P-ISSN : 0853-1692 E-ISSN : 2579-5848
Buletin Al-Turas
website : bit.ly/buletinalturas

Evaluating Machine Translation of Cultural Terms: Readability Comparison Between Google and Yandex

Diana Mentari
Universitas Pendidikan Indonesia
Bandung, Indonesia
dianamentari90@gmail.com

Abstract

Purpose
This study aimed to analyze the readability of Google Translate (GT) and Yandex Translate (YT) translation results on dialogue texts containing cultural terms from the book Antologi Cerita Anak Indonesia (ACAI). This study evaluated the effectiveness of the Neural Machine Translation (NMT) approach in GT and the Hybrid Machine Translation (HMT) approach in YT in conveying text meanings clearly and comprehensibly to readers.

Method
This research employed a cloze test involving 28 participants aged 18-24 years, along with a questionnaire to assess user preferences regarding GT and YT translation results. Text readability was analyzed using the Flesch-Kincaid Grade Level and Gunning Fog Index to measure the linguistic complexity of the translations.

Results/Findings
The study results show that GT's readability reaches 81.1%, while YT's readability is 74.5%, both categorized as the independent level according to Rankin & Culhane's (1969) criteria. Additionally, 80% of the 20 questionnaire respondents stated that GT's translations were clearer than those of YT. Analysis using the Flesch-Kincaid Grade Level and Gunning Fog Index shows that the readability level of GT and YT translations is classified as advanced suitable for readers with a minimum education level equivalent to a bachelor's degree.

Conclusion
This study showed that GT has a higher readability level than YT, which might be because of its use of NMT, producing more natural sentence structures. Meanwhile, YT, which also relied on SMT, translates based on statistical patterns, making its translations more rigid. Although both systems could produce comprehensible translations, they still struggled with accurately translating cultural terms without additional context. Therefore, human involvement remained essential to improving accuracy and contextual appropriateness in machine translation.

Keywords
Comparison, Readability, Google Translate, Yandex Translate, Cultural Terms.

*) Corresponding Author

DOI: https://doi.org/10.15408/bat.v31i1.44469 *Print ISSN: 0853-1692 E-ISSN: 2579-5848*

Abstrak

Tujuan
Penelitian ini bertujuan untuk menganalisis keterbacaan hasil terjemahan Google Translate (GT) dan Yandex Translate (YT) pada teks dialog yang mengandung istilah budaya dari buku Antologi Cerita Anak Indonesia (ACAI). Selain itu, penelitian ini mengevaluasi efektivitas pendekatan Neural Machine Translation (NMT) pada GT dan Hybrid Machine Translation (HMT) pada YT dalam menyampaikan makna teks secara jelas dan dapat dipahami oleh pembaca.

Metode
Metode penelitian ini menggunakan tes cloze yang melibatkan 28 peserta berusia 18-24 tahun serta kuesioner untuk menilai preferensi pengguna terhadap hasil terjemahan GT dan YT. Selain itu, keterbacaan teks juga dianalisis menggunakan Flesch-Kincaid Grade Level dan Gunning Fog Index guna mengukur tingkat kompleksitas linguistik dari hasil terjemahan.

Hasil/Temuan
Hasil penelitian menunjukkan bahwa keterbacaan GT mencapai 81,1%, sedangkan YT mencapai 74,5%, yang keduanya termasuk dalam kategori independent level berdasarkan kriteria Rankin & Culhane (1969). Selain itu, 80% dari 20 responden kuesioner menyatakan bahwa hasil terjemahan GT lebih jelas dibandingkan YT. Analisis dengan Flesch-Kincaid Grade Level dan Gunning Fog Index mengindikasikan bahwa teks hasil terjemahan GT dan YT berada pada tingkat keterbacaan advanced, yang umumnya sesuai untuk pembaca dengan pendidikan minimal setara sarjana.

Kesimpulan
Penelitian ini menunjukkan bahwa GT memiliki tingkat keterbacaan lebih tinggi dibandingkan YT, hal ini dapat dikarenakan penggunaan NMT yang menghasilkan struktur kalimat lebih alami. Sementara itu, YT yang juga mengandalkan SMT, menerjemahkan berdasarkan pola statistik sehingga cenderung menghasilkan terjemahan yang lebih kaku. Meskipun kedua sistem mampu menghasilkan terjemahan yang dapat dipahami, keduanya masih mengalami kesulitan dalam menerjemahkan istilah budaya secara akurat tanpa konteks tambahan. Oleh karena itu, keterlibatan manusia tetap diperlukan untuk meningkatkan akurasi dan kesesuaian kontekstual dalam penerjemahan MT.

Kata Kunci
Perbandingan, Keterbacaan, Google Translate, Yandex Translate, Istilah Budaya .

الملخص

الهدف
هدف هذا البحث إلى تحليل مقروئية نتائج الترجمة في جوجل ترانسليت (GT) وياندیکس ترانسليت (YT) للنصوص الحواريّة التي تحتوي على مصطلحات ثقافية من كتاب *أنتولوجي تشريتا أناك إندونيسيا (ACAI)*. بالإضافة إلى ذلك، يقيّم البحث فعالية نهج الترجمة الآلية العصبية (NMT)في جوجل ترانسليت ونهج الترجمة الآلية الهجينة (HMT) في ياندیکس ترانسليت في نقل معاني النصوص بوضوح وإمكانية فهمها من قبل القرّاء.

الطريقة
يعتمد هذا البحث على اختبار *cloze* بمشاركة 28 متطوعًا تتراوح أعمارهم بين 18-24 عامًا، بالإضافة إلى استبيان لتقييم تفضيلات المستخدمين بشأن نتائج الترجمة في جوجل ترانسليت وياندیکس ترانسليت. علاوة على ذلك، يتم تحليل مقروئية النصوص باستخدام *Flesch-Kincaid Grade Level* و*Gunning Fog Index* لقياس مدى تعقيد اللغة في النصوص المترجمة.

النتائج
أظهرت نتائج الدراسة أن مقروئية جوجل ترانسليت بلغت 81.1%، بينما بلغت مقروئية ياندیکس ترانسليت 74.5%، وكلاهما يقع ضمن مستوى الاستقلال (*independent level*) .وفقًا لمعايير Rankin & Culhane (1969)بالإضافة إلى ذلك، أفاد 80% من بين 20 مشاركًا في الاستبيان بأن ترجمات جوجل ترانسليت كانت أوضح من ترجمات ياندیکس ترانسليت. كما أشارت التحليلات باستخدام *Flesch-Kincaid Grade Level* و*Gunning Fog Index* إلى أن مستوى مقروئية النصوص المترجمة في كل من جوجل ترانسليت وياندیکس ترانسليت يقع ضمن مستوى متقدم، وهو مناسب بشكل عام للقرّاء الذين يمتلكون مستوى تعليميًّا يعادل درجة البكالوريوس على الأقل.

الخلاصة
أظهرت هذه الدراسة أن جوجل ترانسليت يتمتع بمستوى مقروئية أعلى من ياندیکس ترانسليت، وقد يكون ذلك بسبب استخدامه لنظام الترجمة الآلية العصبية(NMT) ، مما يسمح بإنتاج تراكيب جمل أكثر طبيعية. بينما يعتمد ياندیکس ترانسليت أيضًا على الترجمة الآلية الإحصائية (SMT)، والتي تستند إلى الأنماط الإحصائية، مما يجعل ترجماته أكثر صرامة وأقل مرونة. وعلى الرغم من أن كلا النظامين قادران على إنتاج ترجمات مفهومة، إلا أنهما لا يزالان يواجهان صعوبات في ترجمة المصطلحات الثقافية بدقة دون سياق إضافي. لذا، يبقى تدخل العنصر البشري ضروريًا لتحسين الدقة والملاءمة السياقية في الترجمة الآلية..

الكلمات الرئيسية
المقارنة؛ القابلية للقراءة؛ ترجمة جوجل؛ ترجمة ياندكس؛ المصطلحات الثقافية

Print ISSN: 0853-1692 E-ISSN: 2579-5848 DOI: https://doi.org/10.15408/bat.v31i1.44469

INTRODUCTION

The process of conveying cultural quirks and meaning from the source language (SL) to the target language (TL) while preserving an equivalent message is known as translation (Nida & Taber, 1969). One of the significant challenges in translation is preserving cultural nuances, particularly in automated machine translation (MT). Despite advances in Natural Language Processing (NLP), MT still struggles with translating culturally embedded terms accurately (Aleshko-Ozhevskaya & Ternovskaya, 2023).

The evolution of MT began with Statistical Machine Translation (SMT), which relies on probability-based models to generate translations (Sall et al., 2013). However, SMT struggles with contextual accuracy, leading to the adoption of Neural Machine Translation (NMT), which leverages deep learning to enhance fluency and coherence (Nonghuloo & Rao, 2020; Dwivedi et al., 2023) . In contrast, Hybrid Machine Translation (HMT) combines SMT and NMT to improve translation accuracy, particularly in handling complex sentence structures (Dugonik et al., 2023). NMT, used in Google Translate (GT), applies deep learning to improve fluency and coherence in translations (Janiesch et al., 2021). Meanwhile, Yandex Translate (YT) employs a HMT system, which integrates SMT and NMT techniques to enhance translation accuracy, despite these technological improvements, both systems face challenges in translating cultural expressions and terminology (Mentari et al., 2024).

Readability is a crucial aspect of translation quality, as it determines how easily a translated text can be understood by readers. Flesch (1949) defines readability as a combination of text simplicity and reader-friendliness. Various metrics, such as sentence structure, word choice, and coherence, influence readability (Gunning, 1971; Sibeko, 2024). The cloze test is commonly used to measure readability by assessing how well participants can predict missing words in a passage (Taylor, 1953). This study applies the cloze test to evaluate the readability of GT and YT translations of Indonesian texts containing cultural terms.

Previous research on MT has focused on linguistic accuracy and technical aspects rather than readability and user comprehension. Limited studies have examined how different MT systems affect readability, particularly in texts with strong cultural elements. This study aims to fill that gap by comparing the readability of GT (NMT) and YT (HMT) translations of culturally rich Indonesian texts. By analyzing readability scores and user preferences, this research provides insights into the effectiveness of both MT approaches in conveying cultural meanings. The findings will contribute to the development of more advanced translation technologies that better preserve cultural and contextual accuracy.

METHOD

The method used for this study is a qualitative descriptive technique. This kind of method applies to qualitative data in order to characterize, clarify, and comprehend occurrences or phenomena. This study's data collection methods include questionnaires, interviews, document analysis, and cloze tests. The researcher uses questionnaires to get feedback from readers and users of the YT and GT translations in order to help the readability study and identify the machine translation with the highest quality. Participants are chosen through interviews in order to measure readability using the cloze test on the GT and YT translations.

The sources of research data are investigated using document analysis. In order to assess the texts' readability and the reader's comprehension of the GT and YT translations from the data sources, the cloze test is used in this investigation. The results of the cloze test also aid in assessing how well each translation tool captures the original text's meaning and cultural quirks.

When developing cloze tests, the techniques of John Haskall (as cited in Dewi, 2014) and Taylor (1953) are combined. The first step is to select a text that will be assessed for

DOI: https://doi.org/10.15408/bat.v31i1.44469 Print ISSN: 0853-1692 E-ISSN: 2579-5848

readability and is longer than 250 words. Following that, every seventh word, apart from digits, foreign words, and proper names, is replaced with dots or horizontal lines. Participants respond the test considering the context of the sentence. The cloze test results will be evaluated according to the criteria set by Hardjasujana et al. (1999) in order to discover the level of readability and reader comprehension.

Besides qualitative analysis, this study also integrates objective readability metrics to complement subjective assessments. Computational readability formulas such as the Flesch-Kincaid Grade Level and Gunning Fog Index are used to measure text complexity. These formulas evaluate sentence length, word complexity, and syllable count, providing an additional layer of evaluation. Comparative linguistic analysis is conducted to examine structural differences between GT and YT translations, focusing on aspects such as syntax, morphology, structure, and semantic accuracy. The researcher examined the data using the content analysis method (Krippendorff, 2019). This is because the primary focus of this study is the content of dialogue excerpts from the ACAI text, which serves as the source of the research data, and their translations because of automatic translation using GT and YT. Identifying relevant sentences, recording data, categorizing, and assessing interpretations and conclusions are all systematic steps in the content analysis process.

FINDING AND DISCUSSION

The readability quality of translated materials is assessed in this study using dialogue excerpts that contain cultural terms from the Indonesian anthology book ACAI and Achmad Jauhari Umar's (AJU) Arabic translation. These translations are then back-translated into Indonesian using a Neural Machine Translation (NMT) tool called GT and a Hybrid Machine Translation (HMT) tool called YT. Readability is assessed using the cloze test.

The Flesch grade level hypothesis is used by the researcher to assess reading proficiency in order to select the respondents who will take the cloze exam. The researcher uses Flesch's Readable (2020) readability grade level criteria to count the syllables, words, and sentences in the data sources for this study in order to establish the reading level across age ranges.

The data sources used in this study, which are dialogue sentences with cultural terminology, are classified as advanced level based on the measures (Bilal, 2013). According to developmental phases according to school levels, this suggests that the data sources are understood by people with an academic background or those who have finished three prior educational levels elementary school, junior high school, and senior high school (Torgesen dkk., 2017). These individuals range in age from 18 to 24 years (Ajhuri, 2019).

After determining that the participants who can be included in this study are students aged 18-24 years old, the researcher selected again using the Purposive Sampling technique with the determination that the participants would be undergraduate students who had previously used translation tools to translate foreign texts and participate in BTQ (Bimbingan Tahfidz Qur'an or guidance to memorize the Qur'an) activities at Indonesia University of Education in 2024 and these participants were interviewed by the researcher. Then criteria for evaluating the results of the Clowes test based on Rankin & Culhane (1969) criteria were used to measure the reading level of the participants. The evaluation criteria used in this research are shown in the following Table 1.

Table 1. Evaluation Criteria Based on Rankin & Culhane (1969)

Reading Level	Score Range	Description
Independent	>60%	The reader can read fluently, independently, and understand the text without assistance from others.
Instructional	40-60%	The reader requires assistance to understand the text but is still able to learn from it.
Frustration	<60%	The reader experiences serious difficulties in understanding the text and feels frustrated.

Print ISSN: 0853-1692 E-ISSN: 2579-5848 DOI: https://doi.org/10.15408/bat.v31i1.44469

Table 1 shows the categorization of readers' comprehension levels based on the percentage of readability of texts. If the readability percentage is over 60%, the reader is categorized as independent, which means they can read fluently and independently and understand the text with no need help from another party. If the percentage ranges from 40% to 60%, the reader is categorized as instructional, where they need help to understand the text, but can still get information from the text. If the percentage is less than 40%, the reader is categorized as frustrated, indicating that they are having great difficulty understanding the text and feel frustrated without additional support (Shaye, 2021). Based on this categorization, the reader's comprehension level is based on their readability score. An independent level shows good comprehension without the need for help, an instructional level shows comprehension that depends on help or context, and a frustrated level shows significant difficulties in comprehending texts. The results of the findings and their discussion will be explained in the following subchapters.

Readability of GT Translations Based on Neural Machine Translation

As shown in Table 2, the results of the cloze test on texts translated using GT for 5 participants indicate that the average readability of the texts reached 73.14%. This ratio illustrates the extent to which readers can understand the text translated by GT without requiring additional context. According to Qudah et al. (2020), this suggests that the text can be easily understood without help. This data was obtained through a test involving five participants, where they were asked to complete texts with certain keywords removed. The results provide insights into GT's effectiveness in producing texts that can be easily understood by readers.

Table 2. Results of Cloze Test 1

Gender	Age	Cloze Test Result
Male	19	74,3
Female	19	94,3
Female	20	51,4
Female	18	77,1
Female	19	68,6
Average		**73,14**

The average readability ratio in Table 2 places the GT-translated text in the independent category, based on the evaluation criteria of Rankin & Culhane (1969). After obtaining the readability results of the GT text from the first 5 participants, the researcher added 5 new participants to determine whether the cloze test results from the previous 5 participants would affect the overall readability ratio. A total of 5 new participants were added to take the cloze test on the GT-translated text, bringing the total number of participants to 10. The cloze test conducted with 10 participants showed an average readability percentage of 78%, as presented in Table 3. According to Olusola (2013), a readability score of 78% falls within the independent level, indicating that the text can be well understood without requiring additional assistance. This result shows an increase in readability by 4.86% compared to the previous findings. Below are the data obtained from the test results.

This average percentage in Table 3 categorizes the readability of the translated text using GT within the independent level, according to Rankin & Culhane's (1969) evaluation criteria. The researcher then conducted the Clowes test again with 4 new participants. This was done to ensure the reliability and validity of the text readability measure. The results of the Clowes test for the three times adding 14 participants showed that the mean percentage of readability reached 81.1%, indicating a 3.1% increase in readability compared to the previous tests as shown in Table 4.

The average percentage of text reading in Table 4 is 81.1%, indicating that the text translated by GT falls within the independent level (Shaye, 2021). This is in line with the evaluation criteria established by Rankin and Culhane (1969), where independent level

DOI: https://doi.org/10.15408/bat.v31i1.44469 Print ISSN: 0853-1692 E-ISSN: 2579-5848

shows that the reader can understand the text without the need for additional help or guidance. Based on the data obtained, it can be concluded that the percentage of reading the texts translated by GT tends to increase as the number of participants in the cloze test increases. It is possible that each participant has different abilities in completing the missing words, depending on their experience and cognitive background and their language skills. Thus, this diversity in understanding contributes to improving the overall average readability scores of the GT translated texts.

Based on the research conducted on the readability of GT texts on dialog excerpts containing cultural terms from the ACAI book, the average readability score of 14 participants was found to be 81.1% when the Clowes test was administered. Based on Rankin & Culhane (1969) assessment criteria, it can be concluded that the level of readability of the texts falls under the independent classification. That is, the reader can read smoothly and understand the text without needing help from others.

It is also important to mention that even though the readability of texts has reached the independent level, there are some things that need to be taken into account for the results of GT translation. Although machine translation technologies have made great progress in automatic translation, there may still be errors or inaccuracies in interpreting cultural terms (Mentari et al., 2024). Therefore, it is recommended that the reader remain critical and consider the original context and revise texts if necessary to ensure correct understanding of the text. This study provides valuable insights into the effectiveness of MT in certain contexts and can serve as a basis for future research in this area.

Table 3. Results of Cloze Test 2

Gender	Age	Cloze Test Result
Male	19	74,3
Female	19	94,3
Female	20	51,4
Female	18	77,1
Female	19	68,6
Female	19	51,4
Female	20	88,6
Female	22	80
Female	19	100
Female	18	94,3
Average		**78**

Table 4. Results of Cloze Test 3

Gender	Age	Cloze Test Result
Male	19	74,3
Female	19	94,3
Female	20	51,4
Female	18	77,1
Female	19	68,6
Female	19	51,4
Female	20	88,6
Female	22	80
Female	19	100
Female	18	94,3
Male	22	88,57
Female	23	85,71
Female	22	65, 71
Female	23	100
Average		**81,1**

Print ISSN: 0853-1692 E-ISSN: 2579-5848 DOI: https://doi.org/10.15408/bat.v31i1.44469

Readability of YT Translations Based on Hybrid Machine Translation

Analyzing the readability of YT translated texts in sentences containing cultural terms in the ACAI book provides an interesting picture of translation quality. The research results showed a readability rate of 74.5%, which places them in the independent category based on interpreting the cloze test (Rankin & Culhane, 1969). This shows that readers can understand the texts without additional help from other parties or can read them independently.

Based on Table 5, the results of the study on the readability of translated texts using YT through the cloze test with 14 participants, students aged between 18- and 23-years old studying at the undergraduate level at Indonesia University of Education. The participants were selected using the same method as in the GT translated text cloze test, and the cloze test for YT translated texts was conducted using a similar procedure. The results of the cloze test for GT translated texts with 5 participants showed that the average percentage of readability of the texts reached 88.58%.

Table 5. Results of Cloze Test 4

Gender	Age	Cloze Test Result
Female	18	94,3
Female	18	94,3
Female	18	94,3
Female	19	88,6
Female	19	71,4
Average		**88,58**

The average readability percentage in Table 5 was obtained by categorizing the text translated by YT within the independent level. This categorization is based on the evaluation criteria developed by Rankin and Culhane (1969), where the independent level shows that the reader can understand the text without the need for additional help or context. In the first phase of the research, the readability of the text translated by YT was tested using a cloze test involving 5 participants. After getting the initial results, the researcher decided to add 5 more participants to conduct the same test. The purpose of this addition was to check the consistency and validity of the previous results, as well as to see if the average percentage of text readability would change as the number of participants increased. In this way, the research not only relies on data taken from a limited number of participants, but also takes into account how participant differences can affect the average text readability results.

The researcher decided to add 5 new participants to the cloze test conducted on the text translated by YT. With this addition, the total number of participants who took the completion test became 10. After administering the test to all participants, an average readability percentage of 76.01% was obtained as shown in Table 6. This result indicates a decrease in readability percentage by 12.57% compared to the previous average results of the completion test administered to only 5 participants. This decrease shows that participants' level of comprehension of YT-translated texts may vary based on their backgrounds, language skill levels, and personal experiences. This change also highlights the importance of involving more participants in the tests to get a more accurate picture of the readability level of YT-translated texts.

The average readability percentage in Table 6 places the translated texts using YT within the independent level according to Rankin & Culhane (1969). The researcher then conducted a cloze test with 4 new participants to check the reliability and validity of the test results. With the addition of the new participants, the average percentage of readability for the 14 participants was 74.5% as shown in Table 7.

The average readability ratio in Table 7 categorizes the text translated by YT as independent according to Rankin & Culhane's (1969) criteria. The results of the cloze test indicate that despite the presence of cultural terms in the ACAI texts, the high readability level indicates that the translation is relatively uncomplicated without additional help.

DOI: https://doi.org/10.15408/bat.v31i1.44469 Print ISSN: 0853-1692 E-ISSN: 2579-5848

This highlights YT's ability to convey meaning relatively clearly while maintaining the cultural and linguistic complexity of the original text. Although, according to Andriani et al. (2023), YT's translation results exhibit low accuracy, acceptability, and readability.

However, it should be noted that this research was limited to a small sample of respondents belonging to a specific age group and education level. Thus, the findings may not be directly applicable to a larger or more diverse group. This research provides valuable insights into the readability of subtitles using YT in a given context and shows that translated texts can be well understood by readers with higher educational backgrounds.

Table 6. Results of Cloze Test 5

Gender	Age	Cloze Test Result
Female	18	94,3
Female	18	94,3
Female	18	94,3
Female	19	88,6
Female	19	71,4
Female	19	71,4
Female	19	94,3
Male	19	20
Female	20	82,9
Female	21	48,57
Average		**76,01**

Table 7. Results of Cloze Test 6

Gender	Age	Cloze Test Result
Female	18	94,3
Female	18	94,3
Female	18	94,3
Female	19	88,6
Female	19	71,4
Female	19	71,4
Female	19	94,3
Male	19	20
Female	20	82,9
Female	21	48,57
Female	22	57,14
Female	23	82,86
Male	21	45,71
Female	22	97,14
Average		**74,5**

Comparison of Readability Between GT and YT

The results of the cloze test among 28 participants showed independent readability of 81.1% for GT and 74.5% for YT. Independence-level readability refers to the ability of readers to comprehend the GT and YT MT texts that were subjected to the completion test in this study, which are sentences containing dialog segments with cultural terms without the need for external assistance or can be understood autonomously (Hardjasujana et al., 1999).

This study reflects efforts to understand the effectiveness of GT's Neural Machine Translation (NMT) and YT's Hybrid Machine Translation (HMT) in cultural understanding. Translation that is uncomplicated and accurate requires not only the transfer of words between languages, but also the correct conveyance of meanings and intentions. The higher cloze test results for GT compared to YT show that translation using the Neural Machine Translation (NMT) approach has a higher level of readability compared to translation using YT's Hybrid Machine Translation (HMT) approach. Based on the results, the

Print ISSN: 0853-1692 E-ISSN: 2579-5848 DOI: https://doi.org/10.15408/bat.v31i1.44469

researchers disagree with Du & Way (2017); Grundkiewicz & Junczys-Dowmunt (2018), who indicated that HMT results are superior to NMT.

The study also included a questionnaire to assess users' preferences towards translation results from GT and YT. The questionnaire comprised four questions related to participants' use of GT and YT in translating foreign texts and the extent to which they felt these systems help them understand foreign texts. In addition, the questionnaire asked questions about how easy it is to understand the translation results without external help and which of the two systems participants felt provided better and easier-to-understand translations. A pie chart is presented in Figure 1, showing the percentage of users' ratings of GT and YT in terms of quality and ease of understanding.

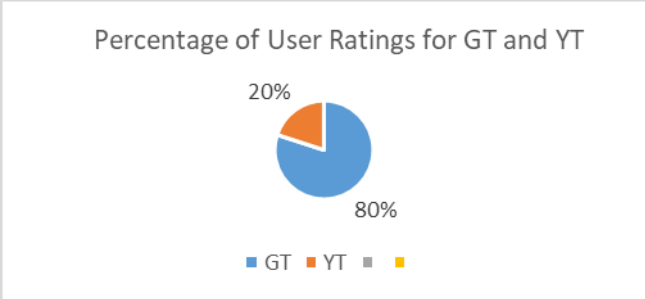

Figure 1. Percentage of user ratings for GT and YT

As available in Figure 1, the results of the questionnaire revealed that the majority of the participants, i.e. 80% of the total 20 participants, believe that the translation results from GT are better and clearer compared to the results from YT. This result reinforces the results of the cloze test conducted on 28 participants where the researcher tested the comprehension of the translated texts produced by both systems. If linked to the Flesch-Kincaid Grade Level and Gunning Fog Index, the translated texts of GT and YT are classified at the Advanced level, which is generally suitable for readers with a minimum education level equivalent to undergraduate (college) students or professionals, and aged 18 years and older (Flesch, 1949). This is also consistent with the research methodology in this study, which selected participants within the 18-24 age range, who had generally completed the three previous levels of education, elementary, junior high, and senior high school, (Torgesen et al., 2017; Ajhuri, 2019).

The measurement of text complexity in this study applies the Flesch-Kincaid Grade Level and Gunning Fog Index formulas. Flesch-Kincaid Grade Level calculates readability based on sentence length and the number of syllables per word (Flesch, 1949; Morony et al., 2015; Ojha et al., 2021; Sibeko, 2024), using the following equation:

$$\textbf{Flesch-Kincaid Grade Level} = (0.39 \times \text{ASL}) + (11.8 \times \text{ASW}) - 15.59$$

where ASL (Average Sentence Length), represents the average sentence length (words per sentence), and ASW represents the average number of syllables per word. Meanwhile, the Gunning Fog Index is used to estimate the number of years of formal education required to understand the text (Gunning, 1971; Gosselin et al., 2021; Yu & Gutt, 2024), using the formula:

$$\textbf{Gunning Fog Index} = 0.4 \times [(\text{ASL}) + 100 \times (\text{PHW})]$$

where PHW (Percentage of Hard Words), represents the percentage of words with three or more syllables (excluding proper nouns, technical terms, and common suffixes).

Although the readability results indicate that the text can be understood independently, they also suggest that the text may still be too complex for readers with lower educational

DOI: https://doi.org/10.15408/bat.v31i1.44469 Print ISSN: 0853-1692 E-ISSN: 2579-5848

levels. Therefore, language simplification or human intervention can be a solution to improve the accessibility of the text for a wider audience.

By combining participants' subjective assessments of translation quality with objective readability metrics such as the cloze test, Flesch-Kincaid Grade Level, and Gunning Fog Index, this study provides a more comprehensive perception of the efficiency and effectiveness of both systems in producing easily understandable texts. These results not only confirm users' preferences but also support the reliability of the objective test results, providing a strong basis for the conclusion that GT tends to be more effective in producing clearer and easier-to-understand translations. Although the study results show that GT has a higher level of readability compared to YT, this does not necessarily mean that GT's translations are better in all translation contexts. It should be noted that higher readability is often associated with easier linguistic comprehension, but this does not imply that GT is error-free. Researcher in this study still found translation errors, especially when dealing with terms that carry specific cultural connotations. For example, local cultural terms such as *egrang* (a traditional game using bamboo stilts for walking), *tradisi sasi*, a customary rule in Maluku and Papua that temporarily restricts the use of natural resources as a conservation effort, (Sumarsono & Wasa, 2019; Muin, Abdul; Rakuasa, 2023), *klepon*, a traditional Indonesian snack made of glutinous rice balls filled with liquid palm sugar and coated with grated coconut,(Sari et al., 2024); *lemper*, a dish made of glutinous rice filled with shredded chicken or beef floss, wrapped in banana leaves,(Prasetiyo, 2022; Alvionita, 2023; Sutami & Kabul, 2024); *haminjon*, a type of benzoin resin from North Sumatra used in traditional rituals and religious ceremonies, (Petersen, 2014; Sormin et al., 2021; Manalu, 2024); *blencong*, an oil lamp used in wayang kulit performances to create dramatic shadow effects, (Imanjaya, 2024); *tambaroro*, the customs and traditions of several communities in the Aru Islands, Indonesia, include songs and dances that are usually performed as part of the traditional Sasi opening ceremony, (Lewerissa et al., 2023); and *mendo*, a traditional Javanese frying technique for half-cooking battered tempeh, (Winarno et al., 2017; Lusiana et al., 2020; Fajarini & Suharto, 2022); etc. These terms were not accurately translated by either system, possibly because they are not fully included in the training data used to develop these systems. These errors in translating cultural contexts show that, although the systems may be relatively effective in addressing general linguistic challenges, they still face limitations in dealing with unique cultural complexities. Based on these findings, the researchers agree with Khoshafah's (2023) opinion that translation systems, including GT and YT, are still unable to accurately translate specific cultural elements, especially when it comes to local cultural terms that are only known in specific regions.

These findings emphasize the importance of human intervention in the translation process, especially when dealing with culturally rich texts. They also highlight the need for further evaluation of the quality of the translations produced by these systems, focusing on aspects such as accuracy, consistency, and appropriateness to the complex cultural context. In a broader context, this study provides insights into the evolution of machine translation technologies and their challenges related to understanding and reproducing cultural meanings and speech acts in an accurate manner. With an understanding of the readability results of texts translated using NMT and HMT approaches, the study shows the importance of driving the evolution of MT technologies to meet users' needs for deeper understanding of cross-cultural texts. Ultimately, this study underscores the necessity of improving MT systems by incorporating more culturally aware translation approaches, enhancing dataset inclusivity, and integrating human expertise to refine the accuracy of translations. Future research should explore hybrid translation models that combine the strengths of NMT and human-assisted translation to better address cultural and contextual challenges in machine translation.

Print ISSN: 0853-1692 E-ISSN: 2579-5848 DOI: https://doi.org/10.15408/bat.v31i1.44469

CONCLUSION

This study shows that both GT and YT can produce comprehensible translations; however, GT has a higher readability level (81.1%) compared to YT (74.5%), both classified as independent, showing that readers can understand the texts without assistance. A total of 80% of participants stated that GT's translations were easier to understand than those of YT. The higher readability of GT may be because of the dominant use of NMT in its system, allowing for more natural sentence structures that align with human language patterns, while YT employs HMT, which still relies heavily on SMT. SMT translates based on statistical patterns without fully considering contextual meaning, making YT's translations often feel more rigid and less coherent. Additionally, GT benefits from a larger training dataset due to its more frequent global usage, enabling it to produce more accurate and fluent translations than YT. However, both systems still struggle to accurately translate local cultural terms without additional context, emphasizing the crucial role of human involvement in ensuring accuracy and contextual appropriateness. To improve cultural translation, MT should not only transliterate cultural terms but also include them in their original script to avoid misinterpretation. Further research is needed to optimize the integration of human intelligence and MT in translating texts with complex cultural elements.

ACKNOWLEDGEMENT

The author would like to extend their gratitude and appreciation to Prof. Dr. Mohamad Zaka Al Farisi and Hikmah Maulani, S.Pd., M.Pd., lecturers in Arabic Language Education at Indonesia University of Education, for their support in the preparation of this article.

REFERENCES

Ajhuri, K. F. (2019). Psikologi Perkembangan: Pendekatan Sepanjang Rentang Kehidupan. In Lukman (Ed.), *Psikologi Perkembangan Pendekatan Sepanjang Rentang Kehidupan*. Penebar Media Pustaka.

Aleshko-Ozhevskaya, S. S., & Ternovskaya, I. L. (2023). Titles of Acts: Translation from English into Russian. *NSU Vestnik*, *21*(4), 144. https://doi.org/10.25205/1818-7935-2023-21-4-143-155

Alvionita, S. (2023). *Cultural Words Translation Analysis From English Into Indonesian in Burnt Movie*. Univeristas Nasional.

Andriani, R., Eriyanti, R. W., & Huda, A. M. (2023). Problem dalam Menggunakan Mesin Terjemahan: Error dalam Menterjemahkan Teks Bahasa Inggris ke dalam Bahasa Indonesia. *INNOVATIVE: Journal Of Social Science Research Volume*, *3*.

Bilal, D. (2013). Comparing Google's Readability of Search Results to the Flesch Readability Formulae : A Preliminary Analysis on Children's Search Queries. *Proceedings of the 76th ASIS&T Annual Meeting*, 1–9.

Dewi, R. P. (2014). Tingkat Keterbacaan Buku Teks cakap Berbahasa Indonesia SMP Kelas VII pada SMP Budya Wacana dan SMP Don Bosco Yogyakarta. *Widya Dharma: Jurnal Kependidikan*, *26*(2), 201–221.

Du, J., & Way, A. (2017). Neural Pre-Translation for Hybrid Machine Translation. *In: MT Summit XVI - 16th Machine Translation Summit*, 27–37.

Dugonik, J., Maučec, M. S., Verber, D., & Brest, J. (2023). Reduction of Neural Machine Translation Failures by Incorporating Statistical Machine Translation. *Mathematics*, *11*

DOI: https://doi.org/10.15408/bat.v31i1.44469 Print ISSN: 0853-1692 E-ISSN: 2579-5848

(11), 2484. https://doi.org/10.3390/math11112484

Dwivedi, R. K., Nand, P., & Pal, O. (2023). Evolution of Machine Translation for Indian Regional Languages using Artificial Intelligence. *2023 International Conference on Disruptive Technologies (ICDT)*, 764–768. https://doi.org/10.1109/ICDT57929.2023.10150776

Fajarini, A., & Suharto, A. W. B. (2022). Early Childhood Learning Based on Local Wisdom of The Banyumas Region in Early Childhood Education Institutions in RA Masyithoh 13 Sokaraja Lor. *International Proceedings of Nusantara Raya*, *1*(1), 186–193. https://doi.org/10.24090/nuraicon.v1i1.126

Flesch, R. (1949). *The Art of Readable Writing*. Harper & Row.

Gosselin, A. M., Maux, J. Le, & Smaili, N. (2021). Readability of Accounting Disclosures: A Comprehensive Review and Research Agenda. *Accounting Perspectives*, *20*(4), 543–581.

Grundkiewicz, R., & Junczys-Dowmunt, M. (2018). *Near Human-Level Performance in Grammatical Error Correction with Hybrid Machine Translation Roman*.

Gunning, R. (1971). *The Technique of Clear Writing* (Revised). McGraw-Hill.

Hardjasujana, A., Suriamiharja, A., Kuswari, U., Koswara, D., & Nurjanah, N. (1999). *Evaluasi Keterbacaan Buku Teks Bahasa Sunda untuk Sekolah Dasar di Jawa Barat*. Pusat Pembinaan dan Pengembangan Bahasa.

Imanjaya, E. (2024). Wayang Kulit Show, Layar Tancap (Traveling Cinema Show), and History of Pre-Cinema. *Journal of Art, Film, Television, Animation, Games and Technology*, *3*(2), 7–19.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electron Markets*, *31*, 685–695. https://doi.org/10.1007/s12525-021-00475-2

Khoshafah, S. A. N. A. K. (2023). Effectiveness of Machine Translation in Rendering Yemeni Culture- Specific Items into English: Sana'ani Dialect as a Case-in-Point. *Sana'a University Journal of Human Sciences*, *5*(1), 637–651.

Krippendorff, K. (2019). *Content Analysis An Introduction to Its Methodology* (4th ed.). SAGE Publication, Inc. https://doi.org/https://doi.org/10.4135/9781071878781

Lewerissa, Y. A., Ayal, F. W., & Letsoin, Y. N. (2023). Efisiensi Kinerja Sasi Teripang Pasir (Holothuria scabra) Desa Tungu Kepulauan Aru. *PAPALELE: Jurnal Penelitian Sosial Ekonomi Perikanan Dan Kelautan*, *7*(1), 67–76. https://doi.org/https://doi.org/10.30598/papalele.2023.7.1.67

Lusiana, Y., Laksono, P. M., & Hariri, T. (2020). Self-Styling, Popular Culture, and the Construction of Global-Local Identity among Japanese Food Lovers in Purwokerto. *I-Pop: International Journal of Indonesian Popular Culture and Communication*, *1*(1), 21–40. https://doi.org/10.36782/i-pop.v1i1.33

Manalu, H. B. (2024). Itak Gurgur and its Meaning in the Batak Cultural Community. *Lincah: Jurnah Inovasi Dan Tren*, *2*(1), 11–18. https://doi.org/https://doi.org/10.35870/ljit.v2i1.2207

Mentari, D., Al Farisi, M. Z., & Maulani, H. (2024). Machine Translation Shifts on The Meaning Equivalence of Cullture Sentence and Illocuttionary Speech Act: Back-Translation. *Journal of Culture, Arts, Literature, and Linguistics*, *10*(1). https://doi.org/10.30872/calls.v10i1.15168

Morony, S., Flynn, M., McCaffery, K. J., Jansen, J., & Webster, A. C. (2015). Readability of Written Materials for CKD Patients: A Systematic Review. *American Journal of*

Print ISSN: 0853-1692 E-ISSN: 2579-5848 DOI: https://doi.org/10.15408/bat.v31i1.44469

Kidney Diseases, *65*(6), 842–850. https://doi.org/https://doi.org/10.1053/j.ajkd.2014.11.025

Muin, Abdul; Rakuasa, H. (2023). Sasi Laut as a Culture of Natural Resources Conservation to Overcome the Tragedy of the Commons in Maluku Province. *International Journal of Multidisciplinary Approach Research and Science*, *1*(3), 277–287. https://doi.org/10.59653/ijmars.v1i03.139

Nida, E. A., & Taber, C. R. (1969). *The Theory and Practice of Translation*. E.J. Brill.

Nonghuloo, M. S., & Rao, N. (2020). Analyses and Modeling of Neural Machine Translation for English to Khasi. *International Journal of Recent Technology and Engineering (IJRTE)*, *9*(2), 115–118. https://doi.org/10.35940/ijrte.B3175.079220

Ojha, P. K., Ismail, A., & Srinivasan, K. K. (2021). Perusal of readability with focus on web content understandability. *Journal of King Saud University - Computer and Information Sciences*, *33*(1), 1–10. https://doi.org/https://doi.org/10.1016/j.jksuci.2018.03.007

Olusola, F. J. (2013). Comparative analysis of readability level of basic six pupils in private and public schools in Ibadan Land. *Journal of Educational Development and Practice*, *4*(1), 105–117. https://doi.org/https://doi.org/10.47963/jedp.v4i.960

Petersen, D. (2014). *The Essential Oils of Indonesia*. American College of Healthcare Sciences.

Prasetiyo, A. (2022). The Phenomenon of Gotong Royong in Java Community: A Case Study Nyumbang. *Indonesian Journal of Multidisciplinary Science*, *1*(7), 792–801. https://doi.org/10.55324/ijoms.v1i7.145

Qudah, K., Muflih, M. K., & Sardiah, S. (2020). The level of readability of the computer sciences textbook among the eleventh grade students in Jordan. *Educational Research and Reviews*, *15*(2), 72–80. https://doi.org/https://doi.org/10.5897/ERR2019.3887

Rankin, E., & Culhane, J. (1969). Comparable Cloze and Multiple-Choice Comprehension Test Scores. *Journal of Reading*, *13*(3), 193–198.

Readable. (2020). *Flesch Reading Ease and the Flesch Kincaid Grade Level*. https://readable.com/readability/flesch-reading-ease-flesch-kincaid-grade-level/

Sall, S. B., Sharma, R., & Shedamkar. (2013). Example Based Machine Translation Using Natural Language Processing. *International Journal of Scientific & Engineering Research*, *4*(8).

Sari, I. J., Syafira, R., Zakkiya, Y. H., Ambarsari, R., & Saputra, O. (2024). Ethnophysics of Klepon: Exploring Physics Concepts in Traditional Pasuruan Snack. *International Journal of Research and Community Empowerment*, *2*(2), 48–55. https://doi.org/https://doi.org/10.58706/ijorce.v2n2.p48-55

Shaye, S. Al. (2021). Readability Level of Arabic Language Textbook of the Sixth Grade in the State of Kuwait. *Journal of Educational and Social Research*, *11*(4), 197–212.

Sibeko, J. (2024). Harnessing Google Translations to Develop a Readability Corpus for Sesotho: An Exploratory Study. *Journal of the Digital Humanities Association of Southern Africa (DHASA)*, *5*(1). https://doi.org/10.55492/dhasa.v5i1.5010

Sormin, E., Harefa, N., Purba, L. S. L., Sumiyati, & Nadeak, B. (2021). Benzoic Acid Isolation from Frankincense. *Proceedings of the 2nd Annual Conference on Blended Learning, Educational Technology and Innovation (ACBLETI 2020)*, 217–221.

Sumarsono, A., & Wasa, C. (2019). Traditional Sasi wisdom in Papua-based nature conservation. *IOP Conference Series Earth and Environmental Science*, 1–6. https://

DOI: https://doi.org/10.15408/bat.v31i1.44469 Print ISSN: 0853-1692 E-ISSN: 2579-5848

doi.org/10.1088/1755-1315/235/1/012092

Sutami, H., & Kabul, A. R. (2024). Culinary for The Spirits : The Reality of Cultural Acculturation of The Chinese Peranakan of Palembang. *MANDARINABLE: Journal of Chinese Studies Language, Literature, Culture, and Journalism*, *03*(01), 88–103. https://doi.org/https://doi.org/10.20961/mandarinable.v3i1.1091

Taylor, W. L. (1953). "Cloze Procedure": A New Tool For Measuring Readability. *Journalism Quarterly : Devoted to Research Studies in The Field of Mass Communications*. https://doi.org/10.1177/107769905303000401

Torgesen, J. K., Houston, D. D., Rissman, L. M., Decker, S. M., Roberts, G., Vaughn, S., Wexler, J., Francis, D. J., Rivera, M. O., & Lesaux, N. (2017). *Academic Literacy Instruction for Adolescents: A Guidance Document from the Center on Instruction*. Center on Instruction (RMC Research Corporation) Description:

Winarno, F. G., Ahnan-Winarno, A. D., & Ahnan, W. W. (2017). *Tempe: Kumpulan Fakta Menarik Berdasarkan Penelitian (Tempeh: A Collection of Interesting Facts based on Research)*. Gramedia Pustaka Utama.

Yu, Y., & Gutt, D. (2024). Review Helpfulness Scores vs. Review Unhelpfulness Scores: Two Sides of the Same Coin or Different Coins? *IEEE Transactions on Engineering Management.*, *71*, 8031–8044. https://doi.org/10.1109/TEM.2024.3384960

Received (19-01-2025) *Accepted (13-04-2025)* *Published (22-04-2025)*