

**COMPARATIVE STUDY ON CHATGPT VS GOOGLE TRANSLATE IN
INDONESIAN-ENGLISH OF BILINGUAL DESCRIPTION OF HISTORICAL
HERITAGE AT MUSEUM NEGERI NUSA TENGGARA BARAT**

Yuni Maulida Afifah SR¹, Baharuddin², Lalu Jaswadi Putera³, Lalu Ali Wardana⁴

¹Universitas Mataram, Jl.Majapahit No. 62, Mataram 83115, Indonesia

E-mail: ¹yunimaulida.asr06@gmail.com

ABSTRACT

*Cultural difference pose challenges for machine translation as it's challenging to find equivalent words to convey cultural concepts. Therefore, this research conducted to find out the accuracy level of two popular machine translation namely Chat-GPT and Google Translate in translating Culture-specific Items from the bilingual description of historical heritage at Museum Negeri Nusa Tenggara Barat. Using HTER by Snover (2006), researcher try to find the level of accuracy of the MT. After long analysis of the accuracy of MT namely Chat-GPT and GT using the HTER theory by Snover, researcher found that between Chat-GPT and GT almost play accurate with the presentation 99.94% and 99.93% while the error rate for Chat-GPT is 0.06% while GT is 0.07%. Furthermore, the quality index (TQI) of the MTs are in the same range, which is **EXCELLENT**.*

Keywords: *Chat-GPT, Google Translate, Culture-specific Items, Translation Accuracy*

A. Introduction

In the age of the advance technology, language barriers are not as challenging as before in terms of communicating between groups or individuals. The existence of machine translation (MT) allows global society to be able to communicate and access lots of information in many different languages. One of the most widely used machine translations by internet users is Google Translate (Çimke&YıldırımGürkan, 2021). Google Translate is a website used to translate multilingual texts including in the form of written text or voice recognition. Several written texts can

be translated using this machine translation like an expression, a text segment, and a web page.

Beside Google Translate, another popular machine translation that have gain wide popularity especially among students is ChatGPT. Also known as Chat Generative Pre-Trained Transformer, ChatGPT is an artificial intelligence (AI) chatbot developed by Open AI that uses a machine-based generative natural language model. ChatGPT has brought an incredible impact on natural language processing (NLP) tasks, and has been trained to follow instructions with human feedback

(Peng et al., 2023). Beside ChatGPT is designed as an intelligent conversational system, it integrates various abilities of natural language processing, including question answering, storytelling, logic reasoning, code debugging, machine translation (Jiao et al., 2023). Natural language processing is a part of Artificial Intelligence used for the analysis, understanding, and production of language that humans use to communicate with computers smart machines. Nevertheless, on his current work Jiao et al., (2023) reveals that ChatGPT with naive prompts has a significant performance advantage over other commercial translation systems such as Google Translate and DeepL Translate.

However, for some people, they might concern about the result of machine translation. It is because the level of accuracy of the translation results is still questionable. Yamada (2019), said that the result of machine translation is considered a draft of the translation. The result still needs to analyze, edit and re-read before using it. In translation activity, a translator should pay more attention to the word choice to make translation results readable and not ambiguous (Sumiati

et al., 2022). This issue existed due to the machine translator concept being a word-for-word translation (Sipayung, 2021).

Translating using machine translation sometimes is not accurate due to the different structure, grammar, lexical, type of text, situation, and also the word choice in a text especially in translating Cultural Specific Items (CSI). Referring to Aixelia, Culture-specific Items defines as "those textually actualized items whose functions and connotations in a source text involve a translation problem is a product of the non-existence of the referred item or of its different inter-textual status in the cultural system of the reader of the target text" (Aixela, 1996) cited in (Khaleghi Zavareh, 2021). He consistently emphasizes that culture-specific items (CSIs) are linguistic elements that pose translation challenges due to variations in cultural comprehension. A culture-specific items refers to an object, concept, practice, or idea that holds significance, relevance, or meaning within a particular culture or group of people. These items are often deeply rooted in the customs, traditions,

beliefs, and values of a specific society or community.

Newmark classify categories of culture-specific item into five categories. The division of CSI's are: (1) *Ecology (flora, fauna, winds, plains and hills)*, (2) *Material Culture (foods, clothes, houses, towns, and transports)*, (3) *Social Culture (work and leisure)*, (4) *Organizations, customs, activities, procedures, concepts (political and administrative, religious, and artistic concepts)*, (5) *Gestures and habits* ((cited in) Khaleghi Zavareh

The fact that language and culture have bound relationship, every culture is stick with social dimension, including people interaction. And in form of communication, they reflect the way their life is. Because both languages have different meanings, cultures, and subsystems, it is challenging to identify equivalent target language words to transfer cultural ideology and concepts into the target language. These various languages may have various categories, prototypes, and semantic properties.(Teilanyo, 2007)

To overcome this problem, the researcher would like to take further investigation into analyzing the

accuracy of the translated text by Google Translate and ChatGPT in translating Culture-Specific Items (CSI) on the description table of historical heritage at Museum Negeri Nusa Tenggara Barat. The result of translation from both machine translations will be compared according to their translation accuracy level.

B. Method

This study used a descriptive qualitative and quantitative approach. Qualitative research views the object as something dynamic, the result of the construction of thoughts and interpretations of the observed phenomena (Creswell, 2012 in Pudjiati et al., 2022). Descriptive qualitative used to describe a natural where this research investigates forms, activities, characteristics, changes, relationship, similarities and differences with another phenomenon (Sukmadinata, 2009). Quantitative results according to (George, 2008) was in form of measurement and expressed in numbers that can be analyze. The reason why researcher uses both research approach because qualitative method is uses to define more detail about the error and define

the quality category in translation product by two machine translation while quantitative method is uses to represent the number of errors made by the machine translation.

The data of this research is the transcription generated by Google Translate and ChatGPT specifically to the Culture-specific items. The researcher will translate the description table of historical heritage taken in Museum Negeri Nusa Tenggara Barat using Google Translate and ChatGPT. The text will translate into English using Google Translate and ChatGPT. Next, the output texts of Google Translate and ChatGPT was sorted to the word, phrase, or paragraph consisting CSI then analyzed based on the result of their translation accuracy.

Among the 57 of bilingual description texts, researcher choose 10 text as a sample and the technic use to choose the text is using purposive sampling because for some of the text, between the Indonesian version and the English version there's intra system shifting which would a bit risky and affect the results of accuracy using HTER analysis.

HTER is an evaluation metric designed to measure the quality of

machine translation by considering the level of changes required by human translators to improve machine translation. This metric is based on the idea that a good machine translation should require minimal human intervention in the form of editing or correction. Snover (2006) defined HTER as the minimum number of edits needed to change a hypothesis so that it exactly match one of the references, normalized by average length of the reference.

HTER tries to minimize the number of edits between references and hypotheses. The equation for the HTER score, where SUB (substitution), INS (insertion), DEL (deletion), and SHIFT (shift) are the number of substitutions, insertion, deletions, and shifts, separately, N (umber) is the average word count of the reference. In HTER process, there are four edits that may include: In the HTER process, there are four edits that may include: INS = Insertion, DEL = Deletion, SHIFT = Shift (In this study, SHIFT is used for Shift for convenience), and SUB = Substitution. The calculation is as follows:

$$\text{HTER} = \frac{\text{SUB} + \text{INS} + \text{DEL} + \text{SHIFT}}{N}$$

The procedures to analyze the data. Using data triangulation and methodological triangulation technique, researcher compared multiple data and uses more than 1 method to analysis the data. Triangulation is a technique that uses a dual-method approach to obtain fully accurate data. It included external elements other than the data itself for the aim of verification or comparison against the data and working as a quality control instrument.

a. Data Triangulation

As what mentioned above, there will be multiple data will be analyze, they are original text in Bahasa Indonesia, human translation in English, translated text from ChatGPT and translated text from Google Translate. The English version consider to be accurate with the consideration that the text is published at Museum also it is a public place whose visitors is not coming from local only but also from abroad. In shorts, the English version is verified.

b. Methodology of Triangulation.

In addition to utilizing qualitative measurement methods, such as calculating HTER scores to assess machine translation accuracy, researcher also employ qualitative

methods. Through qualitative methods, researcher identify the criteria of machine translation results, whether it is categorize as poor, average, good or excellent etc. according to Translation Quality using Translation Quality Index (TQI) by Schiaffino and Zearo (2005). Qualitative methods also allow researcher to provide description of errors made by the machine translation.

Analyzing the data from Chat-GPT start from assessing the text that consist of Culture-specific item into the table analysis. The table analysis consist of (1) original text (Indo. Ver.), (2) machine translation (ChatGPT), (3) human translation (Eng. Ver.), (4) wrong word/phrase, (5) types of error (ins, del, sub, and shift), and (6) error description. Then moved to analyzed the translation product from GT. Next procedure is measuring the error presentation. To find the error presentation for each machine translation GT and ChatGPT, the calculation follows:

$$\text{Error Rate} = \frac{\text{number of errors/edits(Ins + Del + Sub + Shift)}}{\text{total words in sample}}$$

From the error presentation, we can find the accuracy rate with 100 - error rate.

To define the quality of translation, researcher use TQI or (Translation Quality Index) adapted from Schiaffino and Zearo.(2005). In the TQI analysis, the accuracy level obtained from the HTER calculation is interpreted into a descriptive (non-numeric) value. Assessment criteria range from *Negative* (0), *Poor* (1-49), *Low* (50-59), *Improvable* (60-69), *Average* (70-79), *Good* (80-89), up to *Excellent* (90-100).

Table 2.1 TQI Score Criteria By Schiaffino And Zearo 2005

TQI Score	Criteria Range
0	Negative
1-50	Poor
50-59	Low
60-69	Improvable
70-79	Average
80-89	Good
90-100	Excellent

Last step, Comparing the performance of two machine translation namely Google Translate (GT) and ChatGPT in terms of translation accuracy level based on Translation Quality Index (TQI) in translating Culture-specific items in the description table of

historical heritage at Museum Negeri Nusa Tenggara Barat.

C. Research Finding And Discussion

After doing an analysis in total 10 text samples, researcher found 54 items of Culture Specific Items on the bilingual description text of historical heritage at Museum Negeri Nusa Tenggara Barat and has been divided into 5 different classification. Those are:

After doing an analysis in total 10 text samples, researcher found 54 items of Culture Specific Items on the bilingual description text of historical heritage at Museum Negeri Nusa Tenggara Barat and has been divided into 5 different classification. Those are:

Table 3.1 List of CSI found in the sample text

NO	Classifications	Words
1.	Ecology (flora, fauna, winds, plains, hills)	<i>Menjangan, Pelepah, Enau,</i>
2.	Material culture (food, clothes, houses and towns, transport)	<i>Jerat Mayung, Jerat Ayam, Pendiwal, Bedil, Peletan Burung Tekukur, Sarao, Boru Lemba Oi, Ponda Oi Nono, Opeq, Batik, Pihhi Api,</i>

3.	Social culture (work and leisure)	<i>Pande Besi,</i>
4.	Organizations, customs, activities, procedures, concepts, (political and administrative, religious, artistic)	<i>Tembang, Surat Perjanjian Wilayah Sultan Bima Dengan Belanda, Agama Islam Waktu Telu,</i>
5.	Gesture and habits	

Based on the table above, researcher found among 54 items of CSI which the most dominant classification is **Material Culture (food, clothes, houses and towns, transport)** with 48 items, followed by **Organizations, customs, activities, procedures, concepts, (political and administrative, religious, artistic)** with 3 items, then **Ecology (flora, fauna, winds, plains, hills)** with 2 items and the last one is **social culture (work and leisure)** with only 1 item.

Table 3.2 Error Presentation of MT System

Error categories	Presentation (%)	
	Chat-GPT	GT
Insertion	0.83	0.27
Deletion	0.54	1.09
Substitution	4.93	5.75
Shift	0.27	0.27
Total	0.06	0.07

With the 10 texts of sample consist of 365 words, results seems to

show that Chat-GPT doing less in error category than Google Translate (GT) but not that significance. The presentation of error made by Chat-GPT is 0.06% from total 10 texts samples and 0.07%.

Table 3.3 Error Frequency made by MT

Error category	Frequency	
	Chat-GPT	GT
Insertion	3	1
Deletion	2	4
Substitution	18	21
Shift	1	1
Total	24	27

Figure 3.1 Error Presentation made by Chat-GPT

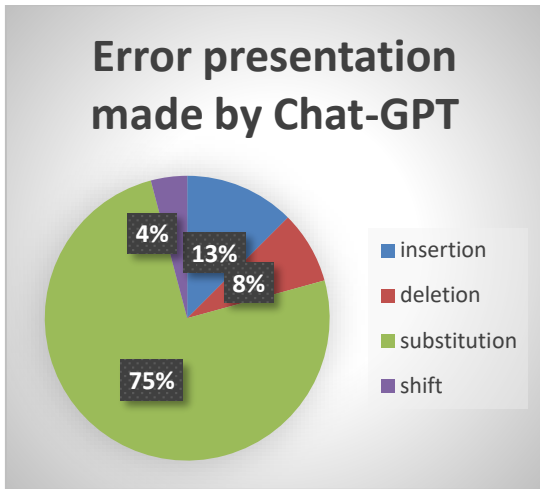


Figure 3.2 Error Presentation made by GT

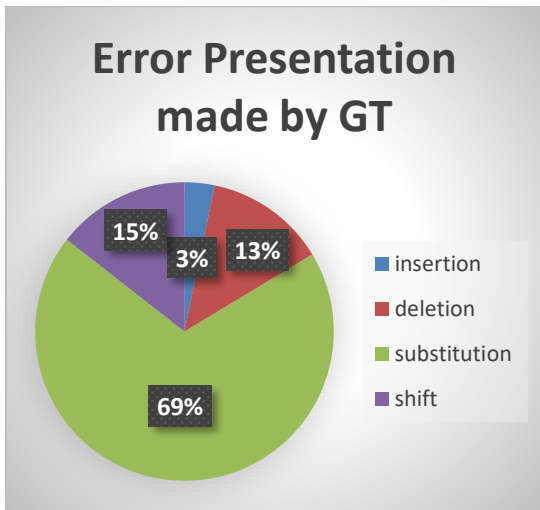
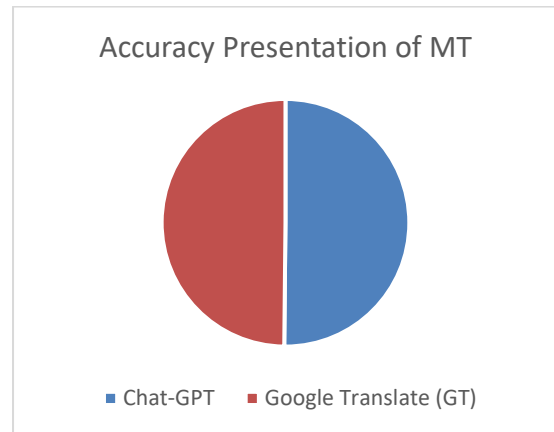


Table 3.4 Accuracy Presentation of MT

MT System	Accurate presentation
Chat-GPT	99.94%
GT	99.93%

Figure 3.3 Accuracy Presentation of MT



According to the data above, the accuracy level between Chat-GPT and Google Translate is almost in the same level that is 99.94% and 99.93%. Between the two MT also made most error in substitution category. To divine the quality of machine translation, researcher use Translation Quality Index (TQI) according to the result of machine translation Google Translate & ChatGPT. Based on the result of accuracy level, the quality of translation from GT and ChatGPT can be interpreted into Transalction Quality using Translation Quality Index (TQI) by Schiaffino and Zearo. In the TQI analysis, the accuracy level obtained from the HTER calculation is interpreted into a descriptive (non-numeric) value. Assessment criteria range of TQI start from *Negative* (0), *Poor* (1-49), *Low* (50-59), *Improvable* (60-69), *Average* (70-79), *Good* (80-

89), up to *Excellent* (90-100). Because the level of accuracy of Chat-GPT and Google translate are in in the same range which Chat-GPT score is 99.94% and Google Translate is 99.93% we can conclude that both of them are classified as **excellent** with the range 90-100.

Error Categories Identified in the Analysis

Human-mediated Translation Edit Rate or HTER is an evaluation metric designed to measure the quality of machine translation by considering the level of changes required by human translators to improve machine-generated translation. The metric is based on the idea that a good translation should require minimal human intervention in the form of editing or correction. There are four category of error identified in analysis. Those are:

1. Insertion

Insertion errors occur when the machine translation missed the words or phrases in the source text or the text reference. Thus, the human translator need to insert or adding the missing word on the target language. Look at the following example:

Text 2 (Chat-GPT)

ST: jerat ayam

TT: (..) chicken snare

HT: trap for wild chicken

Specifically, the word “chicken” referred to the “wild” one not the regular chicken. Chat-GPT translate it to “chicken” only, but to match the meaning in the source text, we should insert the word “wild”, for the complete meaning “wild chicken”. In this case, the text talk about the traditional trap to haunted wild animal because this tools used in the ancient times by the ancestors.

2. Deletion

Deletion errors happened when the machine translation omits words, phrases, or elements that are not present in the reference translator. Deletion errors also occur when there is an extra words, repetition, or any words or phrases that has no relation the reference text. Look at the following example:

Text 2 (GT)

ST: Berburu binatang seperti burung, *menjangan (rusa)*, ayam, dan lain sebagainya,...

TT: Hunting animals such as birds, *deer (deer)*, chickens, etc., as well as utilizing...

HT: Animal hunting such as bird, *deer*, wild chicken, and others as...

The text translated by Google Translate Word repetition occurs when translating the word "menjangan". In order to make the sentence precise and effective, the repetition of the word "deer" must be removed.

3. Substitution

Substitution error occur when the machine translation replace words, phrases or elements in the reference translation with incorrect ones. Pay attention to the following example:

Text 2 (Chat-GPT)

SL: Peletan Burung Tekukur

TL: dove caller

HT: peletan, a trap for a bird

Dove caller is a hand-made wooden call that is perfect for hunting mourning doves. It produces a realistic and authentic sound that will attract doves with ease.. However, in this context, the word "peletan" interpreted as a trap for turtle doves. Peletan is made from wood shape like letter L which then smeared with sap and then placed it on a tree. As for luring the turtledove to come closer, another turtledove is placed in a cage under the trap.

4. Shift

Shift errors involve a change in in the order or positioning of words, phrases, or elements in the machine translation compared to their position in the reference translation. Example:

Text 7 (GT)

SL: *Surat Perjanjian Wilayah Sultan Bima dengan Belanda*

TL: Sultan Bima Territory Agreement Letter with the Netherlands

HT: A Territory Agreement Letter between Sultan Bima and The Dutch
The word "Sultan Bima" was told after the territory agreement letter in the human translation but in GT it is in the first sentence. Thus, "sultan bima" identified as shift error.

E. CONCLUSION

After analyzed the data, researcher came to the conclusion. Based on the data analysis of culture-specific item in the description table of historical heritage at Museum Negeri Nusa Tenggara Barat, 54 items were found. And after long analysis of the accuracy of MT namely Chat-GPT and GT using the HTER theory by Snover, researcher found that between Chat-GPT and GT almost play accurate with the presentation 99.94% and 99.93% while the error rate for Chat-GPT is 0.06% while GT is 0.07%.

Furthermore, the quality index of the MT's are in the same range, which is **EXCELLENT**.

REFERENCES

- Aixela, J. F. (1996). *Culture-specific Items in Translation*.
- Çimke, S., & Yıldırım Gürkan, D. (2021). Determination of interest in vitamin use during COVID-19 pandemic using Google Trends data: Infodemiology study. *Nutrition*, 85. <https://doi.org/10.1016/j.nut.2020.111138>
- Creswell, J. W. (2012). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE Publications.
- George, M. W. (2008). The elements of library research: What every student needs to know. In *The Elements of Library Research: What Every Student Needs to Know*. <https://doi.org/10.5860/choice.46-4758>
- Jiao, W., Wang, W., Huang, J., Wang, X., & Tu, Z. (2023). *Is ChatGPT A Good Translator? Yes With GPT-4 As The Engine*. <http://arxiv.org/abs/2301.08745>
- Khaleghi Zavareh, N. (2021). Translation of Culture-Specific Items from English into Persian: A Case Study of The Secret Garden. *International Journal of Language and Translation Research Summer*, 2021(3), 15–38. https://doi.org/10.12906/978389966737_002
- Peng, K., Ding, L., Zhong, Q., Shen, L., Liu, X., Zhang, M., Ouyang, Y., & Tao, D. (2023). *Towards Making the Most of ChatGPT for Machine Translation*. <http://arxiv.org/abs/2303.13780>
- Pudjiati, D., Lustyantje, N., Iskandar, I., & Fitria, T. N. (2022). Post-Editing of Machine Translation: Creating a Better Translation of Cultural Specific Terms. *Language Circle: Journal of Language and Literature*, 17(1). <https://journal.unnes.ac.id/nju/index.php/LC/article/view/38474>
- Sipayung, K. T. (2021). Translation Shifts on Reference by Machine Translation in Descriptive Text. *Indonesian Journal of EFL and Linguistics*, 6(1). <https://doi.org/10.21462/ijefl.v6i1.362>
- Snover, M., Dorr, B., Schwartz, R., Micciulla, L., & Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. *AMTA 2006 - Proceedings of the 7th Conference of the Association for Machine Translation of the Americas: Visions for the Future of Machine Translation*.
- Sukmadinata. (2009). *Metode Penelitian Pendidikan* (5th ed.). Remaja Posda Karya.
- Sumiati, Baharuddin, & Saputra, A. (2022). The Analysis of Google Translate Accuracy in Translating Procedural and Narrative Text. *Journal of English Education Forum (JEEF)*, 2(1).
- Teilanyo, D. I. (2007). Figurative language in translation: A study of J.P. Clark's the ozidi saga. *Meta*,

52(2).

<https://doi.org/10.7202/016073ar>

Yamada, M. (2019). The impact of Google Neural Machine Translation on Post-editing by student translators. In *The Journal of Specialised Translation Issue*.