Evaluating the Quality of Instagram's Machine Translation in Literary Texts

ABSTRACT

In response to the surge in social media platform users worldwide, numerous platforms including Instagram implemented a feature enabling automatic translation. This feature was designed to expand the scope of information dissemination and enhance cross-cultural communication efficiency. However, the accuracy of machine translation systems utilized by these platforms remains a point of concern. Therefore, this paper aims to scrutinize the reliability of Instagram's "See Translation" feature in the translation of literary texts from Arabic to English. This study aims to analyze translated captions through the identification, classification, and assignment of error types and penalty points, utilizing the MQM core typology. Additionally, the study aims to automatically calculate the Overall Quality Score using the ContentQuo platform, with the ultimate goal of exploring the potential of Instagram machine translation (MT) to produce high-quality translations. Furthermore, the study aims to determine whether Instagram MT can effectively convey the intended message within literary texts. A purposive selection of 30 captions was extracted from an Instagram corpus featuring literary content. The findings of the study indicated that Instagram machine translation demonstrated poor performance in the translation of 90% of the selected data, specifically in three key areas: accuracy, fluency, and style. Of the selected captions, 61 errors were identified, with 26 errors in fluency, 25 errors in accuracy, and 10 errors in style. These errors were found to negatively impact the overall quality of the translations, resulting in the failure to transfer the intended message conveyed within the source texts.

Key words: Arabic Translation; Literary Context; Machine Translation; MQM Analysis; Quality Evaluation

INTRODUCTION

Instagram is a popular social networking platform where users share updates in the form of audio and/or audio-visual elements as posts. These posts are often accompanied by textual details, referred to as captions. According to Dixon's (2023) report on Statista, Instagram is currently the fourth most popular social media platform, with 1.28 billion active users as of January 2022, and is projected to reach 1.44 billion monthly active users by 2025. To address the growing global use of Instagram, Meta (formerly Facebook), the parent company of Instagram, is developing new Machine Translation (MT) innovations and MT systems to facilitate cross-lingual communication among users from different countries. This is aimed at improving the effectiveness of interactions among global users regardless of their language backgrounds.

Although Machine Translation (MT) systems on social media platforms like Instagram provide instant translation results, concerns have been raised about their reliability and the quality of their output, particularly in an Arabic context. While several studies have evaluated Instagram MT output, focusing on Indonesian-English language pairs (Pudjiati & Fadilah, 2017; Susanti, 2018; Larassati et al., 2019; Cahyani et al., 2021), there is a significant gap in understanding the

Arabic-English language pair. As each language has its own unique characteristics, it is important to evaluate different language pairs to enhance the quality of machine translation. Translating Arabic to other languages can be challenging due to the language's unique linguistic and contextual characteristics, as noted by Alkhatib and Shaalan (2018). Thus, this study aims to address this gap by evaluating the quality of Instagram MT's translations of Arabic literary texts into English.

The study's significance lies in its evaluation of machine translation, an essential area of research that helps improve the performance of existing MT systems and understand how they function (Dorr et al., 2011). Regular evaluations of MT systems are crucial to monitor their quality and identify areas for improvement. However, there currently needs to be more standardization for MT quality assessment and error analysis (Trigueros, 2021). Therefore, this study's findings can serve as a valuable reference for the translation technology field in general and for MT evaluation development, translation error analysis methodology, and computational linguistics. Additionally, this study will offer practical benefits by providing insights for Meta developers to improve the algorithms and linguistic architecture of their MT models, as well as for Instagram users to assess the quality and reliability of the instant translations provided by the platform. Given the doubts about the quality of Instagram's MT feature in the Arabic context, this study aims to identify its shortcomings and weaknesses, thereby helping Meta developers improve the feature. The study aims to achieve the following objectives.

This study aims to evaluate the quality of Instagram's Neural Machine Translation (NMT) system by using an analytical error-category approach, which utilizes the MQM system that includes structured translation specifications, an error typology, and a scoring system integrated with the ContentQuo platform. The goal is to assess the NMT system's effectiveness in translating literary texts. Additionally, the study will examine the NMT system's potential to convey the expressive function of literary texts, as defined by Nord's translation function theory.

LITERATURE REVIEW

Machine Translation Overview

Machine Translation is an interdisciplinary paradigm that involves different fields, including Natural Language Processing (NLP), where it is focused on developing and optimizing computer-based translation systems (Ameur et al., 2020). It is also considered a branch of Computational Linguistics, which investigates the use of computer software to translate text from one natural language to another (Arnold et al., 1994; Sipayung et al., 2021). Due to its multidimensional nature, MT is more complex than Human Translation and is continuously evolving with the development of technology. MT is the process of translating text from one language into one or more other languages using computer-based systems and tools, with or without human intervention.

The increasing demand on translation due to the economic globalization exceeded the human capability of handling all the tasks required to be translated led to introducing automatic translation systems resulting in an enormous change in the interrelated fields. MT systems, according to their computational architecture (Chéragui, 2012), are classified into four approaches: Rule-based MT (RBMT) approach, Corpus-based MT (CBMT) approach, Hybrid MT approach, and Neural MT approach (Trigueros, 2022). The first approach was Rule-based MT, which used two linguistic sub-approaches: transfer and interlingua (Chéragui, 2012) that relied on

monolingual and bilingual dictionaries, grammar and transfer rules for generating translations (Espana-Bonet & Costa-jussa, 2016; Castilho et al., 2017; Trigueros, 2022). Later on, Corpus-based MT was introduced as an alternative approach for MT in order to overcome the shortcomings of RBMT (Chéragui, 2012), and was the first approach of data-driven methods that used sophisticated algorithms and mathematical models to automatically learn the translation process from data (Ameur et al., 2020). CBMT used monolingual and bilingual corpora of parallel texts in the translating process (Hutchins, 1995). This approach was divided into two systems: Statistical MT system (SMT) and Example-based MT system (EBMT). The advantage of the system is that it requires less human effort to be automatically trained as well as its solid performance when selecting terms (Hutchins, 2007; Koehn, 2009; Trigueros, 2022). However, it sometimes outputs bad quality translations that are ill-structured or grammatically incorrect attributed to the difficulty in reaching corpora of specific domains or language pairs (Habash et al., 2009; Espana-Bonet & Costa-jussa, 2016; Trigueros, 2022). Nevertheless, corpus-based systems dominated the field for a while as many MT developers adopted the approach to their MT systems, including Google Translate, Facebook, and Instagram. The hybrid MT approach combines both Rule-based MT and statistical MT systems, resulting in a solution that overcomes the deficiencies of each system and produces high-quality translations with a high level of precision (Thurmair, 2009; Hunsicker et al., 2012; Tambouratzis et al., 2014; Trigueros, 2022).

Moreover, most recently, a new data-driven MT approach, called Neural Machine Translation (NMT) is developed with a different mechanism; it works in building and training a single large neural network that reads a sentence and outputs correct translations (Bahdanau et al., 2014; Trigueros, 2022). This system is based on the encoder-decoder model in which the encoder reads the input and encodes it into a fixed length vector while the decoder produces the translation output from the encoder vector (Cho et al., 2014; Bahdanau et al., 2014; Trigueros, 2022). NMT represents the latest development of MT systems, which has become the dominant paradigm that is currently applied in machine translation filed (Ragni & Vieira, 2021; Trigueros, 2022). Furthermore, Trigueros (2022) pointed out that the architecture of NMT is characterized by some advantageous properties that prior MT systems does not own. For instance, it uses less components and processing steps, and it requires less memory than SMT. Moreover, it allows the use of human and data resources more efficiently than RBMT (Cho et al., 2014; Bentivogli et al., 2016; Trigueros, 2022). Furthermore, the findings revealed that NMT output contained fewer overall errors compared to SMT at the accuracy and fluency levels (Wu et al., 2016; Castilho et al., 2017; Moorkens, 2018; Ragni & Vieira, 2021). Such characteristics have pushed Meta, along with many other major companies, such as Google, Systran, and Microsoft (Ameur et al. (2020); Trigueros, 2022) to shift from SMT and RBMT approaches to Neural MT approach. In 2016, six years since it was first launched, Instagram has added an automatic translation system as a feature, called "See Translation" (Veronika, 2017; Arvianti, 2018, Larassati, 2019). Since that, Meta has been continuously striving to come up with new state-of-the art models with the aim of improving the quality of the translation results for better across-nation communication.

NMT of Instagram

In 2017, Meta announced its shift from phrase-based statistical machine translation to neural machine translation (Mannes, 2017), resulting in more accurate and fluent translations (Pino et al., 2017). In 2020, Meta introduced a new neural machine translation model, the multilingual machine translation (M2M-100), which automatically translates between any pair of 100 languages, including 2,200 language directions, without relying on English as an intermediary

source. The M2M-100 model aims to improve translation quality for low-resource languages (Bhattacharyya, 2022). Additionally, Meta developed a single artificial intelligence-based model, the No Language Left Behind (NLLB-200), which translates 200 languages, including those that needed to be adequately addressed by machine translation tools in Instagram. The NLLB-200 model aims to improve the quality of machine translations and facilitate communication worldwide. Meta evaluated the NLLB-200 model using the BLEU algorithm, which measures the quality of translations and reported that it achieved BLEU scores that were 44% higher than any previous record (Meta, 2022).

MT Quality Evaluation

Various studies have evaluated the quality of Instagram machine translation (MT) since its emergence. Analytical approaches have been used to identify linguistic errors in the translations. Pudjiati and Fadilah (2017) identified three types of semantic errors in the output: referential, grammatical, and contextual. Grammatical and contextual errors were the most frequent, while the translation of dictionary meaning performed better. Mawarni et al. (2017) focused on cultural-specific terms (CSTs) and found a loss of meaning in the translations, failing to transfer the expressive meaning to the target culture. In line with previous findings, MT succeeded in translating referential meaning but failed in translating pragmatic meaning. This study aims to investigate whether the new advanced model (NLLB-200) of Instagram MT can improve.

Furthermore, Meilasari (2019) evaluated the accuracy of Instagram MT translations related to ecology and environment vocabulary. The study found that the MT was unreliable, with 40% of the translations being inaccurate and only 24% accurate. Susanti (2018) analyzed Instagram MT translations and identified incorrect and missing words as the most frequent lexical errors. The study also found that the MT tended to use a word-for-word translation method, resulting in a lack of recognition of the text's context and failing to represent the authentic language. Other researchers have compared the quality of Instagram machine translation with human translation. Veronika (2017) compared Instagram MT to human translators and found that Instagram MT made 54 translation errors, while human translators made only 8 errors. Similarly, Arvianti (2018) compared the performance of Instagram MT and human translators in translating formal and informal language. The study found that while Instagram MT produced good translations for formal language, it failed to translate texts written in an informal language. Human translators were better able to recognize languages and better understand context due to their more extensive vocabulary and context understanding. In addition, Instagram MT translations have been compared to output from other MT systems. Larassati et al. (2019) evaluated the output of Google Translate and Instagram MT using NMT and found that both systems had translation errors, with Instagram MT having more errors. The most frequent error types were terminology errors, syntax errors, and literalness, which were related. Similarly, Pujakesuma (2022) found that Google Translate and Instagram MT made similar errors, such as mistranslation, and applied the same translation strategies, such as literal translation. Both studies are consistent with Sujianti's (2012) and Purwaningsih's (2019) findings that literal translation is the dominant technique used by Instagram MT.

Moreover, some researchers have evaluated Instagram MT by exploring its translation strategies. Purwaningsih (2019) investigated the translation strategies used by Instagram MT when translating Indonesian cultural-specific items, specifically Banyumas Batik motifs. The study found that Instagram MT used three techniques, including literal translation, borrowing, and

particularization, with borrowing being the dominant technique for translating cultural items. However, this led to a loss of cultural sense. Purwaningsih recommended that Instagram developers enrich the MT with a more extensive contextual linguistics database to improve the quality of translation results.

Existing literature on Instagram MT has focused on the Indonesian-English language pair, leaving a research gap in Arabic. Ameur et al. (2020) note that many linguistic problems related to Arabic still need further investigation as they pose significant challenges to current Arabic MT systems. Therefore, this study aims to fill this research gap by evaluating the translation quality of Instagram MT for Arabic captions.

Text type

Nord (2005) proposed a tripartite model of the functions of linguistic signs inspired by Bühler's (1934) work, which includes four basic functions of communication in language: referential, expressive, operative, and phatic. The referential function focuses on the meaning or content referred to and represented in informative texts, such as scientific articles and news. The expressive function refers to the emotions and attitudes of the sender towards the referred object, thought, or idea implied in texts of high aesthetic value, such as literary works. The operative or appellative function is concerned with the direction of the text toward the addressee. The phatic function focuses on establishing communication between sender and receiver and attracting the attention of the receiver regarding certain things. This paper focuses on the expressive function implied in literary texts and seeks to explore how Instagram NMT can deal with the unique sentence structures, cultural elements, and aesthetic features present in literary language that have fewer counterparts stored in the MT database. Additionally, the study aims to investigate the extent to which this system can transfer the expressive function implied within literary texts.

METHODOLOGY

Research Design

The present study utilized a qualitative descriptive method and analyzed written content (captions) taken from the @cairo_mockingbird Instagram account, a virtual platform for visual arts and literary writings. The account contains over 12,000 posts, each featuring a photo and a caption written in Arabic (as of last access on 24/3/2023). It is a community platform designed for displaying visual arts and literary writings. To achieve the objectives of the study, the selected data were analyzed by using a non-DEJ-based analytical evaluation method called the Multidimensional Quality Metrics (MQM) core typology. The latest version of MQM (OCT 2021), was developed to allow a harmonization with TAUS DQF (Dynamic Quality Framework) error typology, which resulted in creating a flexible subset to MQM, as shown in (Fig 1) below. A non-DEJ-based evaluation method was used in this study, where the judge (annotator) indirectly expresses judgment on the translation quality. Such evaluation methods are commonly used to evaluate the accuracy and fluency of both human and machine translation results and involve comparing either the source text with the target text or the target text with the translation reference (Chatzikoumi, 2020). In this study, the MQM core typology framework was used as an analytical tool to identify, classify, and annotate errors found in Instagram MT's output. Error analysis is a

means to access the output of machine translation in qualitative terms and can be used to create an error profile for different systems (Stymne & Ahrenberg, 2012).

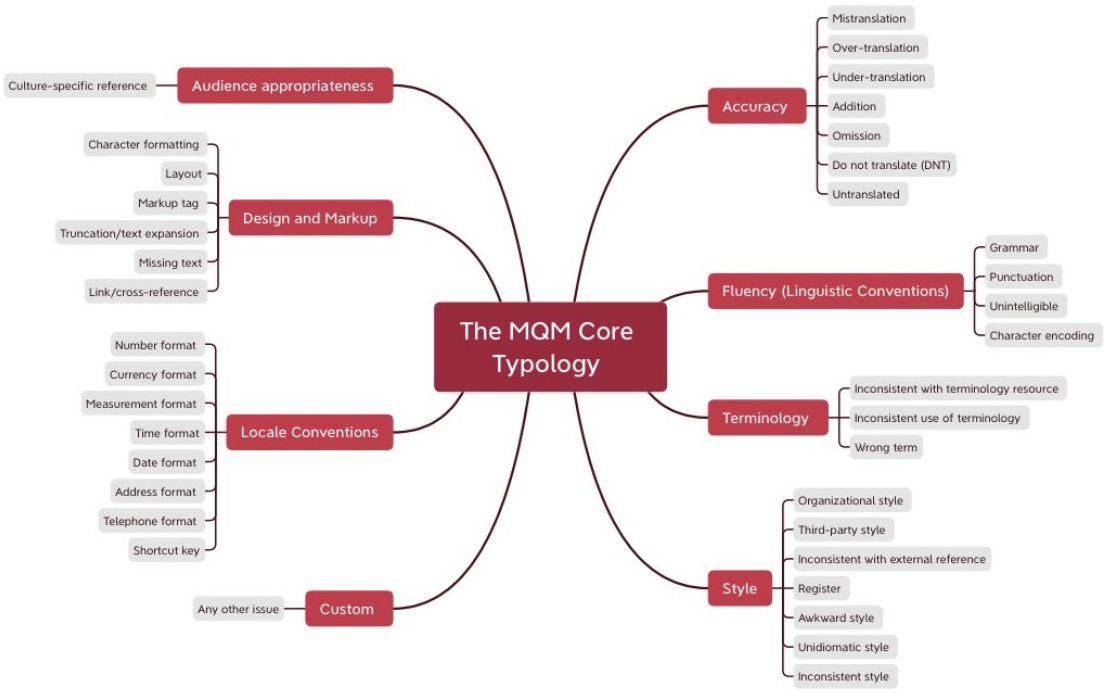

Figure 1. The MQM core (http://www.themqm.info/ (Last access 7/3/2023)).

Translation Quality Evaluation (TQE) Stages.

Stage 1: Preliminary Stage

a) Translation specifications

The following translation parameters were adopted from the structured translation specification framework of 2006 ATSM Standard Guide for Quality Assurance in Translation (ASTM F2575-06). This approach is formulated based on Melby's (2002) paralleled work on Skopos theory and the translation brief, and Nord's extension of Skopos theory (1997), which is known as Functionalism in translation theory and practice. The translation parameters, (Table 1) and (Table 2), adopted in this paper were selected in a way that they realize the objectives of the present study under investigation.

Table 1 Source Content Information

Textual characteristics	
Source language	Arabic (Modern Standard Arabic and Egyptian Arabic)
Text type	Literary texts
Audience	Instagram users who are familiar with Arabic language and culture
Purpose	Expressive function: the text is intended to convey a particular message in the mind of an author in an artistic form.

Specialized language (Subject field)	The captions consist of sayings and texts quoted from novels and other literary sources.
Specialized language (Terminology)	The texts do not include specialized or complicated terminology, but rather everyday use vocabulary. Therefore, it does not require a specialized term base.
Volume	30 captions (388 words)
Complexity	Some captions are written a straightforward form while some others in an artistic style.
Origin	The source texts are captions posted on @cairo_mockingbird Instagram account.

Table 2 Target Text Requirements

Target language	English
Audience	Instagram users who can understand English.
Purpose	Expressive function
Content Correspondence	The ST should be translated accurately and fluently.
Register	Texts written in Modern Standard Arabic should be translated into formal English while texts written in Egyptian Arabic should be translated into informal-colloquial English.
Format	Captions underneath a photo on Instagram
Style	Stylistics should be taken into consideration in translating the ST.

b) Evaluation Metric Design

A metric is a measurement with a specific purpose (Lommel & K. Melby, 2018). To verify the metric of the evaluation, three dimensions were selected from the MQM core typology and its subsets, as shown in Figure 2. The main goal of an MT system is to automatically translate text while preserving its meaning and style, ensuring that the output is as linguistically fluent as possible (Ameur et al., 2020). The evaluation focused on three aspects: accuracy (adequacy), which considers the semantic and pragmatic equivalence of lexis between the source and target texts; fluency, which refers to the linguistic conventions of the target language and naturalness (Chatzikoumi, 2020); and style, which measures the extent to which the translated text uses appropriate language to convey the message effectively. Therefore, the evaluation covered the translated texts' lexical, syntactic, semantic, pragmatic, and stylistic aspects.

The errors extracted from the TT were measured according to the following Error Severity Levels:

1) Minor errors, which do not affect the comprehension of meaning but affect the fluency (Weight: 1)
2) Major errors, which make TT difficult to understand, yet the general message is conveyed. (Weight: 5)
3) Critical errors, which change the meaning of ST and make it incomprehensible or distorted (Weight: 10)

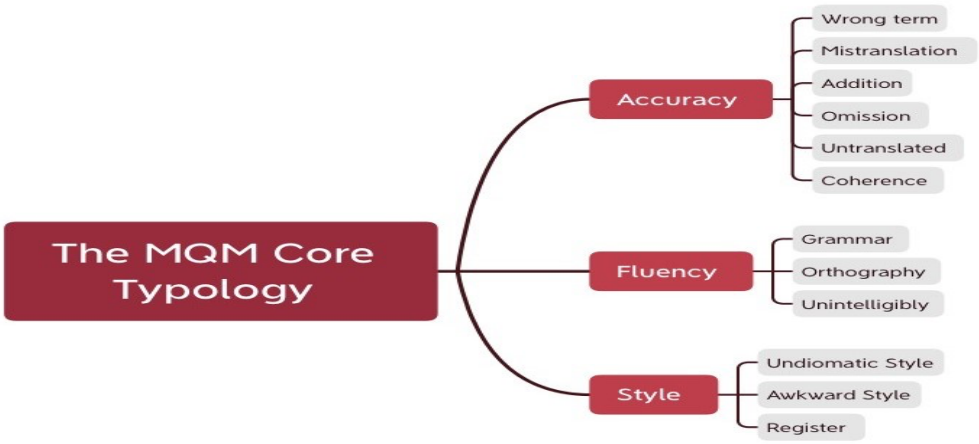

Figure 2. A Metric Designed for the based on the MQM Framework Evaluation in the Present Study

Data Collection

The data collection process includes three phases: (i) selecting the source of data (Instagram account), (ii) selecting the data (captions), and (iii) collecting the selected data. Data source selection phase is determined by the following criteria:

- The source material should contain the data necessary to answer the stated research objectives.
- The data included in the source should not extend beyond the scope of the study.
- The source should be a verified account with a significant number of followers.
- Data included in the account should be in form of captions (texts) and not audio-visual elements.

The researchers selected the @cairo_mockingbird Instagram account as it met the data source selection criteria. The account shares a variety of literary writings daily, providing ample samples for the evaluation and contributing to answering the research questions. All captions are written in Arabic, ensuring the data remains within the study's scope. Additionally, the account is verified with 826 thousand followers (as of last access on 24/3/2023). Finally, the account often uploads literary writings illustrated in a photo with the exact text in a caption below the photo, making the data easily accessible.

In the study, there are two types of data: the primary data, the original captions written in Arabic that is referred to as “Source Text” (ST), and the English machine translations that is referred to as “Target Text” (TT). A purposive sampling was employed in data collection procedure that was carried out in the following steps. Firstly, the researchers read intently all the captions posted in @cairo_mockingbird Instagram account. Secondly, 30 captions, ranging from short to medium-length texts written in Modern Standard Arabic (MSA) and Egyptian Dialect in

form of poetic language, were selected purposively. Thirdly, the researchers collected the translated results after tapping on "انظر للترجمة" or "See Translation" feature set beneath the selected captions that instantly translates the written captions into English, the language in which the researchers have set in their personal Instagram application. Finally, source texts (the captions) and target texts (their translations) are collected and divided into segments. A segment pair, which contains corresponding content (a source text and target text), is called a translation unit (TU).

Stage 2: Error Annotation

In this stage, the annotation was conducted semi-automatically using the harmonized MQM and DQF error typology, i.e., MQM Core Typology integrated with ContentQuo platform. The annotators (two experienced translators along with a skilled linguist) examined the translated text against the source text based on the on-agreed translation specifications, and annotated (i.e., identifying, classifying, and assigning error type and penalty points) errors analytically in accordance with the designed metric.

Stage 3: Automatic Calculation

Through this stage, the Overall Quality Score is calculated automatically by ContentQuo according to the selected scoring model using the following formula: QualityScore = 100 - 100 * (ErrorPoints / Wordcount), then compared to the Threshold Value (100%) to assign a pass/fail rating.

RESULTS AND ANALYSIS

This section shows the results of the analytical evaluation conducted on the Instagram MT translation of 30 captions selected from @cairo_mockingbird account. As illustrated, Instagram MT failed at translating 90% of the data from three different aspects: accuracy, fluency, and style. 61 errors were found in the selected data: 26 errors in Fluency, 25 errors in Accuracy, and 10 errors in Style. Their severity level ranging from minor to critical errors, as illustrated in Figure (4).

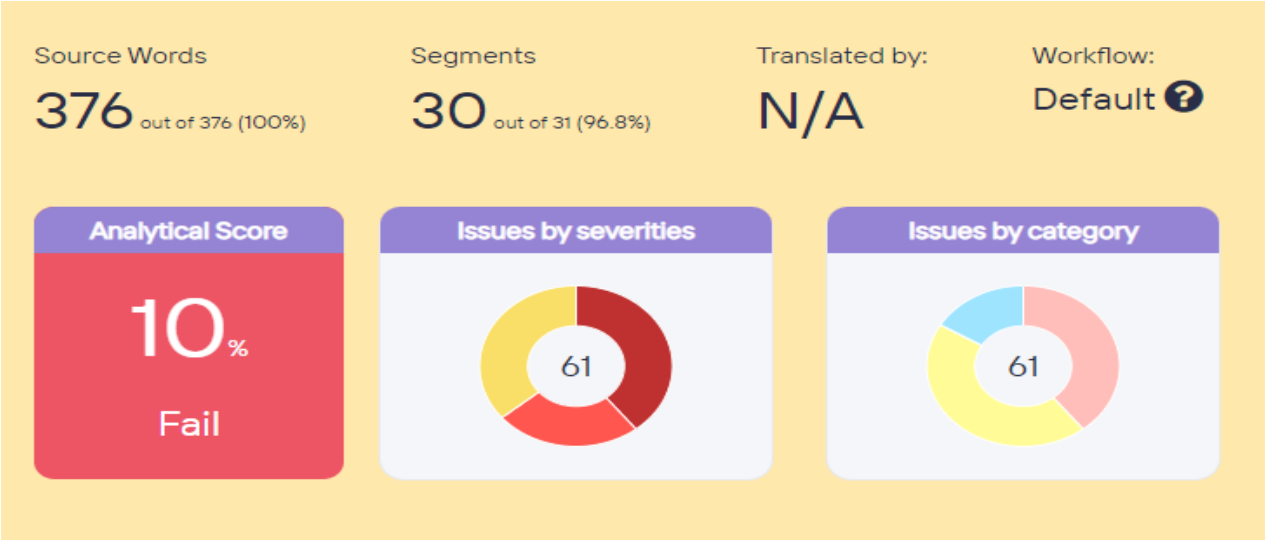

Figure 3 the Overall Quality Score

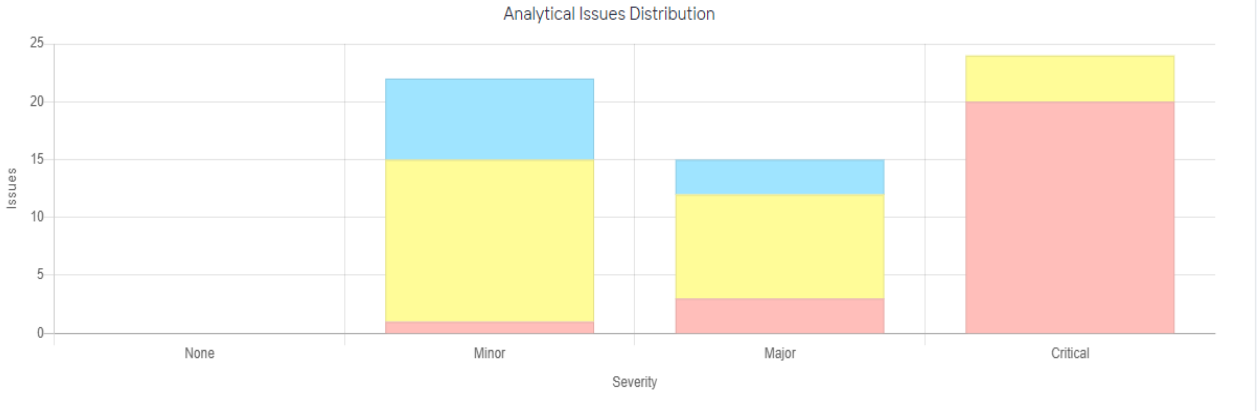

Figure 4 Issue Severity Levels

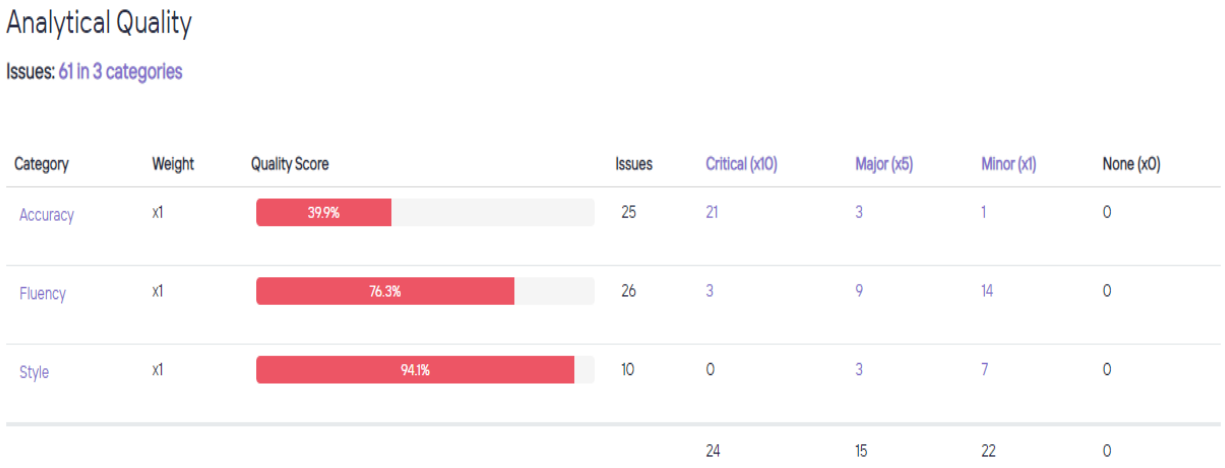

Figure 5

Accuracy

Accuracy categories concerned how the MT system recognized the meaning of the source text and reproduced it in the target text. Based on the results of the TQE, Instagram MT produced a plenty of errors under this category. For instance, Instagram MT system was not able to recognize the exact meaning of the ST term from the context, hence failing to come up with the equivalent that represents it in the TL. As shown in Table (3), The TM was not able to recognize the exact meaning of the polysemous word (بحث search), so it chose the wrong equivalent (research) in the TL.

TU	ST	TT
1	الأمان جميل جداً، أظنه الشعور الوحيد الذي يستحق عناء البحث.	Security is so beautiful, I think it's the only feeling worth the effort to research.

Table (3)

Moreover, in literary texts, authors sometimes represent the message they want to express as a figurative of speech, such as a metaphor. Instagram MT struggled in understanding and translating the metaphors in the source texts. In Table (4), the MT system translated the caption word for word which distorted the intended meaning implied in the metaphor. The vehicles (سماء Sky) and (أرض Earth) that carry the meaning of the topic (God) and (People) indicates that God is above the sky and people are down on the earth. The MT translation did not manage to convey the ground relationship implied.

TU	ST	TT
2	نَلتَمِسُ بالسماء ما تَرفضُ الأرضُ أن تمنحهُ لنا.	We touch heaven what the earth refuses to give us.

Table 4

Furthermore, it was found out that Instagram MT system tended to misread the captions even though they were written in a direct sentence structure and partially or fully vocalized. Unlike English, Arabic language is characterized by having no letters to represent the vowel sounds. Instead, the Arabic writing system uses small signs that are added above or below the letters as vowel sounds called diacritics and the presence of such diacritical signs is known as "Vocalization" (Ameur et al., 2020). Vocalization clarifies the way of reading words and their exact meanings which indeed helps in solving lexical, sematic, and pragmatic ambiguities in translation. The MT system in the present study failed to misread these signs, hence reproducing wrong equivalents. In Table (4), the verb (نَلْتَمِسُ), meaning (ask for) was mistranslated into (touch نَتَلَمَسَ). It can be concluded that the MT system still cannot decide the exact equivalent for a word with or without vocalization.

TU	ST	TT
3	نجيب محفوظ	Najeeb is safe
4	د. جاسم المطوع	Jasim the volunteer - د

Table 5

Additionally, one of the frequent translation errors that Instagram MT produced was the mistranslation of proper names. As demonstrated in Table (5), the MT system transliterated the first names while it translated literally the surnames. This is a common issue occurring when translating from Arabic into English because Arabic language does not have a united system or rules in writing Arabic named entities, such as capitalization. Besides, since Arabic is a highly inflected language with rich lexical variations, it makes the issue even more complicated.

Fluency

Fluency error categories include errors related to the linguistic well-formedness of the translated text, including morphology, syntax, orthography, and sentence readability. The evaluation results showed that fluency errors were the most frequent errors produced by Instagram MT. These errors range from minor errors that only affect the fluency of the TT, to major errors that make the text hard to understand but conveyed the general message, and critical errors that distort the meaning and make the TT unintelligible.

One of the root causes that led to the fluency errors was the loose word order of Arabic language. Unlike English language that has only one rigid SVO word order, Arabic has a flexible

sentence structure that can occur in multiple orders, such as SVO, VSO, OVS, etc. When it comes to machine translation, this issue causes several problems when translating from Arabic into English as the MT systems are built on fixed encoding and decoding mechanisms and algorithms that get confused by the multiple sentence structure (i.e., word order) that Arabic language could take. Therefore, these MT systems fail to produce the Arabic text into the TT. As shown in Table (6), Instagram MT mistranslated the Arabic sentence that has an OVS word order, resulting in an unintelligible output.

TU	ST	TT
5	على دفء العائلة تتكئ البيوت.	Warmth of family leaning homes.

Table 6

Another problem that was noticed in the TQE of Instagram MT translations was that they lacked pronoun-antecedent agreement. In English, both the pronoun and antecedent (i.e., a word to which a pronoun refers) should agree in number, person, and gender. The MT tended to translate each segment of the caption separately. The pronoun (i.e., it) in the target text in Table (7), for instance, contradicts with its antecedent (years) in number. The MT read and translated the two sentences independently out of the context, resulting in incohesive translations.

TU	ST	TT
6	لأعوام تغير الكثير.. أنها تبدل تضاريس الجبال، فكيف لا تبدل شخصيتك؟ - أحمد خالد توفيق	The years changed a lot… It changes the mountains, how can it not change your character? - Ahmed Khaled Tawfiq

Table 7

Furthermore, another frequent issue was errors related to orthography, including target language conventions of writing, such as norms of spelling, hyphenation, capitalization, word breaks, emphasis, and punctuation. These errors might not be critical, but they negatively affect the readability of the translations. It was noticed that Instagram MT tended to imitate the ST writing conventions that caused to produce poor written translations. This strategy might be usable in languages that have similar writing norms, but in our case, the source language and the target language have completely different orthographic systems, it led to considerable issues, such as small letters at the beginning of a sentence and capital letters in the middle of a sentence and lacks of proper punctuation marks, among other things.

Style

Literary writings are expressive texts that value the form of texts. Stylistics was highly considerable in the conducted evaluation since the subject under investigation was literary texts. Several stylistic errors were found in Instagram MT output. The MT system used basic translation strategies, e.g., literal and word-for-word translations with all types of texts, be it informative, expressive, or persuasive. Literal translation could work in informative texts that focus only on the content while in the literary texts that value the form as well, it was one of the cause roots for generating translations that lacked the aesthetic values and had awkward sentence structure, as illustrated in Table (8).

TU	ST	TT
7	لست أفهمُ من معنى الحبِ إلا أن الرُوح قد اهتدت إلى شيء من سِر الإنسانية في إنسانٍ جميل. ـ مصطفى الرافعي	I do not understand the meaning of love, except that the soul has been guided to something of the secret of humanity in a beautiful human being. - Mostafa Al-Rafay

Table 8

Idiomatic expression translation can be problematic, especially in machine translation that most of times these expressions end up to be translated literally. It occurred when there are linguistic or cultural gaps between the SL and TL. However, Instagram MT failed to translation expressions that had one-to-one direct equivalent in the TT. This is clearly demonstrated in Table (9) when Instagram MT translated the ST literally though there was a direct equivalent in English.

TU	ST	TT
8	ما تزرعه اليوم تحصدهُ غداً..	What you plant today you will harvest tomorrow.

Table 9

Another issue commonly found in the output of Instagram MT was the lack of conformity of the register of the ST. The translations seemed to have informal style by using colloquial terms and contractions to reproduce the formality of the ST that represented in in using Modern Standard Arabic. This can be demonstrated in Table (10).

TU	ST	TT
9	مش معنى إن حد شايل الشيلة كويس يبقى الشيلة مش تقيلة!	It doesn't mean that someone is carrying the burden well Then the burden is not heavy!

Table 10

Despite the above-mentioned weaknesses in Instagram MT system, the system has shown improvement in some other aspects. It was able to translate texts written in the Egyptian dialect properly. As shown in Table (10), the MT system managed to recognize the colloquial words (شايل), (الشيلة), and (كويس), and translated them into their proper equivalent terms in English (is carrying), (the burden), and (well).

DISCUSSION

This small-scale exploration questioned whether Instagram MT is capable of producing fluent translations that well maintain the intended message implied within the literary texts to the target language. The results of the evaluation revealed that several translation errors, covering different linguistic aspects, including lexis, syntax, semantics, pragmatics, orthography, and stylistics were produced by the MT system that hindered the process of transferring the purpose of the source texts in fluent well-structured translations which definitely go against the translation specifications that were set by the researchers before the evaluation. These results are in opposition to the concept of translation quality suggested by Koby et al. (2014) who defined translation quality as reproducing accurate and fluent translation results for the target audience that can serve the original

purpose and comply with all other specifications negotiated between the requester and provider, taking into account needs of the end-users. Consequently, Instagram neural machine translation system is not capable of producing well-structured translations that properly convey the intended message and the aesthetic value of the literary texts. Even with the enormous improvements made on Instagram neural MT, these linguistic aspects are still problematic. This is in line with Susanti (2018), who questioned the reliability of translation results produced by Instagram machine translation feature. Instagram MT apparently does not have the flexibility to deal with texts that do not have parallel structures in its database and it cannot recognize the contextual and cultural knowledge implied in the inputs as it only uses literal translation to produce the dictionary meaning of the linguistic units (Purwaningsih, 2019; Omar & Gomaa, 2020). It is clear from the data shown that some of the resultant problems are brought about by translating the texts literally without considering the contextual or cultural references which are essential in literary texts, hence resulting in a loss of the pragmatic sense, and in some cases, the whole meaning get distorted. Likewise, Cahyani et al. (2021) argued that Instagram Machine Translation produced inappropriate translation because it used improper procedures by choosing the lexis in the target language literally through a term that has several synonyms which have different meanings in each use without considering the overall context of the caption. Meilasari (2019) concluded that Instagram translation machine is not a reliable machine for the target language reader who wants to understand certain cultural terms in the source language because the MT only produces the translation product literally based on what is provided by the source text and has no ability to analyze and restructure the text that is being translated. Based on what the results shown, Instagram NMT uses basic translation techniques, such as word-for-word translation without providing additional information or further explanation about the cultural terms being translated.

The poor quality of Instagram MT's Arabic-English translations is due to the significant linguistic and cultural differences between the two languages. Arabic and English belong to different language families and have distinct grammar rules, morphology, semantics, pragmatics, and writing conventions. These differences make it difficult for MT systems to recognize and bridge the gaps, resulting in inadequate translations. Morphologically rich languages like Arabic pose even more significant challenges for MT systems. The flexibility of word order within Arabic also makes it difficult for translation systems to make accurate choices, negatively impacting the quality of translations (Ameur et al., 2020; Omar & Gomaa, 2020). Literary texts pose a greater challenge for machine translation systems due to their unique style and use of figurative language, special diction, and language enhancers that carry implied meanings beyond words and sentences. Arvianti (2018) noted that MT systems have limited vocabulary and context understanding compared to human translators, making it difficult for them to recognize special language. Omar & Gomaa (2020) investigated the challenges of applying MT systems to literary translation and found that although errors may occur, the usefulness of MT systems should not be dismissed.

CONCLUSION

This paper evaluates the output of automatic translation of Arabic literary writings on Instagram. The evaluation involves an analysis of translated texts, identification, classification, and assignment of error types and penalty points using the MQM core typology. The study explores the ability of Instagram MT to convey the implied message in literary texts. The evaluation results reveal multiple translation errors that negatively impact the translations' accuracy, fluency, and

style, hindering the conveyance of the intended message of the source texts. The findings indicate that Instagram MT fails to translate 90% of the data at three levels: accuracy, fluency, and style. Specifically, the selected data exhibited 61 errors, comprising 26 in fluency, 25 in accuracy, and 10 in style. These errors significantly affect the quality of translations, thereby impeding the transference of the intended message embedded within the source texts.

LIMITATIONS AND FURTHER RESEARCH

This study examines the quality of Instagram's AI-based machine translation for translating literary Arabic into English. Further research can investigate other aspects of the Arabic context and compare Arabic linguistic needs to other languages in MT systems. As AI continues evolving, further evaluations are necessary to assess MT applications and texts in different language pairs.

ACKNOWLEDGEMENT

We sincerely thank ContentQuo, a translation quality management SaaS, for their indispensable support in providing localization services and CAT tools to evaluate raw MT and PEMT output using Error Annotation and Rating Scale.

REFERENCES

Alkhatib, M., & Shaalan, K. (2018). The key challenges for Arabic machine translation. *Intelligent Natural Language Processing: Trends and Applications*, 139-156.

Ameur, M. S. H., Meziane, F., & Guessoum, A. (2020). Arabic machine translation: A survey of the latest trends and challenges. *Computer Science Review*, *38*, 100305.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R., & Sadler, L. (1994). Machine translation: An introductory guide. NCC Blackwell.

Arvianti, G. F. (2018). Human translation versus machine translation of Instagram's captions: Who is the best? In *English Language and Literature International Conference (ELLiC) Proceedings* (Vol. 2, pp. 531-536).

Arvianti, G. F. (2018). Human translation versus machine translation of Instagram's captions: Who is the best? In *English Language and Literature International Conference (ELLiC) Proceedings* (Vol. 2, pp. 531-536).

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: a case study. *In Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (pp. 257–267). arXiv preprint arXiv:1608.04631*.

Bhattacharyya, S. (2022, March 7). *Meta's machine translation journey*. Analytics India Magazine. Retrieved March 22, 2023, from https://analyticsindiamag.com/metas-machine-translation journey/#:~:text=Meta%20used%20neural%20machine%20translation,training%20time %20from%2024%20hours.

Cahyani , N. L. D. (2022). *Derivational Affixes Found in the Caption of Selected Posts of @bawabali_official Account on Instagram* (thesis).

Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017a). Is neural machine translation the new state of the art? *The Prague Bulletin of Mathematical Linguistics*, (108).

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, P., ... & Gialama, M. (2017b). A comparative quality evaluation of PBSMT and NMT using professional translators. In *Proceedings of Machine Translation Summit XVI: Research Track* (pp. 116-131).

Chatzikoumi, E. (2020). How to evaluate machine translation: A review of automated and human metrics. *Natural Language Engineering*, *26*(2), 137-161.

Chéragui, M. A. (2012). Theoretical Overview of Machine translation. *ICWIT*, 160-169.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. In *Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation (SSST-8). arXiv preprint arXiv:1409.1259*.

Dorr B., Snover M. and Madnani N. (2011). Chapter 5.1 introduction. In Olive J., McCary J. and Christianson C. (eds), *Handbook of Natural Language Processing and Machine Translation. DARPA Global Autonomous Language Exploitation*. New York: Springer, pp. 801–803.

España-Bonet, C., & Costa-jussà, M. R. (2016). Hybrid machine translation overview. *Hybrid Approaches to Machine Translation*, 1-24.

Habash, N., Dorr, B., & Monz, C. (2009). Symbolic-to-statistical hybridization: extending generation-heavy machine translation. *Machine Translation*, *23*, 23-63.

Hunsicker, S., Chen, Y., & Federmann, C. (2012, June). Machine learning for hybrid machine translation. In *Proceedings of the Seventh Workshop on Statistical Machine Translation* (pp. 312-316).

Hutchins, W. J. (1995). Machine translation: A brief history. In *Concise History of the Language Sciences* (pp. 431-445). Pergamon.

Hutchins, J. (2007). Machine translation: A concise history. *Computer Aided Translation: Theory and Practice*, *13*(29-70), 11.

Koby G.S., Fields P., Hague D., Lommel A. and Melby A. (2014). Defining translation quality. *Tradumàtica* 12, 413–420.

Koehn, P. (2009). *Statistical machine translation*. Cambridge University Press.

Larassati, A., Setyaningsih, N., Nugroho, R. A., Suryaningtyas, V. W., Cahyono, S. P., & Pamelasari, S. D. (2019, September). Google vs. Instagram machine translation: multilingual application program interface errors in translating procedure text genre. In *2019 International Seminar on Application for Technology of Information and Communication (iSemantic)* (pp. 554-558). IEEE.

Lommel, A., & Melby, A. K. (2018, March). Tutorial: MQM-DQF: A good marriage (Translation quality for the 21st Century). In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 2: User Track)*.

Mawarni, B., Pambudi, B. D., & Ghasani, B. I. (2017, December). THE PROBLEM OF―CULTURAL UNTRANSLATABILITY‖ FOUND IN THE ENGLISH TRANSLATION OF JOKOWI 'S INSTAGRAM POSTS. In *UNNES International Conference on ELTLT* (pp. 104-108).

Meilasari, P. (2019). When Instagram translation machine translates ecology terms: Accurate or not? In *The 7th LITERARY STUDIES CONFERENCE* (p. 129).

Melby, A. K. (2002). The mentions of equivalence in translation. *Meta*, *35*(1), 207-213.

Meta AI. (2022, July 6). *200 languages within a single AI model: A breakthrough in high-quality machine translation*. Meta AI. Retrieved March 22, 2023, from https://ai.facebook.com/blog/nllb-200-high-quality-machine-translation/

Mannes, J. (2017, August 3). *Facebook finishes its move to Neural Machine Translation*. TechCrunch. Retrieved March 22, 2023, from https://techcrunch.com/2017/08/03/facebook-finishes-its-move-to-neural-machine-translation/?guccounter=1

Moorkens, J., Toral, A., Castilho, S., & Way, A. (2018). Translators' perceptions of literary post-editing using statistical and neural machine translation. *Translation Spaces*, *7*(2), 240-262.

Nord, C. (2005). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis* (No. 94). Rodopi.

Nord, C. (1997). Functionalist approaches explained. *Manchester, UK: St. Jerome Publishing*.

Omar, A., & Gomaa, Y. (2020). The machine translation of literature: Implications for translation pedagogy. *International Journal of Emerging Technologies in Learning (IJET)*, *15*(11), 228-235.

Pino, J. M., Sidorov, A., & Ayan, N. F. (2017, August 3). *Transitioning to neural machine translation*. Tech at Meta. Retrieved March 22, 2023, from https://tech.facebook.com/artificial-intelligence/2017/8/transitioning-entirely-to-neural-machine-translation/

Pujakesuma, G. A. (2022). *The Performance of Instagram's Auto-Translate and Google Translate in Translating House of Highlight's Instagram Captions* (thesis).

Pudjiati, D., & Fadilah, E. (2017, December). Semantic Errors in the Translation about Actions to Defend Islam in 2016 into English. In *International Conference on Culture and Language in Southeast Asia (ICCLAS 2017)* (pp. 98-102). Atlantis Press.

Purwaningsih, D. R., Sholikhah, I. M., & Wardani, E. (2019). Revealing Translation Techniques Applied in the Translation of Batik Motif Names in See Instagram. *Celt: A Journal of Culture, English Language Teaching & Literature*, *19*(2), 287-301.

Ragni, V., & Nunes Vieira, L. (2022). What has changed with neural machine translation? A critical review of human factors. *Perspectives*, *30*(1), 137-158.

Sabtan, Y. M. N., Hussein, M. S. M., Ethelb, H., & Omar, A. (2021). An evaluation of the accuracy of the machine translation systems of social media language. *International Journal of Advanced Computer Science and Applications*, *12*(7).

S. Dixon. (2023, February 15). *Instagram users worldwide 2025*. Statista. Retrieved March 22, 2023, from https://www.statista.com/statistics/183585/instagram-number-of-global-users/

Sipayung, K. T., Sianturi, N. M., Arta, I. M. D., Rohayati, Y., & Indah, D. (2021). Comparison of Translation Techniques by Google Translate and U-Dictionary: How Differently Does Both Machine Translation Tools Perform in Translating? *Elsya: Journal of English Language Studies*, *3*(3), 236-245.

Sujianti, P. E., Simbolon, M., & Luardini, M. A. (2020). An Analysis of English-Indonesian Translation Procedures Used in Instagram Mobile Application. *Academic Journal of Educational Sciences*, *4*(1), 1-5.

Susanti, E. (2018). *Lexical errors produced by Instagram machine translation* (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).

Stymne, S., & Ahrenberg, L. (2012, May). On the practice of error analysis for machine translation evaluation. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 1785-1790).

Tambouratzis, G. (2014, April). Comparing CRF and template-matching in phrasing tasks within a Hybrid MT system. In *Proceedings of the 3rd Workshop on Hybrid Approaches to Machine Translation (HyTra)* (pp. 7-14).

Thurmair, G. (2009). Comparing different architectures of hybrid machine translation systems. In *Proceedings of Machine Translation Summit XII: Posters*.

Veronika, S. T. E. F. A. N. I. (2018). Instagram Translate's and Human Translation's Performance in Translating the Captions in@ Basukibtp Instagram Account. Published Undergraduate Thesis, Yogyakarta: Sanata Dharma University.

Widiastuti, N. M. A. (2021). Translation Procedures of English Phrasal Verbs into Indonesian on Instagram Captions. In *International Seminar on Austronesian Languages and Literature IX (399-408). Denpasar: Udayana University*.

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