

## ARTICLE

# Semantic Errors in Machine Translation: A Study of Colloquial Arabic-English Captions on TikTok

Shatha Abdullah Alshaye 

Department of English, College of Language Sciences, King Saud University, Riyadh 11652, Saudi Arabia

## ABSTRACT

This study investigates the semantic accuracy of TikTok's machine translation (MT) system, focusing on auto-generated English captions translated from colloquial Arabic content posted by the verified MBC1 and Shahid TikTok accounts. The primary objective is to identify and categorize semantic errors using Sayogie's (2014) framework, which classifies meaning into three interrelated dimensions: grammatical, contextual, and referential. Employing a descriptive qualitative approach, the research analyses a sample of colloquial Arabic captions to evaluate the extent to which TikTok's translation system preserves the intended meaning. Findings reveal a consistent presence of all three error types, underscoring the system's limitations in processing idiomatic, context-dependent expressions typical of colloquial Arabic. These inaccuracies primarily result from the system's reliance on literal translation strategies, which fail to account for figurative language, cultural references, and emotional nuance. While TikTok's MT feature enhances accessibility for multilingual users, it remains inadequate in preserving semantic integrity, particularly in informal, culturally embedded content. To address these challenges, the study recommends developing more context-aware, dialect-sensitive models capable of handling the pragmatic and cultural complexity of colloquial speech. The findings contribute to current debates in MT evaluation by highlighting the need to prioritize semantic adequacy, especially for low-resource and dialect-rich languages such as Arabic.

**Keywords:** Colloquial Language; Machine Translation; Semantic Errors; TikTok; Translation Errors

### \*CORRESPONDING AUTHORS:

Shatha Abdullah AlShaye, Department of English, College of Language Sciences, King Saud University, Riyadh 11652, Saudi Arabia;

Email: [salshaye@ksu.edu.sa](mailto:salshaye@ksu.edu.sa)

### ARTICLE INFO

Received: 25 June 2025 | Revised: 7 July 2025 | Accepted: 15 July 2025 | Published Online: 1 September 2025

DOI: <https://doi.org/10.30564/fls.v7i9.10711>

### CITATION

AlShaye, S.A., 2025. Semantic Errors in Machine Translation: A Study of Colloquial Arabic-English Captions on TikTok. *Forum for Linguistic Studies*. 7(9): 159–171. DOI: <https://doi.org/10.30564/fls.v7i9.10711>

### COPYRIGHT

Copyright © 2025 by the author(s). Published by Bilingual Publishing Group. This is an open access article under the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License (<https://creativecommons.org/licenses/by-nc/4.0/>).

# 1. Introduction

Language plays a fundamental role in human life as it facilitates communication and the acquisition of new knowledge. As social beings, individuals rely on language to interact with others and engage with educational content. Mastery of language is a prerequisite for understanding the materials being studied <sup>[1]</sup>. Language is utilized across various domains such as education, religion, and technology. However, the global linguistic landscape is marked by a vast diversity of languages and cultures, which poses significant barriers to effective communication <sup>[2-4]</sup>.

Translation serves as a crucial tool for overcoming these barriers. Larson <sup>[5]</sup> defines translation as a change of linguistic form, while Sari <sup>[6]</sup> expands this definition to the process of transferring meaning from a source language (SL) to a target language (TL), enabling the intended message to be understood by TL users. In linguistics, semantics, which is concerned with studying meaning, plays a pivotal role in assessing translation quality. An effective translation should convey semantic components such as referential, grammatical, and contextual meanings to ensure accurate interpretation and prevent misunderstandings.

Historically, translation was carried out exclusively by human translators. However, technological advancements have led to the development of machine translation (MT) systems, including but not limited to Google Translate, DeepL, and Bing Translator. These tools significantly facilitate global communication, learning, and information access <sup>[7,8]</sup>. MT refers to computerized systems capable of rendering the closest possible equivalent of the SL in the TL, with or without human input <sup>[3,9,10]</sup>. While MT allows users to choose their preferred languages and receive quick translations, the quality and reliability of its output remain contentious <sup>[11,12]</sup>.

This study investigates the quality of English machine translations of colloquial Arabic captions extracted from the official TikTok accounts of MBC1 and Shahid. Specifically, it examines the semantic accuracy of the translated captions to identify the extent to which TikTok's MT system effectively conveys meaning from the SL. The research highlights limitations associated with MT technologies in the context of informal, user-generated content on social media. The central research question guiding this study is:

What types of semantic errors are present in the English translations produced by TikTok's MT of colloquial Arabic captions?

# 2. Literature Review

Colloquial language, characterized by informal vocabulary, idiomatic expressions, contractions, and context-dependent meanings, poses significant challenges for translation, especially between linguistically and culturally distant

language pairs such as Arabic and English <sup>[13]</sup>. Early research in MT and natural language processing (e.g., Popowich <sup>[14]</sup>) highlighted the limitations in handling colloquial expressions. These challenges are further compounded in Arabic due to its nature, where colloquial varieties (e.g., Egyptian, Levantine, Gulf) differ significantly from Modern Standard Arabic (MSA), and from one another, in phonology, syntax, and vocabulary <sup>[15]</sup>.

Translation between colloquial Arabic and English has remained underexplored until recently, when the rise of user-generated content and social media (e.g., TikTok, Instagram, YouTube) introduced a new urgency to understand how machine and human translators process informal, real-time language. Recent studies (e.g., Mirhashemi, Gholami, & Bahri <sup>[16]</sup>) show that modern MT tools like Google Translate and ChatGPT have improved in handling certain colloquial expressions, particularly in high-resource language pairs like English-Spanish or English-French. However, the translation quality across Arabic-English remains inconsistent, often failing to capture the pragmatics, context, and cultural dimensions embedded in colloquial usage. This is evident in translating slang, idioms, humor, and speech acts, where literal rendering leads to semantic errors or complete miscommunication <sup>[17]</sup>.

Furthermore, the scarcity of parallel corpora for colloquial Arabic and other low-resource language pairs hampers the development of robust MT systems. Researchers such as Fattah & Salih <sup>[18]</sup> emphasize the importance of considering community-specific usage and social context in translation, aligning with the notion that colloquial language is best understood through the lens of shared practices and localized meaning. These findings underscore the need for translation models incorporating sociolinguistic, semantic, and pragmatic dimensions, especially in domains where visual and contextual cues, such as multimedia platforms, are integral to meaning construction.

## 2.1. Machine Translation

The accuracy of machine-generated translations continues to be questioned <sup>[19]</sup>. Numerous studies suggest that MT outputs still fall short of matching the quality of human translations <sup>[20-23]</sup>. This is primarily attributed to the inherent complexity of natural language, including polysemy and context-dependent meaning, which pose significant challenges for MT systems <sup>[24-26]</sup>. Furthermore, the quality of MT results varies significantly across different languages and contexts <sup>[27]</sup>. As a result, current MT technologies often fall short of meeting the criteria for high-quality translation, potentially compromising the intended meaning.

Despite these limitations, the increasing demand for translation tools has led to integrating MT features into social media platforms. Popular platforms like TikTok, Twitter, Instagram, and Facebook now include auto-translation features

to support cross-lingual communication. Social media has become a pervasive medium for global information exchange, with users from diverse professional and personal backgrounds utilizing it to share content and connect with audiences worldwide <sup>[22]</sup>. Many public figures—celebrities, politicians, educators, and entrepreneurs—use social media to promote their ideas and products.

One of the most prominent social media platforms today is TikTok, which allows users to create and share short videos on various topics. Influencers, celebrities, and corporations widely use this platform to engage with audiences. In Arabic-speaking countries, Middle East Broadcasting Center 1 (MBC1) is a prominent and widely recognized television channel renowned for its diverse programming, including news, entertainment, cultural content, talk shows, and drama series. As one of the flagship channels of the MBC Group, MBC1 has played a significant role in shaping media consumption across the Arab world since its launch in 1991. Over the years, it has evolved from a traditional broadcaster into a multimedia brand with a substantial digital footprint. The channel targets a pan-Arab audience, offering content that reflects regional identities and global trends, and often addresses sociopolitical, religious, and cultural issues that resonate with viewers in the Middle East and North Africa (MENA) region.

In alignment with the global shift toward digital media consumption, MBC1 has established a strong presence across various social media platforms, particularly TikTok, which currently has over 7.7 million followers. This substantial following reflects the channel's ability to adapt to changing viewer habits and highlights the growing importance of short-form video content in contemporary media landscapes. MBC1's TikTok content typically includes highlights from popular shows, behind-the-scenes clips, interviews, and culturally relevant discussions, extending its reach beyond traditional broadcast audiences.

In addition to MBC1, MBC Group also owns and operates Shahid, a leading Arabic-language digital streaming platform that offers on-demand access to a wide range of drama series, films, and original content. Shahid likewise maintains a verified TikTok account with over 7.8 million followers, underscoring its strong presence in the social media ecosystem and its appeal to digitally engaged Arabic-speaking audiences.

The extensive online presence of both MBC1 and Shahid enhances their visibility and illustrates how social media platforms are critical tools for building credibility, maintaining public engagement, and reaching linguistically and culturally diverse audiences. Platforms like TikTok have enabled media outlets to transcend geographic and linguistic boundaries, connecting with viewers who may not be fluent in Arabic. Given the large and diverse audiences of MBC1

and Shahid, it is highly probable that many non-native Arabic speakers rely on TikTok's auto-translation features, such as auto-generated captions and subtitle translation, to understand the content. These translation features play a pivotal role in enhancing accessibility, but they also introduce potential challenges related to semantic accuracy, contextual appropriateness, and cultural sensitivity.

TikTok's translation features, including its most recent addition of auto-generated captions, aim to overcome language barriers and expand content reach. These captions are generated through algorithms that transcribe spoken audio into text and then translate it into other languages. While this tool facilitates broader access to content, it also introduces the risk of semantic errors. These may include grammatical inaccuracies, contextual misinterpretations, or erroneous lexical choices, ultimately compromising the quality of the translation and its intended meaning. As noted by Vieira, O'Hagan, and O'Sullivan <sup>[28]</sup>, the presence of errors in outputs produced by even the most advanced MT systems underscores the importance of critically assessing their capabilities and limitations.

## 2.2. Translation Errors

Sayogie's <sup>[29]</sup> framework for translation error analysis offers a comprehensive and multidimensional approach to evaluating translation quality. Sayogie categorizes errors into seven distinct aspects encompassing various linguistic and communicative concerns.

1. **Linguistic Errors:** include issues related to structural transformations such as transposition, modulation, and adaptation. These procedures involve changes in grammatical structure, shifts in perspective or viewpoint, and cultural or contextual adaptations that may be necessary for achieving equivalence in the TL.
2. **Semantic Errors:** This category includes inaccuracies in conveying meaning and is further divided into:
  - Grammatical meaning (e.g., tense, aspect, number)
  - Contextual meaning (the meaning derived from surrounding textual elements)
  - Referential meaning (relation to specific entities or concepts in the real world)
3. **Pragmatic Errors:** the failure to preserve the communicative intent or function of the source text (ST). Pragmatic errors include:
  - Inconsistencies with the writer's intended goal
  - Disagreement in textual meaning, particularly where the tone, level of formality, or purpose is distorted
4. **Naturalness Level:** This refers to how fluent and idio-

matic the translation sounds to native speakers of the TL. Errors in this category indicate a translation that may be grammatically correct but unnatural or awkward.

5. Terminological Errors: These involve misusing or inconsistently using specific terminology, especially in domain-specific or technical texts. Inaccuracies here can severely impact comprehension and credibility.
6. Spelling Errors: Although surface-level, these errors can affect readability and professionalism and are considered part of basic mechanical accuracy in the TT.
7. Textual-Level Errors: These refer to broader issues involving the overall structure and cohesion of the translated text. They include problems with coherence, paragraph organization, logical flow, and unity across sentences and sections.

The body of research on semantic errors in machine translation (MT) remains relatively limited. Addressing this gap, Cao et al. [30] introduce SemMT, a novel testing framework that prioritizes semantic accuracy in evaluating MT outputs. Unlike conventional methods that rely on textual or syntactic comparisons, SemMT focuses on semantic similarity, with particular emphasis on quantifiers and logical relations, elements often overlooked by standard evaluation techniques, despite their potential to cause serious misinterpretations. The approach applies round-trip translation and models sentence meaning using regular expressions or finite automata, allowing for efficient comparisons between source and translated texts. The authors propose three semantic similarity metrics, implemented within the SemMT system. Experimental results show that SemMT outperforms existing techniques, achieving a 34.2% increase in accuracy and a 15.4% improvement in F-score for mistranslation detection. The study also investigates combining metrics for enhanced performance and proposes a method for locating problematic segments in round-trip translations, thereby aiding in translation error diagnosis. Another study by Utami and Utami [31] compared the Indonesian and English versions of the auto-generated translations of Tasya Farasya's TikTok account, identifying semantic inaccuracies that distort intended meanings. Likewise, Läubli et al. [21] argue that sentence-level evaluations are inadequate and call for document-level and meaning-focused assessments that better capture the communicative intent of the ST. These frameworks reflect a broader shift toward semantic adequacy as a central measure of translation quality.

Such developments complement the current study's adoption of Sayogie's [29] model, which emphasizes grammatical, contextual, and referential meaning—three core aspects of semantic integrity that are frequently compromised in machine translations of colloquial Arabic. However, there remains a noticeable gap in research concerning the semantic

accuracy of TikTok's Arabic-English auto-captioning system. This is particularly problematic given the platform's global reach and the pivotal role its embedded MT plays in shaping user interpretation. Further research is therefore needed to examine the types of semantic errors generated, assess their communicative consequences, and propose more context-sensitive and culturally appropriate translation solutions.

### 3. Method

This study adopts a descriptive-qualitative research method to analyze the semantic accuracy of Arabic-to-English translations generated by TikTok's MT system. According to Lambert and Lambert [32], qualitative descriptive studies seek to provide a comprehensive and straightforward account of particular phenomena as experienced or observed in real-world contexts. This research focuses on the semantic accuracy of machine-translated captions, specifically Arabic captions translated into English by TikTok's in-app auto-translation feature.

The primary data consists of auto-generated captions from videos posted on the official MBC1 and Shahid TikTok accounts, which is verified and followed by over 7.7 million users. The analyzed captions appeared in videos related to the *Aser* drama series, published during April and May 2025, a timeframe selected to capture the most recent developments and updates in the performance of TikTok's translation algorithms. *Aser* drama series, delivered in Levantine Arabic, represents a culturally embedded narrative, making it an ideal case for analyzing how TikTok's auto-captioning and translation tools handle colloquial Arabic content. The selection of the *Aser* series provides access to high-quality, relevant data. It offers valuable insights into the performance, limitations, and cultural sensitivity of automated translation systems on social media platforms, especially when dealing with spoken regional dialects such as Levantine Arabic.

Data collection was carried out through direct observation, focusing on the comparison between original Arabic captions and their corresponding English outputs generated by TikTok's automated translation system. The dataset comprised fifteen caption pairs collected from two verified TikTok accounts, namely MBC1 and Shahid. All captions were drawn from *Aser* drama series and were chosen based on clear discrepancies in meaning between the source and target texts. Caption pairs were selected based on three criteria: (1) their use of colloquial Arabic, (2) the presence of emotionally expressive or culturally nuanced language, and (3) their recency, with all captions posted between April and June 2025.

The first ten caption pairs were drawn from MBC1 TikTok account. The selection process prioritized instances where semantic inaccuracies, particularly those related to grammatical, contextual, or referential meaning, were most



evident, in alignment with Sayogie's <sup>[29]</sup> classification framework. To enhance the generalizability of the findings and assess whether the observed semantic issues reflect broader systemic patterns rather than account-specific inconsistencies, an additional five caption pairs were analysed from the verified Shahid TikTok account. This enabled a more robust cross-account comparison, providing further insight into the consistency and nature of semantic errors across different content sources on the same platform. Two main techniques were employed in collecting and analyzing the data:

1. Selection and transcription of fifteen Arabic captions and their English translations from the auto-generated content. The selected captions demonstrate a significant divergence in meaning between the source and target texts.
2. Closely read and highlight translation instances that contain semantic errors, particularly those related to lexical choice and meaning.

Following data collection, the analysis involved a detailed comparison of the source and target captions to assess how accurately the intended meanings were conveyed. Alternative translations were proposed in cases where the machine-generated output failed to preserve the intended meaning. All captions were initially analysed by the author and then cross-verified by an independent coder, an academic colleague with expertise in Arabic-English translation. To ensure consistency in error categorization, both coders applied the same coding scheme based on Sayogie's <sup>[29]</sup> model. Discrepancies were discussed and resolved collaboratively through consensus.

The semantic errors were categorized and analyzed using the semantic dimension of Sayogie's error classification model <sup>[29]</sup>, which includes three core components: grammatical meaning: errors involving tense, number, or word class that affect meaning, contextual meaning: errors arising from misinterpretation of meaning based on surrounding textual cues, and referential meaning: errors in the translation of terms referring to specific entities or ideas. Sayogie's framework <sup>[29]</sup> is particularly suitable for assessing meaning-based translation issues in informal, user-generated content. It allows for a focused and systematic analysis of how MT handles semantic complexity in colloquial Arabic captions within the dynamic context of social media. Sayogie's model <sup>[29]</sup> is beneficial for identifying and analyzing semantic errors in formal and informal texts, including user-generated content such as social media captions, as researchers can identify whether a translation is inaccurate and how and why the inaccuracy occurs.

## 4. Analysis

### 4.1. MBC1 Account

Sayogie

- 1- ST: دموع أسر

TT: **Captivating** tears <sup>[33]</sup>

The translation of the Arabic phrase “دموع أسر” into “Captivating tears” contains significant semantic translation errors. Grammatically, “دموع أسر” is a possessive noun phrase that should be directly rendered as “Aser's tears”, maintaining the syntactic structure of possession. The translated phrase fails to preserve this relationship by replacing the possessive form with an evaluative adjective; thus, constituting a grammatical meaning error. Regarding contextual meaning, the ST is emotionally and personally associated, as it refers to the tears of a specific individual, أسر (Aser), a proper noun. The choice of the adjective “captivating” removes the reference to the person and imposes a subjective tone that may not be present or intended in the original, indicating a loss of contextual fidelity. Additionally, this results in a referential meaning error, as the referent “أسر” is completely omitted from the translation. By failing to preserve the reference to a specific individual, the TT generalizes and abstracts the personal and emotionally charged expression in the source caption. A more accurate and semantically faithful translation would be “Aser's tears”, which maintains the ST's grammatical structure, contextual tone, and referential clarity.

- 2- ST: هدية أسر لناي... فستان أزرق على مقياس حياة

TT: A **captivating** gift for Nai... A blue dress in the size of life <sup>[33]</sup>

The translation demonstrates several semantic translation errors. Firstly, the phrase “هدية أسر” is a possessive construction that grammatically indicates possession, meaning “Aser's gift.” This grammatical structure is not preserved in the TT, where the message was rendered as a vague and stylistically altered phrase, “a captivating gift”, thereby omitting the syntactic and possessive relationship, constituting a grammatical meaning error. Secondly, the translation reflects a contextual meaning error, as the proper names أسر (Aser) and حياة (Hayat) are misinterpreted. Specifically, حياة is incorrectly translated as the abstract noun life, rather than being maintained as a proper name referring to a person. This not only distorts the intended meaning but also erases the personalized elements of the original message. Finally, this misrepresentation leads to referential meaning errors, as the referents of “أسر” and “حياة” were misconstrued in the translation. In the ST, على مقياس حياة clearly refers to a dress that matches Hayat's size; however, in the TT, this reference

is inaccurately rendered as “in the size of life”, an abstract and semantically incoherent expression. To maintain semantic accuracy and communicative clarity, the phrase should be translated as “Aser’s gift for Nai: a blue dress in Hayat’s size” preserving the original referents, grammatical relationships, and contextual dimensions.

3- ST: اسم وملاح أسر ولكن روح مجد لازالت محبوسة بداخله

TT: The name of **captivation** features, but the soul of Majd is still imprisoned inside it<sup>[33]</sup>

The translation of “اسم وملاح أسر ولكن روح مجد لازالت محبوسة بداخله” into “The name of captivation features, but the soul of Majd is still imprisoned inside it” reveals multiple semantic errors. First, a grammatical meaning error is evident in the phrase “The name of captivation features.” The source phrase “اسم وملاح أسر” is a compound subject meaning “the name and features of Aser.” However, the TT renders this into a grammatically incoherent expression where “captivation” is treated as a possessive noun modifying “features”, losing the correct possessive relationship with the proper name أسر (Aser). This mistranslation alters the syntactic structure of the ST and introduces grammatical ambiguity. Secondly, a contextual meaning error arises from treating the proper noun أسر (Aser), which is mistranslated not as a name but as the abstract noun “captivation.” This misreading undermines the emotional and narrative context of the sentence by eliminating personal identity and replacing it with a generic concept; thus, distorting the intended meaning. Finally, a referential meaning error is committed in the same phrase, as أسر is a specific individual being referred to, not an abstract quality. The translation fails to preserve this referent, causing confusion for the target reader. The second part of the sentence, “but the soul of Majd is still imprisoned inside it”, is relatively closer in structure and reference to the original; however, the referent of “it” remains ambiguous and does not clearly point back to Aser in the absence of his name. While the name مجد (Majd) is preserved correctly, the pronoun “it” lacks sufficient clarity and may result in referential ambiguity. A more semantically accurate translation would be: “The name and features are Aser’s, but the soul of Majd is still imprisoned within him.”

4- ST: المهم إنك رجعت، دموع أم مجد

TT: The important thing is that you are back, tears or Majd<sup>[33]</sup>

The translation of “المهم إنك رجعت، دموع أم مجد” as “The important thing is that you are back, tears or Majd” presents several semantic translation errors. From a grammatical meaning perspective, the clause “دموع أم مجد” is mistranslated as “Tears or Majd”, which disrupts the intended syntactic relationship in the ST. In Arabic, “دموع أم مجد” clearly means “The tears of Majd’s mother”, a possessive construction that denotes both ownership and emotional association. The TT, however, uses “or”—a coordinating conjunction that is grammatically incorrect in this context, producing a fragmented and confusing structure. This misrepresentation is

considered a grammatical error as it misconstrues the possessive relation central to the meaning. The translation also contains an error in contextual meaning. The ST phrase shows a strong emotional context by highlighting a character’s return and the emotional response of Majd’s mother. The TT, however, disrupts this emotional meaning by offering an ambiguous expression—“Tears or Majd”—which neither conveys the mother’s grief nor maintains the emotional tone of the original. Additionally, the TT commits a referential meaning error by failing to preserve the referent أم مجد (Majd’s mother). This phrase is not a general reference but a specific identifier that situates the character within a relational framework (i.e., as the mother of Majd). The omission of this familial reference erases key contextual information and reduces the translated message’s interpretive coherence. A more accurate and semantically faithful translation would be: “What matters is that you’re back... the tears of Majd’s mother”.

5- ST: صعب أسر يفتح قلبه غير لحياة

TT: **Saab captivating** opens his heart other than **the life of**<sup>[33]</sup>

The translation of “صعب أسر يفتح قلبه غير لحياة” as “Saab captivating opens hid heart other than the life of” exhibits significant semantic translation errors across all three categories. The grammatical meaning of the source sentence is distorted at multiple levels. The Arabic phrase expresses a complex idea: “It is hard for Aser to open up his heart... to anyone other than Hayat.” However, the TT begins with “Saab”—a transliteration of the Arabic word صعب (hard or difficult)—which is not a recognized English word; thus, making the sentence incomprehensible to target readers. The second clause, “other than the life of”, is incomplete and ungrammatical, lacking a proper object and failing to convey the intended contrastive structure. These elements indicate a breakdown in grammatical coherence. From a contextual meaning perspective, the phrase misrepresents the emotional and narrative context of the original. The ST is a deeply personal and emotional statement about Aser’s emotional vulnerability and his exclusive affection for Hayat. Instead of conveying the exclusivity of Aser’s emotional openness toward Hayat, the TT transfers a confusing and awkward sequence of disconnected words. Moreover, translating أسر (Aser) as “captivating” and حياة (Hayat) as “life” abstracts the proper names into general concepts, stripping the sentence of its relational depth and situational context. This leads to a serious referential meaning error, as the proper names Aser and Hayat, central figures in the source context, are misrepresented. أسر is mistranslated as “captivating” and حياة is rendered as “life”, both of which result in the loss of referents necessary for TT’s clarity. A more accurate and semantically faithful translation would be: “It’s hard for Aser to open up his heart to anyone but Hayat”.

6- ST: ضاع مجد بسبب هذا الخاتم

TT: **Glory** was lost because of this ring <sup>[33]</sup>

The translation of “ضاع مجد بسبب هذا الخاتم” as “Glory was lost because of this ring” contains critical semantic translation errors. From a referential meaning perspective, the primary issue lies in the mistranslation of the proper noun مجد (Majd). Rather than recognizing مجد as a personal name, the translator renders it as the common noun “glory,” which misleads the audience and distorts the intended referent. In the ST, مجد refers to a specific individual named Majd, and the phrase ضاع مجد means “Majd was lost”. Misreading this proper name as a generic term for “glory” undermines the clarity of the reference, resulting in a significant loss of meaning. This error also results in a contextual meaning error. In the ST, the loss described is directly attributed to an individual, with the misfortune surrounding Majd as a person. The phrase بسبب هذا الخاتم (because of this ring) relates the loss to a particular object. By translating مجد as “glory”, the TT changes the focus from a personal loss to the abstract concept of “glory” being lost due to the ring. This shifts the focus from an individual’s misfortune to a more general metaphorical statement about the loss of an abstract quality. As a result, the emotional context of the ST, which revolves around the individual Majd, is significantly altered, and the interpersonal implications of the loss are lost in translation. Furthermore, the TT introduces a subtle grammatical meaning issue. The phrase “Glory was lost” creates a passive construction that fits the general meaning of the ST but fails to preserve the active agency implied in the original Arabic. Using “glory” as the subject of the sentence in English abstracts the action, whereas in Arabic, the subject of the sentence is a person, مجد (Majd), and the loss is directly attributed to him. While this is a subtle grammatical issue, it reflects the underlying semantic change that occurs when the proper name is replaced with an abstract noun. A semantically accurate and contextually faithful translation would be: “Majd was lost because of this ring”.

7- ST: حتى زواج مجد وحياة كان كذبة

TT: **Until** the marriage of Majd and Hayat was a lie <sup>[33]</sup>

The translation of “حتى زواج مجد وحياة كان كذبة” as “Until the marriage of Majd and Hayat was a lie” contains semantic translation errors that fall under all three categories. From a referential perspective, the proper nouns مجد (Majd) and حياة (Hayat) are correctly preserved, indicating that character names are recognized. However, the adverb of degree, حتى is mistranslated. In the source context, حتى means “even,” not “until.” The phrase حتى زواج مجد وحياة should be understood as “even the marriage of Majd and Hayat,” expressing surprise or emphasis that something as significant as their marriage was also part of the deception. Rendering حتى as “until” shifts the reference and misleads the audience into interpreting the sentence as describing a time limit, which is not intended in the ST. This misinterpretation also results in a contextual meaning error. The ST emphasizes the degree

of deception, with حتى conveying that even this authentic event was a lie. The TT, however, changes the focus to a chronological structure (until), suggesting a narrative progression. This shift undermines the emotional function of the original statement. Furthermore, the TT exhibits a grammatical meaning issue. The phrase “Until the marriage of Majd and Hayat was a lie” is grammatically awkward and incomplete in English. The temporal connector “until” typically requires a main clause to follow (e.g., “Until the marriage of Majd and Hayat, everything seemed fine”), but in this case, it is followed directly by “was a lie,” creating a syntactically incorrect construction. The incorrect use of the temporal connector results in a sentence that lacks clarity and coherence. A semantically accurate and contextually faithful translation would be: “Even the marriage of Majd and Hayat was a lie”.

8- ST: ناي عم تمزح ولا جد؟

TT: where are you joking or **grandfather**? <sup>[33]</sup>

The translation of “ناي عم تمزح ولا جد؟” as “where are you joking or grandfather?” contains significant semantic translation errors. From a referential meaning perspective, the primary error stems from the mistranslation of the word جد (jad). In colloquial Levantine Arabic, جد commonly functions as an adverb meaning “serious(ly)” as in “ولا جد؟” or “are you serious?”. However, جد was interpreted as a noun “grandfather,” leading to a nonsensical phrase in English. This misreading of the referent results in a total failure to convey the speaker’s intended meaning and introduces an unrelated concept (a family member) that is absent from the ST. The translation also influenced the contextual meaning. In the ST, the speaker, addressing Nai, is expressing doubt or surprise, asking: “Nai, are you joking or serious?” The TT, by mistranslating جد and omitting key elements ناي of the utterance, fails to capture the pragmatic function of the question and instead produces a phrase with no logical context, “where are you joking or grandfather?”. In terms of grammatical meaning, the TT also presents structural problems. The phrase “where are you joking or grandfather?” is grammatically incoherent in English. The use of “where” suggests a question about location, which is not present in the ST. This likely stems from a misinterpretation of “عم” (a colloquial present-tense expression in Arabic). In reality, “عم” is used to indicate continuous or current action (similar to the English present continuous tense) and should have been rendered as “are you” or “are you currently.” Thus, the TT exhibits a grammatical error in the verb tense and interrogative form, which further undermines its clarity. A semantically accurate and contextually faithful translation would be: “Is Nai joking or serious?”

9- ST: كل من خان، يمحي من الوجدان

TT: **Both Khan**, erased from the conscience <sup>[33]</sup>

The translation of “كل من خان، يمحي من الوجدان” as “Both Khan, erased from the conscience” contains critical semantic translation errors. From a referential meaning standpoint, the

primary issue lies in the mistranslation of خان (Khan). In Arabic, خان is a verb meaning “to betray” or “to be unfaithful”, not a proper noun. However, خان was misinterpreted as a proper name, “Khan,” which misleads the audience and distorts the intended meaning. The phrase كل من خان means “Anyone who betrays” or “Whoever betrays”, where خان is the verb referring to the act of betrayal. The misinterpretation of خان as a name shifts the referent from a general action of betrayal to a specific person named “Khan”, which leads to a misunderstanding of the original message. This error also results in a contextual meaning error. In the ST, the phrase “كل من خان” (Anyone who betrays) implies a general statement about the consequences of betrayal, suggesting that those who betray will be erased from the heart or memory. By translating خان as “Khan,” the TT loses the general, metaphorical meaning and incorrectly suggests that a person named Khan is the subject of the statement. This changes the contextual focus from a condemnation of betrayal to an ambiguous reference to a specific person or name, altering the intended message of the ST. Furthermore, the TT introduces a grammatical meaning issue. In the ST, “يمحى من الوجدان” (is erased from the heart) is a passive construction that fits grammatically in Arabic, conveying that the betrayal leads to the erasure of an individual from the heart or memory. The TT, however, disregards the passive structure and introduces confusion by misapplying خان as a proper name, which affects the clarity of the subject and object and creates grammatical awkwardness in the TT. A semantically accurate and contextually faithful translation would be: “Anyone who betrays is erased from the heart”.

10- ST: الضربة وجعها ثانية، بس الخيانة وجعها عمر

TT: The hit wrenched it **again**, but betrayal, **Omar hurt her** <sup>[33]</sup>

The translation of “الضربة وجعها ثانية، بس الخيانة وجعها عمر” as “The hit wrenched it again, but betrayal, Omar hurt her” contains critical semantic translation errors. From a referential meaning standpoint, the primary issue lies in the mistranslation of عمر (Omar). In the ST, عمر is not a proper noun but a figurative term that refers to “a lifetime” or “for life.” The phrase وجعها عمر means “its pain lasts a lifetime,” where عمر is used metaphorically to signify the enduring nature of betrayal’s pain. However, in the TT, عمر is mistranslated as a person’s name, “Omar,” which misleads the audience and distorts the intended meaning. This also results in a contextual meaning error. The ST contrasts the pain of a physical punch (الضربة وجعها ثانية) “The punch hurts for a second”) with the much longer-lasting emotional pain caused by betrayal (الخيانة وجعها عمر – “Betrayal hurts a lifetime”). Using عمر in the ST is crucial in conveying that the pain of betrayal is enduring and lasts a lifetime. However, the metaphorical concept of lasting pain is lost by translating عمر as a person’s name in the TT. The sentence in the TT, “Betrayal, Omar hurt her,” implies that a person named Omar is responsible

for the pain; thus, it diverts from the intended statement about the emotional consequences of betrayal. Furthermore, the TT introduces a grammatical meaning issue. The phrase “Omar hurt her” in the TT makes a grammatically awkward construction, as it introduces an individual, “Omar,” who does not appear in the ST. The original Arabic structure clearly compares the brief physical pain of a punch with the long-lasting emotional pain of betrayal. In the TT, however, the sentence shifts to focus on “Omar,” introducing confusion about who is causing the pain and disrupting the grammatical structure that conveys the contrast between temporary and enduring pain. A semantically accurate and contextually faithful translation would be: “A punch hurts for a second, but betrayal hurts a lifetime.”

## 4.2. Shahid Account

1- ST: جاد حفيدك

TT: Grandfather of your grandson <sup>[34]</sup>

The translation of the Arabic phrase “جاد حفيدك” into “Grandfather of your grandson” exhibits multiple semantic translation errors that distort the intended meaning. Structurally, the ST is a nominal phrase identifying a person, “Jad”, as the speaker’s or addressee’s grandson. The TT, however, introduces a reversed generational relationship by rendering the phrase as “Grandfather of your grandson,” which is both illogical and syntactically inaccurate. This misrepresentation constitutes a grammatical meaning error, as the TT fails to preserve the correct syntactic roles and introduces a structurally incoherent relationship. In terms of contextual meaning, the ST likely functions as a narrative identifier, referring to “Jad” as someone’s grandson. By altering the familial relation, the TT distorts the context and undermines the intended narrative significance. The contextual cue embedded in the original, anchored in kinship and generational positioning, is lost in translation. Additionally, the TT commits a referential meaning error by omitting the proper noun “Jad.” The ST clearly refers to a specific individual, while the TT abstracts and generalizes the phrase, removing the direct reference to the named person and thereby weakening the personal and narrative impact of the original. A more accurate and semantically faithful translation would be: “Jad, your grandson”.

2- ST: إذا بتحب أخوك بتموت كرماله

TT: If you love your brother, you will die as his caramel <sup>[34]</sup>

The translation of the Arabic phrase “إذا بتحب أخوك بتموت كرماله” into “If you love your brother, you will die as his caramel” introduces multiple semantic. Grammatically, the ST is a conditional sentence composed of two clauses: “If you love your brother” and “you would die for him”. The TT maintains the conditional structure but renders the second clause nonsensically by translating “كرماله” (for him) as “his caramel,” a lexical mistranslation. This results in a grammatical meaning error, as the verb phrase “die as his caramel” is ungrammatical and semantically incoherent in English. From



a contextual standpoint, the ST uses emotionally charged colloquial language to express deep loyalty or self-sacrifice for one's brother, "you would die for him." The TT, however, substitutes the intended idiomatic expression with a meaningless metaphor involving "caramel," stripping the sentence of its emotional force. This constitutes a contextual meaning error, as the original pragmatic intent is entirely lost. Furthermore, the TT exhibits a referential meaning error in its handling of the prepositional phrase "كرماله." In Arabic, this term means "for him" and directly refers to the previously mentioned "brother." By mistranslating it as "his caramel," the TT introduces a lexical item that has no relation to the context. A more accurate and semantically faithful translation would be: "If you love your brother, you will die for him"

3- ST: أنا ما بدّي هيك عيلة

TT: I don't want you to be a family <sup>[34]</sup>

The translation of the Arabic sentence "أنا ما بدّي هيك عيلة" into "I don't want you to be a family" contains multiple semantic errors that distort the intended message. Grammatically, the ST is a negative declarative sentence that translates more accurately as "I don't want a family like this." The TT, however, introduces a second-person subject "you" that is absent in the ST, altering the sentence structure and introducing a new agent. This constitutes a grammatical meaning error. From a contextual meaning perspective, the ST conveys disappointment or emotional rejection of a particular kind of family dynamic. By changing the focus to "you" and turning the statement into an abstract condition "you to be a family", the TT misrepresents the emotional tone and communicative intent. In terms of referential meaning, the ST implies a specific referent, "this family" or "a family like this", anchored in the immediate narrative situation. The TT omits this referential element, substituting it with a general and unrelated phrase "you to be a family", thereby severing the link to the intended referent. A more accurate and semantically faithful translation would be: "I don't want a family like this".

4- ST: بدّي ولادي يكونوا حدّي

TT: I want my children to be my limit <sup>[34]</sup>

The translation of the Arabic phrase "بدّي ولادي يكونوا حدّي" into "I want my children to be my limit" contains severe semantic distortions. Grammatically, the ST is a straightforward colloquial expression meaning "I want my children to be by my side" or "close to me." The TT, however, mistranslates the prepositional phrase "حدّي" as "my limit," confusing the spatial or emotional proximity expressed in Arabic with an abstract boundary. This results in a grammatical meaning error, as the syntactic relationship between the subject and complement is misrepresented, and the lexical choice fails to reflect the original structure. From a contextual meaning perspective, the ST communicates emotional closeness as it expresses a parent's desire for their children to remain near, physically or emotionally. The TT, by contrast, introduces an illogical and contextually incoherent phrase "be my limit".

This marks a clear contextual meaning error, as the emotional register and pragmatic intent of the original are lost. In terms of referential meaning, the ST includes implicit reference to proximity, directed at a parent-child relationship. By interpreting "حدّي" literally as "limit," the TT removes the relational and situational reference embedded in the expression, thereby committing a referential meaning error through incorrect mapping of spatial deixis. A more accurate and semantically faithful translation would be: "I want my children to be by my side".

5- ST: أنا قدرك، مو راغب

TT: I appreciate you in life, I don't want <sup>[34]</sup>

The translation of the Arabic phrase "أنا قدرك، مو راغب" into "I appreciate you in life, I don't want" contains multiple semantic errors that severely distort the intended meaning. Grammatically, the ST consists of two nominal clauses: "أنا قدرك" (I am your destiny) and "مو راغب" (not Raghib). The TT mistranslates the first clause as "I appreciate you in life," introducing an evaluative verb not found in the original. This is due to confusion with the phonetically similar verb "أقدرك" which means (I appreciate you). The second clause is translated as "I don't want," which is both semantically incomplete and grammatically fragmented, lacking an object or referential anchor. These issues reflect a grammatical meaning error, as the translation alters the syntactic structure and fails to deliver a coherent message. From a contextual meaning standpoint, The name "راغب" functions as a proper noun and the phrase implies: "I am your destiny, not Raghib." This expression suggests emotional inevitability while denying romantic alignment with another person. The TT's rendering misses both the rhetorical contrast and the underlying emotional dynamics, resulting in a contextual meaning error. In terms of referential meaning, the original sentence clearly includes named referents, "Raghib" and "I". The TT omits the name Raghib, which is essential to the contrastive structure, and introduces a vague and irrelevant notion of "appreciation." This omission leads to a referential meaning error, as the core personal and narrative references are lost. A more accurate and semantically faithful translation would be: "I'm your destiny, not Raghib".

## 5. Discussion

Based on the analysis of fifteen colloquial Arabic captions translated into English using TikTok's MT feature and evaluated through the lens of Sayogie's <sup>[29]</sup> classification of semantic errors, it can be inferred that all three types of semantic errors—grammatical meaning, contextual meaning, and referential meaning—were present across the dataset. This recurring pattern of semantic inaccuracies indicates that MT systems, particularly in the context of colloquial language on social media platforms, remain significantly limited in their ability to preserve meaning.

**Table 1** presents the frequency of semantic errors identified across the 15 caption pairs analysed, 10 from MBC1 account and 5 from Shahid TikTok account. Each caption contained at least one grammatical, contextual, and referential error, resulting in a total of 53 semantic errors. These captions, drawn from 2025 TikTok uploads, illustrate that TikTok’s Arabic-English MT errors are both systematic and persistent across different sources. The find-

ings suggest that TikTok’s MT is still far from achieving the accuracy required for reliable cross-linguistic communication, especially when dealing with informal, idiomatic, and culturally embedded language such as colloquial Arabic. This highlights the importance of continued development in MT technologies and the need for greater attention to linguistic and contextual sensitivity in automated translation tools.

**Table 1.** Semantic error frequency.

Caption Source	Caption Number	Grammatical Errors	Contextual Errors	Referential Errors
MBC1	1	1	1	1
MBC1	2	1	1	2
MBC1	3	1	1	2
MBC1	4	1	1	1
MBC1	5	1	1	2
MBC1	6	1	1	1
MBC1	7	1	1	1
MBC1	8	1	1	2
MBC1	9	1	1	1
MBC1	10	1	1	1
Shahid	1	1	1	1
Shahid	2	1	1	1
Shahid	3	1	1	1
Shahid	4	1	1	1
Shahid	5	1	1	1

These findings align with the broader concerns raised in recent machine translation research. For instance, Läubli et al. <sup>[21]</sup> argue that sentence-level evaluations are often inadequate, as they fail to account for the communicative intent and coherence of the text as a whole. Instead, they advocate for document-level and meaning-focused assessments that more accurately reflect the intended message and pragmatic function of the source content. This perspective aligns with the observed shortcomings in TikTok’s translations, where sentence-level fluency may mask deeper semantic distortions, particularly in metaphorical or culturally nuanced expressions. Furthermore, Cao et al. <sup>[30]</sup> emphasize the importance of semantic similarity, particularly in the treatment of logical relations and quantifiers, in evaluating the adequacy of MT output. Their proposed SemMT framework demonstrates that surface-level accuracy can be misleading, as translations may appear fluent or structurally sound while failing to preserve core semantic relations. Similarly, in the present study, although some captions exhibited grammatical coherence, deeper analysis revealed significant distortions of meaning, especially in culturally embedded phrases and figurative expressions. The prevalence of contextual and referential errors

supports Cao et al.’s argument that current evaluation practices are insufficient and that robust semantic-based assessment models are essential for identifying meaning-level failures in translation. These findings underscore the need to move beyond superficial fluency metrics and adopt more semantically grounded evaluation methods when assessing MT outputs in culturally and linguistically complex contexts.

To convert the findings into practical recommendations, several improvements to TikTok’s captioning and translation system are recommended. First, incorporating dialectal corpora into the training data can enhance the model’s ability to recognize and accurately translate region-specific vocabulary and syntax, especially in Arabic. Second, the system could benefit from a confidence scoring feature that flags low-certainty segments for user review, helping audiences identify potentially unreliable translations. Third, integrating a lightweight post-editing interface—such as a crowd-editing or feedback option—would allow native speakers to suggest corrections, thereby improving semantic accuracy and cultural appropriateness in real time. These enhancements could contribute significantly to mitigating misinterpretation and improving user trust in MT-generated captions.

## 6. Conclusions

This research evaluated the semantic accuracy of TikTok's MT, specifically focusing on English translations of colloquial Arabic captions posted by MBC1 and Shahid TikTok accounts. Using Sayogie's <sup>[29]</sup> classification of semantic errors- grammatical, contextual, and referential- the study revealed that all three types of errors were present across the sample. The prevalence of these errors highlights the ongoing limitations of MT systems in accurately conveying the intended meaning of colloquial, culturally embedded language. The system's reliance on literal translation strategies contributed significantly to meaning distortion, indicating that current MT models still struggle with the complexities of informal speech, pragmatic cues, and context-dependent expressions common in colloquial Arabic.

Consequently, the growing reliance on MT in such settings highlights the need to critically evaluate its effectiveness in conveying meaning across languages and cultures. As MT becomes increasingly integrated into social media platforms, there is a high risk that mistranslations could lead to miscommunication, cultural misunderstanding, or loss of intended message, especially in user-generated content where informal and idiomatic expressions dominate.

Widespread translation errors on social media platforms such as TikTok raises several ethical concerns that merit closer attention. Mistranslations can lead to significant miscommunication, distortion of cultural meaning, or even the spread of misinformation, particularly when dealing with emotionally charged or culturally specific content. This becomes especially problematic in multilingual, multicultural digital spaces where users may rely on machine-translated captions to form impressions about individuals, communities, or social dynamics. Ethical implications include the erasure of cultural references, the potential reinforcement of stereotypes, and the undermining of trust in digital communication. Platforms and developers must acknowledge these risks and take proactive measures, such as incorporating ethical evaluation criteria, enabling user reporting for mistranslations, and promoting transparency in algorithmic decision-making, to mitigate the harm caused by automated translation errors.

Despite providing valuable insights, the dataset was limited to fifteen captions from two TikTok accounts, which may not fully represent the broader spectrum of colloquial Arabic or other regional dialects. The analysis was also restricted to semantic errors, excluding other potential dimensions of translation quality such as syntactic, pragmatic, or cultural accuracy. Expanding the dataset to include a wider range of dialects and content from multiple social media platforms would enhance the generalizability of the findings. Comparative analyses between human and MT of the same content would also provide deeper insight into the specific challenges MT systems face when dealing with informal and culturally embedded language.

Although metadata such as upload year and account type were recorded, the analysis did not explore their impact due to the limited scope and qualitative focus of the study. Future studies could investigate how additional variables, such as view counts, speaker speed, or the presence of hashtags, may correlate with the frequency and nature of translation errors. Furthermore, adopting interdisciplinary approaches that draw on linguistic theory, computational modelling and sociocultural analysis would contribute to the development of more context-aware and semantically accurate translation models, particularly for low-resource, dialect-rich languages like Arabic. Such efforts are essential for ensuring the accessibility, accuracy, and cultural sensitivity of automated translations in increasingly globalized digital environments.

## Funding

This research article is not funded.

## Institutional Review Board Statement

This article does not require any ethical approval as it does not contain any studies with human participants performed by the author.

## Informed Consent Statement

This article does not contain any studies performed by the author with human participants.

## Data Availability Statement

Data are available online in the MBC1 TikTok account via <https://www.tiktok.com/@mbc> and Shahid TikTok account via <https://www.tiktok.com/@shahid>.

## Acknowledgments

The author gratefully acknowledges the institutional support provided by King Saud University.

## Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

## Abbreviations

Machine Translation (MT)  
Middle East Broadcasting Center 1 (MBC1)  
Middle East and North Africa (MENA)  
Modern Standard Arabic (MSA)  
Source Text (ST)  
Target Text (TT)  
Source Language (SL)

Target Language (TL)

## References

- [1] Utami, N.M.V., Irwandika, G., 2021. Semantic errors in the translation of the Hindu's Instagram account. *English Language Teaching, Literature, and Translation (ELTLT)* 2021. 10(1), 84–88.
- [2] Bowker, L., 2020. Chinese speakers' use of machine translation as an aid for scholarly writing in English: A review of the literature and a report on a pilot workshop on machine translation literacy. *Asia Pacific Translation and Intercultural Studies*. 7(3), 288–298. DOI: <https://doi.org/10.1080/23306343.2020.1805843>
- [3] Kumar, M.A., Premjith, B., Singh, S., et al., 2019. An overview of the shared task on machine translation in Indian languages (MTIL)- 2017. *Journal of Intelligent Systems*. 28(3), 455–464. DOI: <https://doi.org/10.1515/jisys-2018-0024>
- [4] Kunchukuttan, A., Bhattacharyya, P., 2022. Machine translation and transliteration involving related and low-resource languages. CNC Press: Boca Raton, FL, USA.
- [5] Larson, M.L., 1998. *Meaning-based translation: A guide to cross-language equivalence*. University Press of America: New York, NY, USA.
- [6] Sari, D.M., 2019. An error analysis on students' translation text. *Eralingua: Jurnal Pendidikan Bahasa Asing dan Sastra*. 3(2), 65–74. DOI: <https://doi.org/10.26858/eralingua.v3i2.8658>
- [7] Lee, S.M., 2020. The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*. 33(3), 157–175. DOI: <https://doi.org/10.1080/09588221.2018.1553186>
- [8] Khoong, E.C., Rodriguez, J.A., 2022. A research agenda for using machine translation in clinical medicine. *Journal of General Internal Medicine*. 37(5), 1275–1277. DOI: <https://doi.org/10.1007/s11606-021-07164-y>
- [9] Gupta, V., Thakral, K.S., 2019. Divergence Issues in machine translation for English-Punjabi language. *Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA)*; Indore, India; 14–16 February 2019.
- [10] Liu, S., Sun, Y., Wang, L., 2021. Recent advances in dialogue machine translation. *Information*. 12(11), 1–2. DOI: <https://doi.org/10.3390/info12110481>
- [11] Hoi, H.T., 2020. Machine translation and its impact in our modern society. *International Journal of Scientific & Technology Research*. 9(4), 1918–1921.
- [12] Ying, C., Shuyu, Y., Jing, L., et al., 2021. Errors of machine translation of terminology in the patent text from English into Chinese. *ASP Transactions on Computers*. 1(1), 12–17. DOI: <https://doi.org/10.52810/TC.2021.100022>
- [13] Kostadinova, V., Yáñez-Bouza, N., Dreschler, G., et al., 2019. *I English Language. The Year's Work in English Studies*. 98(1), 1–166.
- [14] Popowich, F., 1996. A chart generator for Shake and Bake Machine Translation. In: *Advances in Artificial Intelligence: 11th Biennial Conference of the Canadian Society for Computational Studies of Intelligence, AI'96 Toronto, Ontario, Canada, May 21–24, 1996 Proceedings*, 11, 97–108. Springer: Berlin/Heidelberg, Germany.
- [15] Rozovskaya, A., Sproat, R., Benmamoun, E., 2006. Challenges in processing colloquial Arabic. *Proceedings of the International Conference on the Challenge of Arabic for NLP/MT*; London, UK; 23 October 2006. pp. 4–14.
- [16] Mirhashemi, Z., Gholami, M., Bahri, H., 2024. A Comparative Study on Translation of Persian Colloquialism into English by ChatGPT and Other Translation Platforms. *Iranian Journal of Translation Studies*. 22(87). Available from: <https://journal.translationstudies.ir/ts/article/view/1203> (cited 25 May 2025).
- [17] Sun, Z., Zemel, R., Xu, Y., 2022. Semantically informed slang interpretation. *arXiv preprint. arXiv:2205.00616*.
- [18] Fattah, B.O., Salih, S.M., 2023. Drawing a Demarcation Line between Two Overlapping Colloquial Elements: The Case of Idioms and Clichés. *Koya University Journal of Humanities and Social Sciences*. 6(1), 65–76.
- [19] Nasution, D.K., 2022. Machine translation in website localization: Assessing its translation quality for language learning. *AL-ISHLAH: Jurnal Pendidikan*. 14(2), 1879–1886. DOI: <https://doi.org/10.35445/alishlah.v14i2.1308>
- [20] Läubli, S., Sennrich, R., Volk, M., 2018. Has machine translation achieved human parity? A case for document-level evaluation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*; Brussels, Belgium; 31 October–4 November 2018. Association for Computational Linguistics: Brussels, Belgium. pp. 4791–4796. DOI: <http://dx.doi.org/10.18653/v1/D18-1512>
- [21] Läubli, S., Castilho, S., Neubig, G., et al., 2020. A set of recommendations for assessing human-machine parity in language translation. *Journal of Artificial Intelligence Research*. 67, 653–672. DOI: <https://doi.org/10.1613/jair.1.11371>
- [22] Putri, A.T., Setiajidi, H.H., 2021. Instagram translate and human translation in the English captions of Jokowi's account: An analysis of Koponen's error category. *English Language and Literature International Conference (ELLiC) Proceedings*. 4, 432–436.
- [23] Utami, N.M.V., Jayantini, I.G.A.S.R., Pratiwi, Y., 2021. Lexical analysis of semantic errors found in the translation of Joko Widodo's Instagram account. *English Language and Literature International Conference (ELLiC) Proceedings*. 4, 291–297.
- [24] Susanti, E., 2018. *Lexical errors produced by Instagram machine translation [Ph.D. Thesis]*. Universitas Islam Negeri Maulana Malik Ibrahim: Malang, Indonesia. Available from: <http://etheses.uin-malang.ac.id/id/eprint/14231> (cited 20 May 2025).
- [25] Tan, X., Chen, J., He, D., et al., 2019. Multilingual neural machine translation with language clustering. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*; Hong Kong, China; 3–7 November 2019. DOI: <https://doi.org/10.18653/v1/D19-1089>
- [26] Han, L., Jones, G.J., Smeaton, A.F., 2021. Translation quality assessment: A brief survey on manual and automatic methods. *Proceedings for the First Workshop on Modelling Translation: Translatology in the Digital Age*; Saarbrücken, Germany; 31 May 2021. pp. 15–33.
- [27] Fan, A., Bhosale, S., Schwenk, H., et al., 2021. Beyond



- English-centric multilingual machine translation. *The Journal of Machine Learning Research*. 22(1), 4839–4886.
- [28] Vieira, L.N., O'Hagan, M., O'Sullivan, C., 2021. Understanding the societal impacts of machine translation: A critical review of the literature on medical and legal use cases. *Information, Communication & Society*. 24(11), 1515–1532. DOI: <https://doi.org/10.1080/1369118X.2020.1776370>
- [29] Sayogie, F., 2014. *Teori dan praktikum penerjemahan*. Transpustaka: Tangerang Selatan, Indonesia.
- [30] Cao, J., Li, M., Li, Y., et al., 2022. SemMT: a semantic-based testing approach for machine translation systems. *ACM Transactions on Software Engineering and Methodology (TOSEM)*. 31(2), 1–36.
- [31] Utami, N.P.L.D., Utami, N.M.V., 2023. Unveiling Semantic Errors Found in Lexical Translations of Tasya Farasya's TikTok Account. *Lingua Cultura*. 17(2), 219–225.
- [32] Lambert, V.A., Lambert, C.E., 2012. Qualitative descriptive research: An acceptable design. *Pacific Rim International Journal of Nursing Research*. 16(4), 255.
- [33] MBC1 [@MBC1], 2025. TikTok. Available from: <https://www.tiktok.com/@mbc> (cited 10 May 2025).
- [34] Shahid [@shahid], 2025. TikTok. Available from: <https://www.tiktok.com/@shahid> (cited 10 June 2025).