeepL's ability to recognize capital letters as a characteristic. According to Newmark (1988), proper names are indicated by capital letters. Furthermore, a word or phrase that begins with a capital letter tends to be retained by DeepL instead of being translated literally, as it serves as a signal for DeepL and other NMT systems to identify proper nouns, as explained by Mayhew et al. (2019).

The accuracy of MT was also noted by recent studies, stating that proper nouns pose some issues for MT systems. However, it can be tackled by using strategies to improve the accuracy. For instance, Takai et al. (2021) use Factored Transformer with named entity (NE) features when processing the translation of proper nouns. As a result, the accuracy of translation achieves a 9.6-point. Thus, another researcher, Sharma et al. (2023), scored 99.5% accuracy for the proper nouns that focus on person, location, and organization names. They use Stanford's named entity recognizer (NER) to process the proper nouns. The findings from those scholars align with this study, which is still struggling to achieve a 100% score. The findings on successful retention and challenges in translation are illustrated in Tables 6 and 7 below.

Table 6. Successful Retention of Proper Nouns

No.	Source Text	DeepL
6.	Aku harus membuang pikiran-pikiran buruk itu, **Hyang Widhi**.	I have to get rid of those bad thoughts, **Hyang Widhi**.
166.	Tari yang diajarkan Kambren adalah **Legong Keraton**.	The dance that Kambren taught was **Legong Keraton**.
213.	Bukankah ini sebuah prestasi? Apalagi Ketu ingin mereka berdua mengenakan busana tari **Oleg**. Tari yang penuh sentuhan emosi, sentuhan nafsu berahi. Tari percintaan.	Isn't this an achievement? Moreover, Ketu wanted them both to wear **Oleg's** dance outfit. A dance full of emotion, a touch of lust. A dance of romance.

As can be seen in Table 6, the proper nouns were successfully retained. DeepL can maintain proper nouns without altering their meaning. By preserving the terms, the foreign reader might avoid the cultural misinterpretations in the sentences. Furthermore, the raters agreed that the proper nouns should be kept, so that they are given the highest score, Acceptable (4), except for the proper noun of *Hyang Widhi*. Rater 1 suggested translating *Hyang Widhi* as 'the Supreme Being.'

Table 7. Challenges in Translating Cultural Proper Names

No.	Source Text	DeepL
74.	Ada yang bergeser dari wujud yang diberikan **Sang Hyang Hidup** padanya.	Something had shifted from the form that the **Living God** had given him.
114.	Sering sekali Sekar bertanya pada **Sang Hidup**, dosa apa sesungguhnya yang telah diperbuat perempuan ini hingga tak ada...	So many times Sekar asked **the Living One** what sins this woman had committed that there was...

No.	Source Text	DeepL
120.	Sering Telaga berpikir bagaimana caranya agar **Sang Hyang Hidup** bisa dibohongi.	Often Telaga thinks about how to make **the Living Hyang** be lied to.

Table 7 illustrates how DeepL faces the challenges in translating proper nouns. The two proper nouns that failed to be retained were *Sang Hyang Hidup* and *Sang Hidup*. *Sang Hyang Hidup* is translated as 'the Living God,' and *Sang Hidup* is 'the Living One.' Both terms are commonly used as everyday language for references to God in Hinduism, and these phrases are not part of theological terms, according to the author of the novel, Oka Rusmini (personal communication, June 3, 2025). Moreover, the translation of *Sang Hyang Hidup* as 'the Living God' does not fully reflect Hinduism and Balinese traditions because it is known in the context of Christianity (Hahn, 2010).

Additionally, *Sang Hidup* and *Sang Hyang Hidup*, which have been translated into 'the Living One' and 'the Living Hyang,' respectively, are literal translations. By having this translation, it causes the loss of cultural and spiritual nuances that are expressed in the original context. Bali, as a Hindu belief, has several different names for the gods, and in Balinese culture, this is an essential element in the cultural and spiritual sacredness (Eiseman, 2009). By eliminating this element in its literal translation, it shows that DeepL can translate technically, but it eliminates the sacredness in the word or phrase. Failing to consider this element leads to a less accurate translation in the context of religious texts and cultural literature.

The phrase *Sang Hidup*, which has been translated into 'the Living One,' is a literal translation. Thus, it leads to the loss of cultural and spiritual nuances that are expressed in the original context. According to KBBI (n.d.-b), *Sang* is defined as *kata yang dipakai di depan nama orang, binatang, atau benda yang dianggap hidup atau dimuliakan* (back translated: word used before the name of a person, animal, or object that is considered alive or revered). DeepL translates *Sang* into 'One,' which aligns with Merriam-Webster (n.d.-b)'s meaning as the person or thing singled out or considered as unique or distinct (back translated: *orang atau sesuatu yang dianggap unik atau berbeda*).

Similarly, *Hidup* is translated as 'Living,' which is correct in terms of literal translation. However, DeepL's translation overlooks the theological aspect that is inherent in the concept of Hinduism's life in Bali as sacred and divine. In the phrase of *Sang Hyang Hidup*, the word *Sang* is eliminated but *Hyang* is preserved, so it is translated as 'the Living Hyang.' This is another example of how DeepL fails to capture the spiritual aspects of Hinduism in Bali. As a result, the foreign readers may misinterpret sacred terms if cultural context and nuances are not accounted for.

One possible reason for this inconsistency is the existence of the word ‘Living,’ which is a common word. The word ‘Living’ is part of the high-frequency used vocabulary because of its appearance in English dictionaries, such as Oxford (Oxford University Press, n.d.). Although it starts with a capital letter, DeepL treats Hidup as 'Living.' The system considers that it is a common lexical vocabulary

instead of a proper noun. This overlap between cultural proper noun nuances and common vocabulary might create confusion for the MT, especially when the ST does not explicitly mark the context. Furthermore, DeepL's failures in some cases demonstrate the system's limited cultural sensitivity and underscore the need for human editing to ensure the integrity of cultural meaning in literary texts.

Loanwords

A loanword is defined as a foreign word that is kept either in the original form or adapted to fit the phonology of the target language (Newmark, 1988). In the novel *Tarian Bumi*, the loanword is observed to fall under the category of Material Culture (Artifacts), namely *kebaya*, six times. *Kebaya* is a type of women's garment or clothing that is commonly worn in Malaysia, Indonesia, and other Southeast Asian nations (Oxford Learner's Dictionaries, n.d.).

Table 8. Loanword

No.	Source Text	DeepL
82.	…memakai **kebaya**…	…wearing **kebaya**…
197.	"Jangan **kebaya** yang itu, Meme. Terlalu resmi."	"Not that **kebaya**, Meme. It's too formal."
273.	**Kebaya** Telaga robek.	Telaga's **kebaya** was torn.

Kebaya, which has been declared by UNESCO as a part of Indonesia's cultural heritage (UNESCO, 2024) and is known globally, is now also considered a lexicalized term in many dictionaries. For instance, Spanish (Reverso, n.d.) and English (Oxford Learner's Dictionaries, n.d.). Along with this, DeepL also apparently recognizes that the word is now part of the dictionary's lexicon. Therefore, DeepL retains the word as it is. In six of its appearances, DeepL does not make changes to the word, as seen in Table 8 above.

CONCLUSION

This study contributes to translation studies by demonstrating the importance of considering cultural meaning and contextual sensitivity when translating cultural words with a focus on repetitive rich points, proper nouns, and loanwords. Then, to explore DeepL's performance in assessing cultural translation, raters' consensus and raters' dissent are included. It concludes that DeepL's algorithm seems to emphasize contextual sensitivity when processing terms with dense cultural nuances. The performance of DeepL is generally reliable. However, some parts may require human translation to ensure the readers can understand the result without losing the cultural nuances.

Based on the above findings, this study has implications in various areas, including translation business development and machine translation enhancement. In the first area, DeepL shows that its translation accuracy for cultural words is still lacking, so the role of human translators is still needed to ensure more accurate

translations. In addition to that, in translating texts with cultural nuances, human translators must have a deep understanding of the cultures to avoid losing them in the narration. The next one is machine translation enhancement. There is a tendency for DeepL to demonstrate some shortcomings by losing cultural contexts in the translations. Thus, DeepL must add more cultural datasets to the system to tackle this issue and improve its overall performance.

From this study, the author asserts the limitations. First, of the five categories of cultural words, due to the time frame in this research. Because of that, it only applies to two categories, namely Organizations, Customs, Activities, Procedures, and Concepts, as well as Material Culture (Artefacts). Second, the MT that is utilized in this research is DeepL, while there are many other MTs in the market. This research only utilized one type of machine translation, DeepL, while there are various other platforms available, such as ChatGPT and GT. Therefore, it is recommended to include several types of machine translation so that the translation results can be compared more comprehensively. Additionally, DeepL is an AI-based MT that evolves. Because of that, time-bound factors might affect the results in this study, meaning the same source texts may have different outputs at different times. DeepL might process the maturity in the systems. User customization and settings might be another impact on the output. Nevertheless, excluding these factors from the study further limits the generalizability of the results. Lastly, this study utilizes the qualitative method. For future research, using quantitative methods, such as the BLEU score, would be more valuable because it provides a structured comparison for a larger data set and different MTs.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

Author Contributions
FA designed the study, gathered the data, drafted the paper, and revised the manuscript. MFR acted as the principal advisor, supervised the project workflow. Both authors approved the final version for submission.

Disclosure and conflicts of interest
The authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Declaration of generative artificial intelligence (AI) and AI-assisted technologies
The authors declare that no generative artificial intelligence (AI) or AI-assisted technologies were used during the preparation or writing process of this manuscript.

REFERENCES

Amazon.com. (n.d.). *Tarian Bumi (Indonesian edition)*. Retrieved June 30, 2024, from https://www.amazon.com/Tarian-Bumi-Indonesian-Oka Rusmini/dp/6020339157

Bassnett, S. (2014). *Translation studies*. Routledge.

Bowker, L., & Ciro, J. B. (2019). *Machine translation and global research: Towards improved machine translation literacy in the scholarly community*. Emerald Publishing Limited. https://doi.org/10.1108/9781787567214

Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to human and machine translation quality assessment. In J. Moorkens, S. Castilho, F. Gaspari, & S. Doherty (Eds.), *Translation quality assessment: From principles to practice* (pp. 9–38). Springer. https://doi.org/10.1007/978-3-319-91241-7_2

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, *108*(1), 109–120. https://doi.org/10.1515/pralin-2017-0013

Caswell, I. (2024). *110 new languages are coming to Google Translate*. Google. Retrieved May 25, 2024, from https://blog.google/products/translate/google-translate-new-languages-2024/

Catford, J. C. (1978). *A linguistic theory of translation*. Oxford University Press.

Chesterman, S. (2021). Through a glass, darkly: Artificial intelligence and the problem of opacity. *The American Journal of Comparative Law*, *69*(2), 271–294. https://doi.org/10.1093/ajcl/avab012

Danesi, M. (2010). *Pengantar memahami semiotika media*. Jalasutra.

DeepL. (n.d.). *DeepL Pro*. Retrieved January 10, 2024, from https://www.deepl.com/en/pro

Echols, J. M., Collins, J. T., Shadily, H., & Wolff, J. U. (1997). *Kamus Indonesia Inggris: An Indonesian-English dictionary*. Gramedia.

Egwuogu, C. (2024). *DeepL vs Google Translate: Which is the better translator?* Centus. Retrieved May 25, 2024, from https://centus.com/blog/deepl-vs-google-translate

Eiseman, F. B. (2009). *Bali: Sekala & Niskala: Essays on religion, ritual, and art*. Tuttle Publishing.

Gaspari, F., Almaghout, H., & Doherty, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. *Perspectives*, *23*(3), 333–358. https://doi.org/10.1080/0907676X.2014.979842

Gramedia. (n.d.). *Tarian Bumi*. Gramedia. Retrieved July 6, 2024, from https://www.gramedia.com/products/tarian-bumi-3

Hahn, S. (2010). *Signs of life: 40 Catholic customs and their biblical roots*. Doubleday.

Hasyim, M., Saleh, F., Yusuf, R., & Abbas, A. (2021). Artificial intelligence: Machine translation accuracy in translating French-Indonesian culinary texts. *International Journal of Advanced Computer Science and Applications*, *12*(3), 186–191. https://doi.org/10.14569/IJACSA.2021.0120323

Hidalgo-Ternero, C. M. (2021). Google Translate vs. DeepL: Analysing neural machine translation performance under the challenge of phraseological variation. *MonTI Monografías de Traducción e Interpretación, Special Issue* (6), 154–177. https://doi.org/10.6035/MonTI.2020.ne6.5

Jakobson, R. (2012). On linguistic aspects of translation. In L. Venuti (Ed.), *The translation studies reader* (3rd ed., pp. 126–131). Routledge.

Karabayeva, I., & Kalizhanova, A. (2024). Evaluating machine translation of literature through rhetorical analysis. *Journal of Translation and Language Studies*, *5*(1), 1–9. https://doi.org/10.48185/jtls.v5i1.962

KBBI. (n.d.). Griya. In *Kamus Besar Bahasa Indonesia Daring*. Retrieved May 29, 2025, from https://kbbi.web.id/griya

KBBI. (n.d.). Sang. In *Kamus Besar Bahasa Indonesia Daring*. Retrieved June 3, 2025, from https://kbbi.web.id/sang

Krippendorff, K. (2022). *Content analysis: An introduction to its methodology*. SAGE Publications.

Larson, M. L. (1998). *Meaning-based translation: A guide to cross-language equivalence* (2nd ed). University Press of America.

Macketanz, V., Burchardt, A., & Uszkoreit, H. (2018). *TQ-AutoTest: Novel analytical quality measure confirms that DeepL is better than Google Translate* [White paper]. Globalization and Localization Association (GALA). Retrieved May 29, 2024, from https://www-live.dfki.de/fileadmin/user_upload/import/10174_TQ-AutoTest_Novel_analytical_quality_measure_confirms_that_DeepL_is_better_than_Google_Translate.pdf

Marlina. (2024). *Resensi novel Tarian Bumi karya Oka Rusmini*. Mediajabar.com. Retrieved June 30, 2024, from https://www.mediajabar.com/sastra/resensi-novel-tarian-bumi-karya-oka-rusmini

Mayhew, S., Tsygankova, T., & Roth, D. (2019). *NER and POS when nothing is capitalized* [Preprint]. *arXiv*. https://arxiv.org/abs/1903.11222

Merriam, S. B., & Tisdell, E. J. (2015). *Qualitative research: A guide to design and implementation* (4th ed). John Wiley & Sons Inc.

Merriam-Webster. (n.d.). Arrack. In Merriam-Webster.com dictionary. Retrieved July 1, 2025, from https://www.merriam-webster.com/dictionary/arrack

Merriam-Webster. (n.d.). One. In *Merriam-Webster.com dictionary*. Retrieved July 15, 2025, from https://www.merriam-webster.com/dictionary/one

Microsoft. (2023). *Real-time, translated conversations with captions*. Microsoft Translator for Consumers. Retrieved May 29, 2024, from https://www.microsoft.com/en-us/translator/personal/

Moorkens, J. (2017). Under pressure: Translation in times of austerity. *Perspectives*, *25*(4), 464–477. https://doi.org/10.1080/0907676X.2017.1285331

Moorkens, J. (2018). What to expect from neural machine translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, *12*(4), 375–387. https://doi.org/10.1080/1750399X.2018.1529802

Munthe, I. B., Sipayung, K., & Lestari, F. D. (2023). Comparing the translation accuracy between Google Translate and professional translator. *Innovative: Journal of Social Science Research*, *3*(6), 890–903. https://j-innovative.org/index.php/Innovative/article/view/6342

Nababan, M., Nuraeni, A., & Sumardiono. (2012). Pengembangan model penilaian kualitas terjemahan. *Kajian Linguistik Dan Sastra*, *24*(1), 39–57. https://doi.org/10.23917/kls.v24i1.101

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*. Prentice Hall International.

Nida, E., & Taber, C. (2003). *The theory and practice of translation (fourth impression)*. Brill.

Olk, H. M. (2013). Cultural references in translation: a framework for quantitative translation analysis. *Perspectives*, *21*(3), 344–357. https://doi.org/10.1080/0907676X.2011.646279

Oxford Learner's Dictionaries. (n.d.) Kebaya. In *Oxford Learner's Dictionaries dictionary*. Retrieved January 10, 2025, from https://www.oxfordlearnersdictionaries.com/definition/english/kebaya?q=kebaya

Oxford University Press. (n.d.). *The Oxford 3000™ (American English)*. Oxford University Press. Retrieved June 3, 2025, from https://www.oxfordlearnersdictionaries.com/external/pdf/wordlists/oxford-3000-5000/American_Oxford_3000.pdf

Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th annual meeting of the Association for Computational Linguistics* (pp. 311–318). Association for Computational Linguistics. https://doi.org/10.3115/1073083.1073135

Partami, N. L., Sudiana, I. M., Sukayana, I. N., & Purwiati, I. A. M. (2016). *Kamus Bali—Indonesia*. Balai Bahasa Bali.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice*. SAGE Publications, Inc.

Reverso. (n.d.). Kebaya. In *Reverso Dictionary*. Retrieved June 10, 2025, from https://dictionary.reverso.net/english-definition/kebaya

Rysz, P. (n.d.). *Find out what Google Translate means - in online marketing. Click and read more*. Localo. Retrieved May 9, 2024, from https://localo.com/marketing-dictionary/what-is-google-translate

Sharma, R., Katyayan, P., & Joshi, N. (2023). Improving the quality of neural machine translation through proper translation of name entities. In *2023 6th International Conference on Information Systems and Computer Networks (ISCON)*, 1–4. https://doi.org/10.1109/ISCON57294.2023.10111938

Takai, K., Hattori, G., Yoneyama, A., Yasuda, K., Sudoh, K., & Nakamura, S. (2021). Named entity-factored transformer for proper noun translation. In S. Bandyopadhyay, S. L. Devi, & P. Bhattacharyya (Eds.), *Proceedings of the 18th International Conference on Natural Language Processing (ICON)* (pp. 7–11). NLP Association of India (NLPAI). https://aclanthology.org/2021.icon-main.2/

UNESCO. (2024). *Kebaya: Knowledge, skills, traditions and practices. Representative List of the Intangible Cultural Heritage of Humanity*. Retrieved July 5, 2025, from https://ich.unesco.org/en/RL/kebaya-knowledge-skills-traditions-and-practices-02090

Venuti, L. (2008). *The translator's invisibility: a history of translation*. Routledge.

Wiggins, D. (2023, March 6). *What is neural machine translation (NMT)?* Omniscien Technologies. Retrieved May 28, 2024, from https://omniscien.com/faq/what-is-neural-machine-translation/

Wyndham, A. (2021). *Inside Deepl: The world's fastest-growing, most secretive machine translation company*. Slator. Retrieved May 23, 2024, from https://slator.com/inside-deepl-the-worlds-fastest-growing-most-secretive-machine-translation-company/

Xamrokulovna, O. G. (2023). Proper noun and its interpretation in linguistics. *JournalNX - A Multidisciplinary Peer Reviewed Journal*, *9*(11), 178–182. https://repo.journalnx.com/index.php/nx/article/view/5147

Yıldız, M. (2021). A translation quality assessment tool proposed. *Amasya Üniversitesi Sosyal Bilimler Dergisi (ASOBİD)*, *10*, 237–266.