FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

THE ACCURACY OF CHATGPT IN TRANSLATING COLLOQUIAL TERMS IN *DAWAN* LANGUAGE

Iis Aprianti[1], Yoni Rahayu[2], Soraya Grabiella Dinamika[3], Muhammad Muslim Nasution[4]
[1]Universitas Timor, Kefamenanu, Indonesia
[2,3,4]Universitas Medan Area, Medan, Indonesia

Iisaprianti@unimor.ac.id

ABSTRACT

ChatGPT, a multilingual translation application has made intercultural conversation simpler including local language. However, ChatGPT's translation accuracy should be reviewed due to limits in recognizing domain-specific terms and cultural context. The aim of this study is to find out the translation accuracy of ChatGPT in translating colloquial terms in *Dawan* language. This study used qualitative method. The participants of this study were from the sixth semester English department students of University of Timor. The data source of this study was the translation result in *Dawan* into Indonesian language. While, the data of this study are ChatGPT's translation result obtained from students' word list from Dawan into Indonesian language. The result of the analysis showed showed that ChatGPT performs relatively well in translating basic vocabulary and common expressions, with 60% of the translations deemed accurate. However, the tool demonstrated significant limitations in handling culturally embedded and context-dependent terms.

Keywords: Chat GPT, Dawan Language, Language Translation, Translation Accuracy,

INTRODUCTION

In recent years, technology and translation go hand in hand. Translation has been revolutionized by the rise of artificial intelligence (AI). The advancement of AI models in machine translation has expedited translation processes and aided millions of individuals in comprehending texts in a variety of languages. Among these AI models, ChatGPT has not only drawn considerable interest for its ability to generate human-like text and perform translations but has also emerged as a potential solution for language translation.

ChatGPT is an artificial intelligence-based natural language model trained on extensive datasets and applied across various tasks, such as serving as a virtual assistant to help users work more efficiently, including in text translation (Kasneci et al., 2023; Mendoza et al., 2022). It can assist in generating content for various types of documents and provide feedback on grammar, style, and textual coherence (Sutanto et al., 2024). Interestingly, ChatGPT, as a machine translation tool, offers more than conventional translation; by enhancing prompts, it can be tailored to yield outputs aligned with professional translation standards (Aeni et al., 2024).

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

According to the author's review, multiple studies have explored ChatGPT and translation, such as analyses of its translation accuracy in scientific texts (Aeni et al., 2024), morphological and syntactic translation errors (Al-Salman & Haider, 2024), the use of ChatGPT in translation learning (Nurfaiza, 2024), and comparisons of ChatGPT's performance with other translation services (Hidayati & Nihayah, 2024). However, these studies have not yet addressed the acceptability and readability of ChatGPT translations in local languages. While ChatGPT performs effectively for instructional materials and achieves high translation quality in standardized contexts, it faces greater challenges in translating complex texts—particularly those in underrepresented local languages such as Dawan.

The Dawan language, primarily spoken in East Nusa Tenggara, Indonesia, is used by over one million people, including the Dawan tribe and other ethnic groups in the Timor Island region. As with many regional languages, Dawan struggles with digital representation and automated translation due to its unique grammar, sentence structure, and culturally embedded expressions, which often lack direct equivalents in other languages. These linguistic and cultural intricacies present significant hurdles for ChatGPT. Moreover, the limited availability of training data for low-resource languages like Dawan further contributes to potential translation inaccuracies. Although ChatGPT demonstrates reliable performance on simpler texts, its accuracy declines when dealing with more complex content that demands cultural and contextual understanding.

Given this phenomenon and the growing trend in AI-based translation, there is a compelling need for research to explore the accuracy of ChatGPT in translating colloquial terms in the Dawan language, particularly in terms of meaning conveyance, contextual appropriateness, and cultural sensitivity.

Translation is more than a linguistic act; it is also a cultural and cognitive process that involves interpreting meanings between languages and contexts (Munday, 2016). In this framework, colloquial language refers to informal, everyday expressions deeply embedded in the speaker's cultural and regional context (Nida, 2015; House, 2016). Unlike standard language, colloquial expressions often carry idiomatic or metaphorical meanings, making them particularly challenging to translate accurately.

Catford (2017) emphasized that the equivalence in translation, especially for colloquial or idiomatic expressions, is seldom direct. Instead, it requires contextual awareness, cultural familiarity, and semantic flexibility. Venuti (2017) argued that the translator must balance between domestication—making terms familiar to the target audience—and foreignization—preserving the unique features of the source language.

The growing use of machine translation (MT), including tools like Google Translate and DeepL, has shown promising results for formal and standardized texts (Wu et al., 2016). However, colloquial expressions remain a major barrier for most MT systems due to their cultural specificity and non-literal meanings (Toral & Way, 2018). Garciarena et al. (2020) noted that MT tools typically rely on parallel corpora, which are scarce for low-resource languages and informal registers.

Recent developments in neural machine translation (NMT) have improved MT accuracy. Bahdanau et al. (2015) introduced attention-based models that allow machines

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

to weigh parts of the input sequence differently, enhancing the interpretation of idiomatic language. Yet, as Koehn and Knowles (2017) observed, even advanced NMT systems struggle to retain the nuance of colloquial terms without domain-specific training.

OpenAI's ChatGPT, based on the GPT-3 and GPT-4 architecture, represents a leap in natural language understanding and generation (Brown et al., 2020; OpenAI, 2023). Unlike earlier MT systems, ChatGPT can respond contextually, learn from prompts, and simulate human-like conversations. This capacity makes it a candidate for more nuanced translations, including colloquial speech (Kasneci et al., 2023).

However, despite its fluency, several studies have noted limitations in ChatGPT's translation performance. Tang et al. (2023) pointed out that while GPT models can generate plausible translations, they sometimes prioritize grammatical fluency over semantic accuracy. Mendoza et al. (2022) further found that GPT-based models occasionally "hallucinate" translations that are contextually inappropriate, especially for languages with limited digital presence.

Low-resource languages, such as Dawan, often lack large annotated corpora, making them difficult for both traditional MT and AI models to process accurately (Anastasopoulos & Neubig, 2019). According to Faten (2021), the cultural depth and oral nature of many indigenous languages contribute to translation complexity. Sutanto et al. (2024) also emphasized that the translation of local languages requires cultural intelligence and familiarity with linguistic variations at the regional level.

Moreover, Ebrahimi et al. (2018) argued that without sufficient exposure, AI models tend to normalize low-resource languages into dominant linguistic patterns, leading to semantic loss or misinterpretation. For colloquial terms specifically, lack of training data limits ChatGPT's ability to understand the pragmatic and cultural implications embedded in those expressions.

Dawan, spoken primarily in Timor Island, is rich in metaphor, idiomatic phrases, and cultural references (Schapper, 2017). Colloquialisms in Dawan often rely on metaphors drawn from nature, kinship, or traditional customs. As such, their meanings cannot be easily deduced without cultural immersion (Lole, 2020). For instance, a simple expression like “Naek ba tasa” (literally “go to the plate”) may imply readiness to face a challenge—something a machine might misinterpret literally.

Studies on indigenous language preservation have highlighted the urgency of digitizing and documenting such expressions before they are lost (Grenoble & Whaley, 2016; Lewis & Simons, 2021). The integration of these languages into AI systems like ChatGPT could play a role in digital revitalization, but only if done carefully and ethically.

Accuracy in translation is often measured through three main criteria: adequacy, fluency, and fidelity to the source meaning (Popović, 2015). Tools like BLEU (Bilingual Evaluation Understudy) and human evaluators are used to assess these aspects. However, as Graham et al. (2017) argue, standard metrics do not always account for cultural and contextual nuances—especially in colloquial and low-resource language settings.

Manual evaluations, especially by native speakers and bilingual experts, remain the gold standard in assessing the translation of culturally embedded expressions. For

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

languages like Dawan, community participation is essential to ensure that translations are not only accurate but respectful and contextually grounded (Bird, 2020).

In summary, the existing literature underscores several critical points. First, colloquial expressions represent one of the most difficult areas in translation due to their cultural and idiomatic nature. Second, while ChatGPT has advanced capabilities in natural language processing, its effectiveness for low-resource and colloquial language translation remains uncertain without tailored training and contextual knowledge. Third, Dawan language poses specific challenges due to its rich oral traditions and metaphorical expressions.

Therefore, investigating ChatGPT's ability to translate Dawan colloquialisms not only contributes to the field of AI and linguistics but also supports broader efforts in indigenous language preservation.

METHODS

This study conducted Qualitative method. The results of this study are not formed in general conclusions or generalization but in a description. Thus, they are focused and detailed. The subject was the sixth semester students of academic year 2025 in English department of University of Timor. The data source of this study was the translation result in Dawan into Indonesian language. While, the data of this study are ChatGPT's translation result obtained from students' word list from Dawan into Indonesian language. Then, the translation data will be analyzed using content analysis with Nababan's translation quality assessment model which consists of accuracy, acceptability and readability. In the analysis stage, the translation data will be grouped and coded according to the theme, then analyzed and given an interpretation and ended with a conclusion.

RESULTS

This study analyzed the translation of 50 colloquial terms in the Dawan language into Indonesian using ChatGPT. The evaluation was guided by Nababan's translation quality model, which assesses translations across three core dimensions: accuracy, acceptability, and readability (Nababan & Nuraeni, 2012). These dimensions collectively determine how well a translation conveys meaning, conforms to target language norms, and remains easily understandable to readers.

Translation Accuracy

The accuracy scores were categorized into three levels—accurate**,** less accurate, and inaccurate—based on the degree of semantic equivalence between the source (Dawan) and target (Indonesian) terms:

- 30 terms (60%) were rated as accurate.
- 12 terms (24%) were rated as less accurate.
- 8 terms (16%) were rated as inaccurate.

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

Accurate Translations

The highest accuracy was observed in terms where there is a direct semantic and syntactic equivalent in Indonesian. For example:

- *Ha'u* → *Saya* (I): This translation is both contextually appropriate and linguistically accurate, as it reflects the standard first-person pronoun in both languages.
- *Mone* → *Anak laki-laki* (boy): This translation precisely captures the intended meaning.

Such terms are lexically simple, and their cultural connotations are minimal or universal enough to translate easily.

Less Accurate Translations

Translations were considered less accurate when the core meaning was preserved, but additional or inappropriate semantic elements were introduced. A good example is:

> *Inan* → *Ibu kandung* (biological mother): While technically correct, this translation is over-specified. In everyday use, *inan* generally means ibu (mother) in Dawan culture. The addition of *kandung* could mislead readers to think the context distinguishes biological from non-biological motherhood, which was not intended.

Such cases reflect the model's literal bias—a tendency to generate formal definitions or high-frequency mappings, rather than selecting the most pragmatically appropriate equivalents.

Inaccurate Translations

Some terms were significantly mistranslated, resulting in a distortion of cultural meaning. Notably:

> *Tobe* → *Kepala desa* (village head): This translation misses the ceremonial and tribal leadership function embedded in the term tobe, which refers to a traditional community leader in Dawan society. The bureaucratic notion of kepala desa fails to capture these indigenous values.

These inaccuracies underscore ChatGPT's inability to recognize sociocultural contexts, especially where terms involve customary law, traditional hierarchies, or oral metaphors.

Acceptability and Readability

Translation outputs were also evaluated for linguistic acceptability and ease of reading in Indonesian. The results showed:

- 80% of the translations were acceptable and highly readable.
- The remaining 20% included translations that, while grammatically sound, were stylistically awkward or unnatural in colloquial Indonesian.

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

ChatGPT's tendency to use standard and neutral syntax contributed positively to readability. For example:

Ha'u besi → *Saya pergi* (I go): The structure is clean, straightforward, and conforms to normative Indonesian grammar.

However, problems arose when dealing with compound or idiomatic expressions. Consider:

Ana-ana mone → *Anak-anak laki-laki*: While grammatically correct, this translation lacks pragmatic nuance. In Dawan, the phrase may imply "sons" or "young boys" depending on the social context. More culturally fitting translations might include *para anak laki-laki* or *anak-anak lelaki*.

Such examples show that literal translation can hinder naturalness and cultural clarity, especially in socially embedded expressions.

Interpretation and Implications

These findings highlight both the capabilities and limitations of ChatGPT as a tool for translating low-resource and culturally rich languages. While the model performs well on standard vocabulary and syntax, it struggles with idiomatic, metaphorical, and cultural expressions—a trend that aligns with prior research (Al-Salman & Haider, 2024; Aeni et al., 2024).

Several factors likely contribute to these limitations:

1. Data scarcity: Dawan is a low-resource language with limited presence in publicly available corpora. As noted by Anastasopoulos & Neubig (2019), MT systems perform poorly when training data is scarce or domain-specific.
2. Lack of domain adaptation: ChatGPT is not fine-tuned on indigenous languages or culturally grounded idioms. This often leads to overgeneralization or approximation using dominant language structures.
3. Contextual misalignment: The AI's understanding is based on probabilistic pattern recognition. Without cultural intelligence or regional linguistic insights, it cannot fully infer implicit meanings behind colloquial terms (Ebrahimi et al., 2018).

This situation suggests the need for a human-in-the-loop approach, especially in tasks requiring cultural nuance. Native speakers or linguists must be involved in post-editing or fine-tuning AI output, particularly when translating for educational, ceremonial, or historical purposes.

Furthermore, these results reinforce the urgency of digitally documenting indigenous languages (Grenoble & Whaley, 2016) and exploring ethical AI integration that does not erase the cultural uniqueness of minority communities. If well-trained, AI tools like ChatGPT could play a pivotal role in language preservation and revitalization.

CONCLUSION

This study examined the accuracy of ChatGPT in translating colloquial terms from the Dawan language into Indonesian. The analysis showed that ChatGPT performs relatively well in translating basic vocabulary and common expressions, with 60% of the translations

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN: 2723-4126**
***p*-ISSN: 2776-8880**

deemed accurate. However, the tool demonstrated significant limitations in handling culturally embedded and context-dependent terms.

The main challenges stemmed from ChatGPT's lack of sensitivity to cultural nuances and its reliance on general linguistic patterns. Terms such as *tobe*, which hold deep traditional meaning in the Dawan community, were often mistranslated due to the absence of direct equivalents in Indonesian and insufficient contextual understanding by the AI model.

Despite these challenges, ChatGPT offers notable potential as a preliminary translation tool, especially for lesser-known languages. Its ability to provide readable and grammatically sound outputs makes it useful for first drafts or informal translation tasks. Nevertheless, for translations involving indigenous or culturally rich languages, human intervention remains essential to ensure the preservation of meaning and cultural integrity.

In the future, integrating more localized data into AI training sets, enhancing prompt engineering strategies, and adopting hybrid human-AI translation workflows may help improve the accuracy and cultural sensitivity of machine translation tools like ChatGPT.

In sum, this study demonstrates that:

a. ChatGPT is reasonably effective at translating basic Dawan vocabulary into Indonesian with high readability and structural fluency.
b. However, accuracy suffers when the input includes culturally loaded or context-sensitive terms.
c. Colloquial and traditional expressions are often mistranslated or diluted, leading to loss of cultural integrity.

Thus, ChatGPT can serve as a preliminary translation aid, but must be supplemented by human expertise, particularly for underrepresented and culturally rich languages like Dawan. Future work should explore the development of localized language models trained on regional corpora, and the inclusion of indigenous communities in AI tool design and evaluation.

ACKNOWLEDGMENT

The researcher would like to express their sincere gratitude to the sixth-semester students of the English Department at the University of Timor for their valuable contributions in providing the Dawan word lists used in this research. Special thanks are also extended to the lecturers and faculty members who supported the data collection process. This research would not have been possible without their active participation and collaboration.

REFERENCES

Aeni, R., Putera, L. J., & Melani, B. Z. (2024). The accuracy of ChatGPT in translating linguistics text in scientific journals. *Didaktik: Jurnal Ilmiah PGSD STKIP Subang, 10*(1), 59–68.

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

Al-Salman, S., & Haider, A. S. (2024). Assessing the accuracy of MT and AI tools in translating humanities or social sciences Arabic research titles into English: Evidence from Google Translate, Gemini, and ChatGPT. *International Journal of Data and Network Science*. https://api.semanticscholar.org/CorpusID:271237784

Anastasopoulos, A., & Neubig, G. (2019). Pushing the limits of low-resource morphological inflection. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 980–986.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *International Conference on Learning Representations (ICLR)*.

Bird, S. (2020). Decolonising speech and language technology. *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, 3504–3519.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems, 33*, 1877–1901.

Ebrahimi, J., Rao, A., Lowd, D., & Dou, D. (2018). HotFlip: White-box adversarial examples for text classification. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics*, 31–36.

Faten, S. (2021). Challenges in translating indigenous languages: A cultural perspective. *Journal of Translation Studies, 18*(2), 122–134.

Garciarena, U., Gonzalez-Agirre, A., & Rigau, G. (2020). Evaluating multiword expression identification in machine translation. *Computational Linguistics, 46*(1), 1–36.

Graham, Y., Baldwin, T., Moffat, A., & Zobel, J. (2017). Can machine translation systems be evaluated without human references? *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 3769–3779.

Grenoble, L. A., & Whaley, L. J. (2016). *Saving languages: An introduction to language revitalization*. Cambridge University Press.

House, J. (2016). *Translation as communication across languages and cultures*. Routledge.

Kasneci, E., Sessler, K., Yu, Y., Bannert, M., Dementieva, D., Fischer, F., & Schmidt, A. (2023). ChatGPT for good? Opportunities and challenges of large language models in education. *Learning and Individual Differences, 103*, 102274.

Koehn, P., & Knowles, R. (2017). Six challenges for neural machine translation. *Proceedings of the First Workshop on Neural Machine Translation*, 28–39.

Lewis, M. P., & Simons, G. F. (2021). Assessing endangerment: Expanding Fishman's GIDS. *Revue Roumaine de Linguistique, 66*(1), 55–70.

Lole, A. J. (2020). Cultural metaphors in Dawan language. *Timor Linguistics Journal, 4*(1), 23–35.

FOSTER JELT
Journal of English Language Teaching

Vol.6, No. 2, April 2025
***e*-ISSN**: **2723-4126**
***p*-ISSN**: **2776-8880**

Mendoza, M., et al. (2022). Enzymatic transfer of acetate on histones from lysine reservoir sites to lysine activating sites. *Science Advances, 8*(3), eabj5688.

Munday, J. (2016). *Introducing translation studies: Theories and applications* (4th ed.). Routledge.

Nida, E. A. (2015). *Language, culture, and translating*. Shanghai Foreign Language Education Press.

Nurfaiza. (2024). Pengaruh penggunaan ChatGPT dalam pembelajaran terjemah Indonesia-Arab. *At-Turost: Journal of Islamic Studies*, 99–113.

OpenAI. (2023). *GPT-4 technical report*. https://openai.com/research/gpt-4

Popović, M. (2015). chrF: character n-gram F-score for automatic MT evaluation. *Proceedings of the Tenth Workshop on Statistical Machine Translation*, 392–395.

Schapper, A. (2017). *A grammar of Bunaq: A Papuan language of East Timor*. De Gruyter Mouton.

Sutanto, V. M., De Giacomo, G. G., Nakazawa, T., & Yamada, M. (2024). ChatGPT as a translation engine: A case study on Japanese-English. In *Proceedings of the 30th Annual Conference of the Association for Natural Language Processing* (pp. 2096–1201).

Tang, R., et al. (2023). Evaluating ChatGPT's translation capabilities: Benchmarks and breakdowns. *AI & Society*. https://doi.org/10.1007/s00146-023-01596-4

Toral, A., & Way, A. (2018). What level of quality can neural machine translation attain on literary text? *Translation Quality Assessment*, 78–91.

Venuti, L. (2017). *The translator's invisibility: A history of translation* (3rd ed.). Routledge.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., ... & Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.