

FROM WORDS TO ERRORS: ANALYZING LEXICAL CHALLENGES IN INSTAGRAM'S MACHINE TRANSLATION OF 'NYTIMES'

I Putu Riky Wiwekananda^{1*} & Ni Made Verayanti Utami²
^{1,2}English Study Program Faculty of Foreign Languages
Universitas Mahasaraswati Denpasar, Indonesia

Corresponding Author: Author's Name, Ni Made Verayanti Utami E-mail: verayanti.utami@unmas.ac.id

ARTICLE INFO	ABSTRACT
Received: 28-11-2024 Revised: 15-12-2024 Accepted: 25-12-2024 Published: 30-12-2024 Volume: 9 Issue: 1 DOI: https://doi.org/10.33019/lire.v9i2.386	In the current digital era, disseminating information via social media is very easy and quick, but this can also increase the spread of wrong and inaccurate information to the public caused by translation errors in automatic translations on social media. The main objectives of this research are 1. Identify the types of translation errors found in Instagram Machine Translation in Translating the Caption of the "NYTimes" Instagram Account. 2. search for the most dominant types of lexical errors made by Instagram Machine Translation in Translating Caption of "NYTimes" Instagram Account. The theory of error classification by Vilar et al. (2006) is applied to identify three main types of errors: unknown word, missing word, and incorrect word. The research method involves analyzing 100 translated news captions. There are three types of Lexical Error found in Instagram machine translation when translating captions from the NYTimes Instagram account (1) Incorrect Word (4 instances, 50%) (2) Omitted or Missing Word (2 instances, 25%) (3) Unknown Word (2 instances, 25%). The most common type of error identified in the translations is the use of incorrect words, accounting for half of all errors. This indicates that the machine translation often fails to select the appropriate words to convey the original meaning effectively.
KEYWORDS	
<i>Lexical errors, Automatic translation, Social media</i>	

1. INTRODUCTION

In the current digital era, social media plays a crucial role in spreading information, shaping narratives, and impacting public opinion. From breaking news to viral trends, social media platforms are crucial conduits for global information exchange. According to Monthly Active Users (MAU), alongside other media platforms such as Facebook, YouTube, and WhatsApp, Instagram holds more than 2 billion active users worldwide (Rizaty, 2022). However, this phenomenon presents both advantages and challenges. While it facilitates rapid information dissemination, it also amplifies the spread of misinformation and disinformation. The accessibility and user-generated nature of social media content makes it vulnerable to manipulation and exploitation. Consequently, critical thinking and digital literacy have become essential skills for navigating the vast sea of online information. Nevertheless, social media remains a pivotal force in shaping how information is shared, consumed, and interpreted in today's interconnected world.

36



Ensuring the accuracy and reliability of information circulated on social media through translated news captions is crucial. Catford (1974:20) states that translation is the substitute of textual material in one language by equivalent textual material in another language. Machine translation is a translation tool used to facilitate the translation process (Holmes in Munday: 2008). According to Vilar et al. (2006), lexical errors in translations can be categorized into three main types: unknown words, missing words, and incorrect words. The lexical errors can be found in many media, such as social media caption, including news caption. As emphasized by Sager (1989), there is no definitive benchmark for translation quality; rather, translations should be judged by their suitability for their intended purpose. Brown (as cited in Wijayanto, 2020) explains that although errors cannot correct themselves, they can be observed, analyzed, and classified through a process known as error analysis.

2. LITERATURE REVIEW

This study provides some related studies for idea development. The first study is a thesis by Rohmani (2021) entitled Common Lexical Errors Made by Machine Translation in Cultural Text. This study focused on 553 pages of the Arok Dedes story, using a simple random sampling technique to select samples. The results identified 9 out of 21 types of lexical errors, with calque being the most dominant. This study is similar to Rohmani's in its focus on identifying and categorizing lexical errors. However, while Rohmani analyzed a cultural text, this study examines news-related content translated by Instagram Machine Translation, with a particular focus on errors during a period of global news coverage. Additionally, both studies emphasize the identification of dominant error types, though the content and context differ.

The second study is a journal article by Nadilla and Lubis (2023), entitled Lexical Errors Made by Instagram Machine Translation in Translating the Account of "CNN Indonesia" News Article. This study depicted the types of errors made by Instagram Machine Translation, finding 4 missing errors, 10 incorrect words, and 8 unknown words. Similar to Nadilla and Lubis's research, this study also analyzes Instagram Machine Translation's performance, specifically focusing on lexical errors. Both studies use Vilar et al. (2006) as a theoretical framework, but they differ in the datasets analyzed: Nadilla and Lubis examined CNN Indonesia's content, while this study focuses on NYTimes captions during a specific period of intense news activity.

Another related study is a thesis by Utami et al. (2021), titled Lexical Analysis of Semantic Errors Found in the Translation of Joko Widodo's Instagram Account. This research concentrated on semantic errors in lexis made by Instagram Machine Translation in Joko Widodo's Instagram posts. While both this study and Utami et al.'s research investigate Instagram Machine Translation, Utami et al. focus on semantic errors, whereas this study delves into lexical errors. The difference lies in the scope of analysis: Utami et al. centered on semantic accuracy, while this study is concerned with the lexical precision of translations, particularly in news contexts.

The fourth study is a journal article by Alasta and Sujarwati (2021) titled Lexical Errors Produced by Google Translate in Translating "Putri Serindang Bulan" to English Language. This study analyzed the lexical errors produced by Google Translate using Vilar et al.'s (2006) error categories. This study shares a theoretical foundation with Alasta and Sujarwati, as both use Vilar et al.'s framework to analyze lexical errors. However, this study examines Instagram Machine



Translation rather than Google Translate, and it focuses on translating news content instead of a literary work.

The last study is by Prasetyo (2023), titled *An Analysis of the Error Translation in Movie Trailers by YouTube Auto-Translate*. This research discussed common errors made by YouTube Auto-Translate in translating movie trailers into Indonesian. Both this study and Prasetyo's work analyze machine translation errors, using Vilar et al.'s (2006) theory. However, while Prasetyo focused on audiovisual content from YouTube, this study examines Instagram Machine Translation's textual content, specifically in the context of news captions during a significant global event. Instagram's automatic translation feature can clearly be very helpful for people to interpret the captions displayed, but there are many discrepancies and word errors due to the poor quality of machine translation which causes the translation machine to not be able to clearly analyze what is being translated. Therefore, this research was conducted to identify several translation discrepancies such as unknown words, missing and incorrect words and to determine the most dominant translation errors that occurred in the NYTimes Instagram account caption as the main data.

Evaluating the accuracy of English to Indonesian translations using Instagram's translation tool offers valuable insights into the challenges and limitations of machine translation systems, particularly in handling intricate linguistic nuances and contextual subtleties inherent in news content.

Examining lexical errors produced by Instagram's translation machine in news captions sheds light on phonetic designs, social subtleties, and semantic complexities that contribute to interpretation errors. Such insights are essential for refining machine translation algorithms, improving cross-cultural communication, and ensuring the faithful dissemination of news in multilingual contexts. In this manner, investigating lexical errors in machine-translated news content on Instagram is crucial for addressing challenges related to misinformation, advancing linguistic accuracy, and fostering a more informed and interconnected global community.

3. METHODOLOGY

The research design of this study employed a quantitative methodology to elucidate the data. Quantitative information was gathered to investigate the phenomenon of Instagram Machine Translation output. The data of this study is primarily textual, comprising phrases, sentences, and paragraphs. The data source for this research was the captions from the "NYTimes" account on Instagram, which were translated by Instagram Machine Translation. The data was extracted from captions originally in English and translated into Indonesian using Instagram Machine Translation. The NYTimes Instagram account is chosen due to its wide reach and substantial follower base of over 18 million people, which ensures significant audience engagement and relevance. Furthermore, as the account regularly shares captions on major global events, its content offers valuable insights into machine translation performance in high-profile, contextually complex scenarios, and their impact on audience comprehension. There are some ways to gather the data by reading the caption of NYTIMES's Instagram account, translating the captions using Instagram machine translation. taking notes of the lexical error that machine translation produced from translating the caption.



The data collection took place from February 2024 to June 2024, a period notably marked by extensive news coverage of the conflict between Russia and Ukraine. This period was chosen due to the global significance of the issues, as they influenced public discourse and highlighted the role of key figures in shaping narratives. The analysis aimed to observe the direction of the news and its potential impact on the interpretation of events during this critical time. Following the collection of all data, the researcher examined the texts or captions translated by Instagram Machine Translation on the "NYTimes" Instagram account. The translation results were then compared to the source language by the researcher. The next step involved examining the lexical errors produced by Instagram Machine Translation, specifically focusing on the error categories outlined by Vilar et al. (2006), which include unknown words, missing words, and incorrect words.

4. RESULTS AND DISCUSSION

This study investigates the lexical errors produced by Instagram Machine Translation when translating captions from the "NYTimes" Instagram account from English to Indonesian. Using the framework proposed by Vilar et al. (2006), the analysis identified three main types of lexical errors: unknown words, missing words, and incorrect words. These errors highlight the limitations of machine translation in handling the nuanced language used in news captions, which often includes domain-specific terminology and concise expressions. Moreover, the findings underline the importance of understanding error patterns in social media content, as these platforms increasingly serve as primary sources of information for global audiences. With over 18 million followers, the "NYTimes" Instagram account reaches a diverse and extensive audience, making the accuracy and clarity of its translated captions crucial for effective communication. By focusing on this account, the study sheds light on how machine translation systems handle high-profile news content and the challenges of ensuring contextually appropriate translations for such a broad audience. With more than 50 posts published every week, the account generates a substantial volume of content, each caption translated into multiple languages by machine translation systems. This high posting frequency provides a rich dataset to analyze the prevalence and patterns of lexical errors, offering valuable insights into the system's performance in managing a large-scale translation workload.

4.1 Table 1. Types of Lexical Error

No	Types	Frequency	Percentage
	Unknown Word	2	25%
	Missing Word	2	25%
	Incorrect Word	4	50%
	Total	8	100%

Based on the data presented in the table, the analysis of Instagram machine translation when translating captions from the "NYTimes" Instagram account reveals several errors. Out of 100 data points analyzed, a total of 8 errors were identified, categorized into three types. Incorrect Word (4 instances, 50%): This type of error involves the machine translation substituting a word



with an incorrect one, which changes the intended meaning of the original text. Omitted or Missing Word (2 instances, 25%): These errors occur when the machine translation fails to include a word or words from the original text, leading to incomplete translations that may lack clarity or context. Unknown Word (2 instances, 25%): This type of error occurs when the machine translation encounters a word or phrase that it cannot accurately translate, resulting in a gap or inaccurate representation in the target language.

4.2 Lexical Errors

Lexical errors refer to inaccuracies in the translation of individual words or phrases, often resulting from a lack of understanding of their meanings or context. Vilar et al. (2006) categorize lexical errors in machine translation into various types, including missing words, untranslated terms, and incorrect word choices. These errors can compromise the overall clarity and effectiveness of the translated text.

4.2.1 Table 2. Data Classification of Unknown Word

Source Language (English)	Target Language (Indonesia)	Lexical Error Produced by IMT Based on Vilar et al (2006) Categories
Those who emerge successfully from the audition process must then endure boot camp .	Mereka yang berhasil keluar dari proses audisi kemudian harus bertahan di boot camp .	Unknown Word

Based on the provided data, there seems to be an issue with the translation of the term "boot camp" from English into Indonesian. The term "boot camp" refers to a rigorous training program characterized by intense, disciplined activities. It is commonly associated with military training or intense fitness programs where participants undergo challenging physical and mental exercises.

According to the data, Instagram's machine translation did not provide a suitable Indonesian equivalent for "boot camp." This failure in translation could potentially lead to confusion among Indonesian readers, particularly those who are unfamiliar with the English term and its specific connotations. To address this, it is important to find an accurate Indonesian translation that effectively conveys the concept of a rigorous and disciplined training program. In Indonesian, alternative translations for "boot camp" could include "program pelatihan intensif," "latihan ketat," or "latihan disiplin militer," preferred when addressing an audience less familiar with English or technical jargon, where a direct translation like "program pelatihan intensif" or "kamp pelatihan" can clarify its meaning.

4.2.2 Table 3. Data Classification of Missing Word

Source Language (English)	Target Language (Indonesia)	Lexical Error Produced by IMT Based on Vilar et al (2006) Categories
A sprawling network of opponents of Donald Trump	Sebuah jaringan luas lawan Donald Trump	Missing Word



The oldest Ukrainians whose towns have been bombarded	Ukraina tertua yang kota-kota telah dibom	Missing Word
--	--	--------------

In this case, the translation from the Source Language (SL) to the Target Language (TL) did not accurately convey the intended meaning due to the omission of the preposition of. The phrase "sebuah jaringan luas lawan Donald Trump" was translated as "a sprawling network against Donald Trump," which suggests that the network itself actively opposes Trump. This translation lacks the correct interpretation of of, which is essential for clarity and accuracy. The correct translation should include the preposition dari to properly convey possession or affiliation. Therefore, the sentence should read as "Sebuah jaringan luas dari lawan Donald Trump," which translates to "A sprawling network of opponents of Donald Trump." The inclusion of "dari" ensures grammatical precision and avoids ambiguity. Without "dari", the phrase wrongly emphasizes the network's adversarial action. In contrast, using "dari" clarifies that the network is composed of individuals or entities who oppose Trump, rather than describing the network itself as actively opposing him. This distinction is crucial for accurately conveying the intended meaning and ensuring the translation reflects the possession or affiliation of the network with respect to Donald Trump.

The word "orang" (people) from the translation of the phrase "The oldest Ukrainians whose towns have been bombarded" from the original language (SL) to the target language (TL). The phrase "orang Ukraina" (Ukrainians) was not fully translated into the target language, which affects the clarity and specificity of the sentence. Including the word "orang" before "Ukraina" (Ukrainians) is crucial because it explicitly identifies the people or individuals affected by the bombings. "Orang Ukraina" translates directly to "Ukrainians" in English, indicating the nationality and thereby providing a clearer understanding of the subject of the sentence. When "orang" is omitted from the translation, the sentence loses specificity regarding the individuals impacted by the bombings. This omission can lead to confusion or ambiguity for readers who rely on the translation to understand the precise subject matter. To ensure accuracy and clarity in translation, it is important to include all relevant terms and phrases that provide essential information about the subject being discussed. In this case, including "orang" before "Ukraina" in the Indonesian translation helps maintain the intended meaning and effectively communicates the specific group of people affected by the bombings.

4.2.3 Table 4. Data Classification of Incorrect Word

Source Language (English)	Target Language (Indonesia)	Lexical Error Produced by IMT Based on Vilar et al (2006) Categories
A new AI prep program, the largest of its kind in the United States	Sebuah program persiapan AI yang baru, yang terbesar dari seumpamanya di Amerika Serikat	Incorrect Word
Construction slowed, but then picked up .	Konstruksi melambat, tetapi kemudian diambil .	Incorrect Word



In this case, the data reveals an error in translation from the Source Language (SL) to the Target Language (TL). The phrase "its kind" in English refers to the types or categories of AI preparation programs in the USA. However, when translated into Indonesian as "seumpamanya," it implies making a comparison rather than categorizing types or categories. The phrase "seumpamanya" in Indonesian denotes a comparative context rather than specifying categories or types. This mistranslation alters the original meaning and context intended in the source text, potentially leading to confusion or misinterpretation for the reader. To accurately convey the intended meaning in Indonesian, it would be more appropriate to use terms like "jenis-jenis" or "kategori-kategori" to clearly indicate the types or categories of AI preparation programs being referred to. This adjustment ensures that the translation reflects the original context accurately and effectively communicates the intended information to the target audience. The revised translation using "jenis-jenis" or "kategori-kategori" is more accurate than the original "seumpamanya." The term "seumpamanya" in Indonesian is primarily used in comparative contexts, suggesting "such as" or "like," which doesn't fit the intended meaning of categorizing different types or categories. On the other hand, "jenis-jenis" or "kategori-kategori" directly convey the sense of types or categories, aligning better with the original meaning in English. This adjustment ensures that the translation correctly reflects the original context, which is about distinguishing between different types or categories of AI preparation programs, rather than making a comparison. As a result, using "jenis-jenis" or "kategori-kategori" avoids the potential for confusion and provides a clearer, more accurate translation that aligns with the source text's intended meaning.

The phrase "picked up" when translated into Indonesian as "diambil". This translation can cause confusion because it does not accurately convey the intended meaning within the context of the news. The phrase "picked up" in English typically means to increase or improve, especially in the context of progress or speed. In the context of construction progress, it suggests an increase in pace or activity. To clarify why the translation of "picked up" as "diambil" might not be ideal, it's essential to consider the context in which the phrase is used. In English, "picked up" can imply improvement or acceleration, especially when referring to progress, as in "construction picked up." This means that the pace or activity of the construction has increased or improved, often indicating growth in efficiency or speed. In this context, translating "picked up" as "diambil" would be confusing because diambil literally means "taken" or "picked up physically," which does not align with the intended meaning of progress or improvement. A more suitable translation in Indonesian would be "meningkat" (which means "increased" or "improved") or "berkembang" (which means "developed" or "progressed"). These words convey the idea of improvement in pace or activity, which fits better with the context of construction progress. For example, if the sentence in English reads, "The construction picked up pace last week," the appropriate Indonesian translation would be "Pembangunan meningkat minggu lalu" or "Pembangunan berkembang minggu lalu," both of which clearly communicate the intended meaning of an increase in speed or progress in the construction work.

5. CONCLUSION

Based on the classification of lexical errors in the automatic translation of news captions from English to Indonesian using Instagram's machine translation tool, several significant observations



can be made. The analysis, guided by Vilar et al.'s (2006) categorization framework, revealed recurring issues such as unknown words, missing words, and incorrect word choices.

The most common type of error identified in the translations is the use of incorrect words, accounting for half of all errors. This indicates that the machine translation often fails to select the appropriate words to convey the original meaning effectively. Omitted or missing words and encountering unknown words also contribute significantly to the errors, highlighting challenges in maintaining accuracy and completeness in the translated captions.

Linking these errors to their impact, the consequences of inaccurate translations are far-reaching. Incorrect word choices can lead to misinterpretations of news, particularly in contexts where precise language is essential. For example, in political or sensitive issues, improper word selection can cause confusion, shifting the tone or meaning of the message. Missing words can make the translation feel incomplete, while unknown words may prevent comprehension altogether. This undermines the ability to effectively communicate with a diverse, global audience.

To improve the accuracy of machine translations on social media platforms like Instagram, it is essential to develop algorithms that better understand linguistic nuances and cultural contexts. Addressing these challenges will not only improve the quality of information dissemination but also contribute to more effective cross-cultural communication and understanding.

In conclusion, while machine translation technology shows promise in facilitating global communication, ongoing refinement and adaptation are necessary to mitigate errors and ensure faithful transmission of information across languages. This analysis could be further validated through comparison with other studies on machine translation in social media, such as the works by Moorkens et al. (2018) and García et al. (2020), which have highlighted similar issues related to cultural nuances and context. Additionally, incorporating empirical validation through human evaluation or crowdsourcing would help reduce subjectivity, providing more reliable insights for improving machine translation technology.

REFERENCES

- Alasta, K., & Sujarwati, I. (2021). Lexical Errors Produced by Google Translate in Translating “Putri Serindang Bulan” to English Language. *Jadila: Journal of Development and Innovation in Language and Literature Education*, 2(2), 200-211.
- Alverina, C. (2021). The Analysis Of Lexical Errors Produced By Fikinakii’s Instagram Account Translated By Instagram Translation Machine. <https://repositori.buddhidharma.ac.id/1774/>
- Catford, J. C. 1974. *A Linguistic Theory of Translation*. Oxford: Oxford University Press
- Hartono, H. (2017). Analisis Kesilapan Terjemahan Bahasa Inggris Ke Dalam Bahasa Indonesia. *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, 2(1), 40–46. <https://doi.org/10.22219/kembara.v2i1.4042>
- Nadilla, R., & Lubis, S. (2023). Lexical errors made by instagram machine translation in translating the account of “CNN Indonesia” news article. *LingPoet: Journal of Linguistics and Literary Research*, 4(1), 30-45.
- New York Times (NYTIMES) Instagram Account <https://www.instagram.com/nytimes/?hl=id>



- Popović, M. (2018). Error classification and analysis for machine translation quality assessment. *Translation quality assessment: From principles to practice*, 129-158.
- Prasetio, N., & Wahyuningsih, N. S. (2023). An Analysis of the Error Translation in Movie Trailers by Youtube Auto-Translate. *Eligible: Journal of Social Sciences*, 2(2), 264-278.
- Pratama, M. R. A., & Dewi, M. K. (2022). Students' translation Errors In Translating News Texts Based On Vilar Et Al.(2006). *Philosophica: Jurnal Bahasa, Sastra, dan Budaya*, 5(1), 31-36.
- Rizaty, M. (2022). Pengguna Instagram Indonesia Terbesar Keempat Di Dunia. DataIndonesia.Id. <https://dataindonesia.id/digital/detail/pengguna-instagramindonesia-terbesar-keempat-di-dunia>
- Sholikhah, N. F. M. A., & Indah, R. N. (2021). Common lexical errors made by machine translation on cultural text. *Edulingua: Jurnal Linguistiks Terapan dan Pendidikan Bahasa Inggris*, 8(1), 39-50. <http://repository.uin-malang.ac.id/9330/>
- Utami, N. M. V., Jayantini, I. G. A. S. R., & Pratiwi, Y. (2021, April). Lexical analysis of semantic errors found in the translation of Joko Widodo's Instagram account. In *English Language and Literature International Conference (ELLiC) Proceedings* (Vol. 4, pp. 291-297).
- Vilar, D., Xu, J., D'Haro, L., Haro, D., & Ney, H. (2006). *Error Analysis and Statistical Machine Translation Output*. Universidad Politecnica de Madrid.
- Wijayanto, P. W. (2020). *The Translation Analysis of English Imperatives into Javanese by Google Translate*. 021, 210–217. <https://doi.org/10.5220/0009018902100217>.
- Yunidar, Rahmatika, L., & Suharja, A. A. (2024). The Integration of Indonesian Culinary Content in Indonesian for Foreigners' Textbooks "Sahabatku Indonesia". *KEMBARA: Jurnal Keilmuan Bahasa, Sastra, Dan Pengajarannya*, 10(1), 87–99. <https://doi.org/10.22219/kembara.v10i1.28213>

