UNVEILING SEMANTIC ERRORS FOUND IN LEXICAL TRANSLATIONS OF TASYA FARASYA’S TIKTOK ACCOUNT

Ni Putu Laksmi Dewi Utami¹; Ni Made Verayanti Utami²*

¹-²Faculty of Foreign Languages, Mahasaraswati Denpasar University
Jl. Kamboja No. 11A, Dangin Puri Kangin, Denpasar, Bali 80233, Indonesia
¹laksmidewi7@gmail.com; ²verayanti.utami@unmas.ac.id

Received: 12th September 2023/Revised: 01st November 2023/Accepted: 04th November 2023


ABSTRACT

This research focused on the semantic errors in lexis that occurred in the translation by TikTok machine translation. It became the main issue in translation studies because the accuracy of translation produced by machine translation was still questionable and debatable. This research aimed to identify the types of semantic error in lexis made by TikTok auto machine translation found in Tasya Farasya TikTok’s account and suggested a more appropriate translation. Methodologically, the research applied a descriptive qualitative study to analyze the error in translation produced by TikTok machine translation. The theory proposed by Sayogie (2014) was used to classify the data based on semantic aspects: grammatical meaning, contextual meaning, and referential meaning. The research results show that three types of errors are found, and the most frequent error found is an error in contextual meaning. TikTok machine translation is incapable of translating accurately because it does not know the context of the situation and translates it literally. Based on research findings, TikTok users cannot entirely rely on machine translation because it still has weaknesses in translating several terms. Thus, it is highly important that TikTok should evaluate and improve the quality of the machine translation.

Keywords: translation, semantic errors, TikTok machine translation

INTRODUCTION

Language is essential in human life as social creatures in order to communicate with each other and also learn new information. People require language as an introduction to learning a specific language, and they must understand the language in order to comprehend the contents of the learning (Utami & Irwandika, 2021). People use language in various fields, such as education, religion, technology, etc. However, the world has many different languages and cultures, creating fantastically rich linguistic diversity. Thus, it results as a barrier to communication (Bowker, 2020; Kumar et al., 2019; Kunchukuttan & Bhattacharyya, 2022).

According to Larson (1998), translation is essentially a change of form. It is further defined by Sari (2019) as converting one language (Source Language/SL) to another (Target Language/TL) in order for the target language to recognize the intended message in the source language. In the field of linguistics, there exists a branch called semantics, which is concerned with the study of meaning. A good translation at least contains semantic aspects such as referential, grammatical, and contextual meanings. It is important to ensure the meaning of SL is accurately conveyed and avoid misunderstanding or confusion in TL.

In the past, the translation process was performed by a human translator; however, due to the development of technology, there are machine translations such as Google Translate, DeepL Translate, Bing, and many more options for translating texts. Today, machine translation plays an important role in how people communicate, work, travel, learn, access information, and more throughout the world (Lee, 2020; Khoong & Rodriguez, 2022). The term Machine Translation (MT) can be defined as computerized systems that produce the most similar meaning of the source
language with or without human assistance (Gupta & Thakral, 2019; Kumar et al., 2019; Liu, Sun, & Wang, 2021). Both SL and TL users can choose their preferred language. Afterward, machine translation will transfer the meaning from SL to TL quickly. Even though the machine translation system has evolved and has the ability to produce a good translation, it cannot be an indicator that the output is reliable, and people should not be dependent on this (Hoi, 2020; Ying et al., 2021).

Up to the present time, the accuracy of translation produced by machine translation is still questionable and debatable (Nasution, 2022). Several studies indicate that MT outputs are still far from being equal to human outputs (Läubli, Sennrich, & Volk, 2018; Läubli et al., 2020; Putri & Setiajid 2021; Utami, Jayantini, & Pratiwi, 2021). It becomes an issue due to the complexity of natural languages, where many words have multiple meanings and different translation versions (Susanti, 2018; Tan et al., 2019; Han, Jones, & Smeaton, 2021). In addition, even though machine translation continues to develop, its results are not evenly distributed (Fan et al., 2021). Thus, it is a concern that machine translation could not fulfill the requirements of good translation and would affect the output. However, due to the high demand, this machine translation has also been added to social media as an in-app auto-translation feature. Some social media that provide that feature are TikTok, Twitter, Instagram, and Facebook.

Social media is used massively nowadays to deliver messages and share information globally. Users can create and follow multiple social media accounts (Putri & Setiajid, 2021). Many entertainers, politicians, educators, and economists share their thoughts or information through their own social media. Moreover, creators and brands use social media as a platform to advertise and sell their products. One of the most popular social media platforms recently is TikTok.

TikTok is a social media platform that allows users to create and post short videos on various topics. With the rise of TikTok, almost all the known celebrities, influencers, and big companies can be found on this platform. One of the most followed influencers in Indonesia is Tasya Farasya. She is known as a beauty content creator who often talks about beauty matters such as reviewing make-up and skincare. The use of social media for a beauty influencer is significant because, through it, they can spread their influence in a particular field to their followers. Through social media, a beauty influencer can gain popularity and trust from the public. Now, she is followed by 3.4 million people on TikTok and continues to grow.

With that massive number of followers from different regions and cultures, there is a high probability that non-native speakers will use the TikTok auto-translation feature. TikTok translation helps content creators break into new markets where language would have been a barrier. As a result, more people from various demographics will get to see and understand the translated video. TikTok provides many translation features; the most recent one is auto-generated captions. If the user turns this feature on, an algorithm will automatically transcribe the audio into text. However, the translation product made by TikTok machine translation has the probability of producing semantic errors such as using wrong grammar and even wrong context. In this case, machine translation cannot meet the requirements of a good translation, which will affect the result and the meaning of the target language. The fact that even well-developed MT systems possess substantial flaws emphasizes the significance of recognizing the capabilities and limitations of this quickly developing technology (Vieira, O’Hagan, & O’Sullivan, 2021).

There is much research related to semantic errors in translation. One of them is the research conducted by Pudjiati and Fadilah (2017). This research examines semantic errors in translated captions in the Hizbut Tahrir Indonesia (HTI) Instagram account. It has been found that the translated captions contain semantic errors in grammatical, contextual, and referential meaning. However, the research on semantic errors in TikTok machine translation, especially in auto-generated captions, has never been done. The situation mentioned above is undoubtedly undesirable in matters of translation. Moreover, research on the quality of TikTok machine translation may not have been conducted as extensively as the other MT. Thus, this research is conducted to uncover the types of translation errors produced by TikTok machine translation based on semantic aspects and suggest a more appropriate translation.

**METHODS**

The research uses descriptive-qualitative research since it aims for descriptive analysis of Indonesian-English error translation phenomena produced by TikTok machine translation. Lambert and Lambert (2012) have stated that the goal of qualitative descriptive studies is a comprehensive summarization, in everyday terms, of specific events experienced by individuals or groups of individuals.

The data in this research is the auto-generated captions taken from Tasya Farasya’s TikTok videos that provide the feature. This account has been verified by TikTok, and it has 3.4 million followers currently. The source language in Indonesia is translated into the target language in English. The Indonesian-English translation is done by the TikTok machine translation. The instrument used for collecting the data in this research is direct observation. It is observed directly through the caption and the translation, which is done by TikTok machine translation. The captions used as the data are chosen randomly from the videos related to skincare and make-up posted in Tasya Farasya’s account from May until August 2023, which provided auto-generated captions. The data from the last three months is chosen to give a more accurate representation of the newest updates in machine translation.
Two techniques are used in collecting the data. The first technique is transcribing the Indonesian and English versions of the auto-generated captions into word document files. The second technique is reading and highlighting the data representing the semantic error in lexis. After the data are collected, they are analyzed further by comparing the Indonesian and English versions of the auto-generated captions regarding the meaning and suggesting more appropriate ones. The semantic errors are categorized and analyzed based on the theory proposed by Sayogie (2014).

RESULTS AND DISCUSSIONS

In this research, the type of errors is categorized and analyzed using Sayogie’s theory (2014). He divides it into seven aspects: linguistics (transposition, modulation, and adaptation), semantics (grammatical meaning, contextual meaning, and referential meaning), pragmatics (writer’s goal of source language agreement, textual meaning agreement), the level of naturalness, particular terminology, spelling use, and textual level. This research applies semantic aspects, which are grammatical meaning, contextual meaning, and referential meaning. The data presented in Table 1 are the translation errors selected from the auto-generated captions.

From the data in Table 1, there are three errors of grammatical meaning, five errors of contextual meaning, and two errors of referential meaning. An error in contextual meaning is the most frequently encountered error in the translation produced by the TikTok machine translation. Furthermore, the data above are treated by comparing the translation between SL and TL, analyzing the translation errors from the semantic point of view, elaborating between theories, and suggesting a more appropriate translation.

According to Kridalaksana (2008), grammatical meaning is the relationship between the language elements in the larger unit, such as the word “can” means dapat or kaleng, depending on the position of the word in a sentence. In the translation process, the meaning of a word may differ from its grammatical

Table 1 The Semantic Error in Lexis Found in Tasya Farasya TikTok Account

<table>
<thead>
<tr>
<th>No</th>
<th>Source Language</th>
<th>Target Language</th>
<th>GM</th>
<th>CM</th>
<th>RM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yang harus kalian perhatikan adalah mereka tuh selalu matte complexion fokusnya tuh malah di glitter.</td>
<td>What you should note is that they are always matte complexion. The focus was glittered.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Tapi emang kalau aku pribadi bukan selera aku sih, kayak complexion-nya se-matte ini.</td>
<td>But personally do not taste me, like this matte complexion.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>By the way aku udah pernah sih nge-review lip cream velvet Brasov ya.</td>
<td>By the way I have review lip cream velvet Brasov.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Ini super powdery tapi kayanya pigmentasinya memang nggak terlalu heboh deh.</td>
<td>I dress up naturally and wear nothing on her lips, means we’re gonna lip swatch.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Langsung aja ya, kalau udah ageliat aku dandan natural dan bhirinya nggak pakai apa-apa berarti kita akan lip swatch.</td>
<td>I still use the ponds. Honestly make a shower too last night with the hope of the height and my skin coming back soon.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Aku masih pakai si Ponds.Jujurly buat mandi juga semalem dengan harapan tinggi badan aku dan kutil aku segera kembali ya.</td>
<td>Wear thin it can be a powder, if overwritten to the extreme, it can look like illuminating powder.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Pakai tipis dia bisa jadi bedak, kalau ditimpan tipina yang sampai agak ekstrim, dia bisa terlihat seperti kayakin illuminating powder.</td>
<td>I want poison in the middle of Bali, you know? And this is really comfortable. Badly beautiful color.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Aku pengen nge-racun di tengah Bali, tahu nggak sih. Pakai Wardah ya dan ini tuh nyaman banget. Cantik paruh warnanya.</td>
<td>Don’t just focus on keeping your face, so his face is glowing but his skin is neglected.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Jangan cuma fokus menjaga wajah kalian, jadi wajahnya doang yang glowing tapi kulitnya ini terbengkalai gitu.</td>
<td>Because he’s velvet too so he won’t crash too much if on the eyelids.</td>
<td>Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: GM (Grammatical Meaning), CM (Contextual Meaning), RM (Referential Meaning), Y (Yes)
meaning if it stands alone. The change in grammatical structure often happens to produce a reasonable translation. However, it is only applicable without changing or eliminating the idea of meaning contained in the SL. It can be seen that three out of ten data contain errors in grammatical meaning.

Data 1
TL : What you should note is that they are always matte complexion. The focus was glittered.

This data contains an error in grammatical meaning. In the SL, the word ‘glitter’ works as a noun. TikTok machine translation makes an error in translating the whole sentence into passive voice and changes the word class of the word ‘glitter’ from noun to verb. In SL, there is no subject that receives an action. The point of the sentence is to highlight that the focus is the glitter. Meanwhile, TL has a subject that is the focus and receives an action. This grammatical structure is incorrect and gives the translation a whole new meaning.

Data 2
SL : Tapi emang kalau aku pribadi bukan selera aku sih, kayak complexion-nya se matte ini.
TL : But personally do not taste me, like this matte complexion.

Data 2 has a similar error to Data 1 in which the change of word class happened, but it does not match the meaning. In the SL, the noun phrase ‘selera aku’, which refers to the creator’s taste in a make-up look, is translated into ‘taste me’ with a completely different meaning. The word ‘selera’ in TL is a noun, while ‘taste’ in TL is a verb. The correct translation of the noun phrase ‘selera aku’ would be ‘my taste’.

Data 3
SL : By the way aku udah pernah sih nge-review lip cream velvet Brasov ya.
TL : By the way I have review lip cream velvet Brasov.

This data is also categorized into grammatical meaning errors. The phrase “I have” shows possession or something acquired. In the SL, it is shown that the creator informs the audience of a past or completed action, which is reviewing the product. For something already done in the past, it should use past perfect tense. The past perfect tense is formed by using the word ‘had’ followed by the past participle of the verb. However, in the TL, it is translated into ‘I have review’, which is grammatically incorrect.

Based on Table 1, several data contain errors in contextual meaning. Contextual meaning is the relationship between the utterance and the situation in which it is used. Every time people use language, their words appear in a specific setting and have a unique context that shapes the meaning of the utterances. Several factors, such as cultural beliefs and social norms, affect how the language is used and how it is interpreted. From the ten data analyzed in Table 1, this is the most frequent error produced by TikTok machine translation.

Data 4
SL : Ini super powdery tapi kayaknya pigmentasinya memang nggak terlalu heboh deh.
TL : This is super powdery but I don’t think the pigmentation will be too excited.

In order to make a correct translation, the context of the situation must be known. In this case, an error is found in the translation produced by TikTok machine translation. Data 4 contains a translation error in contextual meaning in translating the term ‘heboh’ into ‘excited’. In the video, the creator applies a blush-on and reviews the product’s pigmentation. According to Cambridge Dictionary, the term ‘excited’ means feeling very happy and enthusiastic; meanwhile, the term ‘heboh’ is often used as a slang word that refers to something that is dramatic or excessive. Thus, the meaning of SL and TL does not match.

Data 5
SL : Langsung aja ya, kalau udah ngeliat aku pandan natural dan bibirnya nggak pakai apa-apa berarti kita akan lip swatch.
TL : I dress up naturally and wear nothing on her lips, means we’re gonna lip swatch.

This translation has errors in contextual meaning. The phrase ‘dress up’, according to the Cambridge Dictionary, means to put on special clothes in order to change someone’s appearance. In the Indonesian language, the word ‘dandan’ in KBBI means to put on some clothes, accessories, and make-up. The video shows the creator only showing her face with some make-up on. She specifically refers to the word ‘dandan’ as she has already put some make-up on her face. However, TikTok machine translation fails to deliver the specific meaning of the term.

Data 6
TL : I still use the Ponds. Honestly make a shower too last night with the hope of the height and my skin coming back soon.
There are three errors found in this data categorized in contextual meaning. The context of the video is that the creator explains a cleansing product she has been using. She also used it last night and hoped her skin would return to normal (regarding the condition or skin tone). The first error found in the translation produced by TikTok auto-machine translation is translating the word ‘buat’ into ‘make’. Even though in the Indonesian language, the word ‘buat’ means to produce or create something, it cannot be translated literally. The correct term to replace it would be using the preposition ‘for’, which means intended to be given to or used by someone or something, according to Cambridge Dictionary. The second one is that the word ‘kembali’ is mistakenly translated into ‘coming back’ without seeing the whole context. It should be replaced with ‘recovered’. It is because the skin cannot go anywhere and return to its original place. The last one is TikTok machine translation, which translated the adjective word ‘tinggi’, which means greater than usual level into a noun word ‘height’, which means the distance from the top to the bottom of something. The phrase is supposed to be ‘high hope’.

Data 7

SL : Pakai tipis dia bisa jadi bedak, kalau ditimpa-timpa yang sampai agak ekstrim, dia bisa terlihat seperti kayak illuminating powder.

TL : Wear thin it can be a powder, if overwriten to the extreme, it can look like illuminating powder.

In this data, there is an error in contextual meaning, which can be seen in the translation of the word ‘ditimpa’ that is translated into ‘overwriten’. The context here is that the creator shows the audience how to wear a compact powder by applying it on her face. She then elaborates that adding more of this product would make it look like illuminating powder. The word ‘ditimpa’ refers to adding more layers, in this case, more compact powder. Meanwhile, according to the Cambridge Dictionary, the word ‘overwriten’ (past participle of overwrite) means to replace something with a different one or to write over the surface. Thus, the meaning of SL and TL does not match. It should be replaced with the term ‘applied more’.

Data 8


TL : I want poison in the middle of Bali, you know? And this is really comfortable. Badly beautiful color.

A few words in the Indonesian language have newer meanings and are used as slang words by youngsters. Examples can be seen in this data, such as the use of the word ‘racun’ and ‘parah’. According to KBBI Daring, ‘racun’ means a substance that can cause illness or death (if eaten or inhaled), and ‘parah’ means very bad or difficult to overcome. The creator uses those two terms in the video, and the machine translation fails to recognize the correct meaning and translates it literally. This kind of error is categorized as a contextual meaning error. The word ‘racun’ or ‘ngeracun’ (verb) in this context means to persuade someone to buy something, which is the lip product that the creator wears in the video. The second one is the word ‘parah’, which is translated into the adverb ‘badly’ when the real meaning of this term in this context is equal to the word ‘very’. The creator tells the audience that the product she is wearing has a gorgeous color.

Larson (1998) has stated that referential meaning is something that is referenced or referenced directly, which may take the form of objects, events, attributes, or particular relations that can be seen or imagined by words or sentences. Reference itself is a linguistic form that enables someone to identify something depending on the speaker’s intention so the listener can identify the speaker’s intention. There are two out of ten data that contain errors in referential meaning.

Data 9

SL : Jangan cuma fokus menjaga wajah kalian, jadi wajahnya doang yang glowing tapi kulitnya ini terbengkalai gitu.

TL : Don’t just focus on keeping your face, so his face is glowing but his skin is neglected.

This data shows an error in referential meaning. An error is found in referential meaning, which is translating the suffix -nya into ‘his’. In this case, the creator refers to the things that have already been discussed, which are ‘face’ and ‘skin’. Thus, the translation made by the TikTok translation machine shows different meanings, and it can be replaced by using the determiner ‘the’ instead of using a personal pronoun.

Data 10

SL : Karena dia velvet juga, jadi dia nggak bakal terlalu nge-crease juga sih kalau di kelopak mata.

TL : Because he’s velvet too so he won’t crash too much if on the eyelids.

In this case, the error in referential meaning is using the word ‘dia’ as a pronoun. The difference is that in this context, the creator uses ‘dia’ to refer to an object, the product she was using. However, TikTok machine translation does not recognize it and translates it into ‘he’, which refers to a male person. The correct pronoun would be ‘it’ instead of ‘he’.

In conclusion, there are errors in semantic
aspects found in the translation results carried out by TikTok machine translation. The most frequent error that occurs is an error in contextual meaning. TikTok machine translation is incapable of translating accurately because it does not know the context of the situation and translates it literally. Contrary to the findings of research conducted by Pudjiati and Fadilah (2017), the most frequently encountered error type on Instagram machine translation is the error in grammatical meaning.

For other researchers, research on semantic errors in TikTok’s machine translation offers an opportunity to explore and address the challenges in machine translation technology within the context of a popular global platform. It can serve as a foundation for further investigations into improving automated translations and enhancing cross-cultural communication in digital environments. It can also support TikTok’s global expansion and drive technological advancements in language technology.

CONCLUSIONS

This research focuses on the semantic errors in lexis that occurred in the translation by TikTok machine translation. It becomes the main issue in translation studies because the accuracy of translation produced by machine translation is still questionable and debatable. This research aims to identify the types of semantic errors in lexis made by TikTok’s automatic machine translation found in Tasya Farasya TikTok’s account and suggest a more appropriate translation. According to the results and discussions, it can be concluded that all the data contain errors in grammatical meaning, contextual meaning, and referential meaning. From the ten data analyzed, there are three errors of grammatical meaning, five errors of contextual meaning, and two errors of referential meaning. An error in contextual meaning is the most frequently encountered error in the translation produced by the TikTok machine translation. This phenomenon occurs because machine translation does not have the quality that human translators have. Human translators have the capability of obtaining background knowledge, such as the speaker, the concepts, and previous utterances, in order to acknowledge the speaker’s intended meaning.

This research makes a significant contribution to the study of digital communication and language technology, focusing on the critical issue of semantic errors in translations generated by TikTok’s machine translation systems. As TikTok continues to expand its global user base, the need for accurate and contextually meaningful translations becomes increasingly important. This investigation delves into the nature and prevalence of semantic errors in TikTok’s automated translations, shedding light on the challenges inherent in preserving the intended meaning across different languages and cultural contexts. Through rigorous analysis and innovative solutions, this research advances the understanding of the complexities of machine translation, ultimately contributing to more effective cross-cultural communication on TikTok and similar platforms. By addressing and mitigating semantic errors, it can facilitate more nuanced and culturally sensitive interactions, fostering a more prosperous and inclusive digital environment for users worldwide.

While this research makes valuable contributions, it is essential to acknowledge its limitations. Firstly, the scope of the research may be constrained by the available data, which could be limited to the specific languages and content sampled on TikTok during the research period. This limitation may not capture the full spectrum of translation errors or the platform’s evolution over time. Additionally, the effectiveness of proposed solutions to address semantic errors in TikTok’s translations may need further validation and practical implementation, which falls beyond the scope of this research. Real-world factors, such as the constantly evolving algorithms and user-generated content on TikTok, may impact the long-term applicability and success of these solutions.

REFERENCES


Kumar, M. A., Premjith, B., Singh, S., Rajendran, S., & Soman, K. P. (2019). An overview of the shared task on machine translation in Indian languages (MTIL-).
Unveiling Semantic Errors .... (Ni Putu Laksmi Dewi Utami; Ni Made Verayanti Utami) 225


