



Albariq: Jurnal Pendidikan Bahasa Arab, 7 (1), 2026, 18-32

DOI: <https://doi.org/10.24239/albariq.v1i1.1>

E-ISSN: 2746-3362

Received: May 26, 2026

Revised: July 28, 2026

Accepted: July 29, 2026

Published: July 31, 2026

## Arabic Linguistic Error Analysis in TikTok Auto-Generated Captions on the @syukriarabs Account

Nur Ilahi Ulfa<sup>1</sup>, Sinta Ramandani<sup>2</sup>, Ari Khairurrijal Fahmi<sup>3</sup>, Najmi Haqiqi<sup>4</sup>

<sup>1234</sup>UHAMKA, Indonesia

Corresponding E-mail: [nulfa100@gmail.com](mailto:nulfa100@gmail.com),

### Abstract

*The development of Artificial Intelligence (AI) through TikTok's auto-caption feature offers conveniences in digital communication and Arabic language learning; however, the system frequently encounters errors when processing complex Arabic linguistic structures. While previous studies have extensively discussed Arabic error analysis in academic texts and machine translation applications, research exploring real-time linguistic errors in TikTok's auto-caption feature remains limited. This study aims to analyze Arabic linguistic errors in TikTok's auto-caption feature on the @syukriarabs account through the lens of Nahwu (syntax) and Sharf (morphology). This study employs a descriptive qualitative approach with a case study design. Data were obtained from 10 videos uploaded by the @syukriarabs account, accessed on May 3, 2026, and collected over a two-week period using the observation and note-taking technique. The sampling utilized purposive criteria, selecting videos featuring standard Arabic (Fusha) with clear audio articulation. The data were analyzed using S.P. Corder's Error Analysis theory by comparing original audio texts, generated auto-captions, correct grammatical forms, and specific error types. The results indicate that the most dominant errors occur within the phonetic and morphological domains, particularly in speech sound misdetection, wazan alterations, tanwin omission, pronoun shifts, and syntactic inaccuracies. Furthermore, lexical and word segmentation errors were identified, causing semantic distortion. This study concludes that TikTok's auto-caption system remains oriented toward sound-matching rather than a comprehensive understanding of Arabic linguistic context.*

**Keywords:** TikTok auto-caption, Arabic speech recognition, linguistic errors, Arabic morphology, Arabic syntax

### Abstrak

Perkembangan Artificial Intelligence (AI) melalui fitur auto-caption TikTok memberikan kemudahan dalam komunikasi digital dan pembelajaran bahasa

Arab, namun sistem tersebut masih sering mengalami kesalahan dalam memproses struktur linguistik Arab yang kompleks. Penelitian sebelumnya banyak membahas analisis kesalahan bahasa Arab pada teks akademik dan aplikasi penerjemahan otomatis, tetapi kajian mengenai kesalahan linguistik fitur auto-caption TikTok secara real-time masih terbatas. Penelitian ini bertujuan menganalisis kesalahan linguistik bahasa Arab pada fitur auto-caption TikTok di akun @syukriarabs melalui kajian Nahwu dan Sharf. Penelitian menggunakan pendekatan kualitatif deskriptif dengan desain studi kasus. Data diperoleh dari 10 video unggahan akun @syukriarabs yang diakses pada 3 Mei 2026 dan dikumpulkan selama periode dua minggu melalui teknik simak dan catat. Pengambilan sampel menggunakan kriteria purposive sampling, yaitu memilih video berorientasi bahasa Arab standar (Fusha) dengan artikulasi audio yang jelas. Data dianalisis menggunakan teori Error Analysis S.P. Corder dengan membandingkan teks audio asli, hasil auto-caption, bentuk kaidah yang benar, dan jenis kesalahan. Hasil penelitian menunjukkan bahwa kesalahan paling dominan terdapat pada ranah fonetik dan morfologi, terutama pada kesalahan deteksi bunyi huruf, perubahan wazan, penghilangan tanwin, perubahan dhamir, serta ketidaktepatan sintaksis. Selain itu, ditemukan kesalahan leksikal dan segmentasi kata yang menyebabkan distorsi makna. Penelitian ini menyimpulkan bahwa sistem auto-caption TikTok masih lebih berorientasi pada pencocokan bunyi dibandingkan pemahaman konteks linguistik bahasa Arab secara menyeluruh.

**Kata Kunci:** Auto-caption TikTok, Pengenalan Suara Bahasa Arab, Kesalahan Linguistik, Morfologi Arab, Sintaksis Arab

## Introduction

The rapid advancement of digital technology has significantly transformed the landscape of human communication, particularly within the context of social media platforms like TikTok, which has now emerged as a primary hub for the dissemination of visual and audio content.<sup>1 2</sup> Amidst these conveniences, the integration of Artificial Intelligence (AI) through auto-caption and automatic translation features powered by Neural Machine Translation (NMT) is viewed as a cutting-edge solution to overcoming language barriers.<sup>3 4 5</sup> For Arabic language

---

<sup>1</sup> Wiwik Romauli Nababan and others, 'Tantangan Bahasa Di Era Digital Terhadap Kesalahan Berbahasa Dalam Komunikasi Media Sosial', *Jurnal Bahasa Daerah Indonesia*, 1.3 (2024), 1–9.

<sup>2</sup> Rosita Eka Sukmawati and Nanda Nini Anggalih, 'Perancangan Konten Media Sosial Tiktok Sebagai Promosi Desa Wisata Kampung Majapahit Bejjong', *Jurnal Barik*, 6.2 (2024), 199–213.

<sup>3</sup> Bashar Alhafni and others, 'Advancements in Arabic Grammatical Error Detection and Correction: An Empirical Investigation', *Journal Arxiv*, 2.05 (2023), 1–19.

<sup>4</sup> Mukhlis Amien, 'Sejarah Dan Perkembangan Teknik Natural Language Processing (Nlp) Bahasa Indonesia: Tinjauan Tentang Sejarah, Perkembangan Teknologi, Dan Aplikasi Nlp Dalam Bahasa Indonesia', *Journal Arxiv*, 1.2007 (2023), 1–7.

<sup>5</sup> Satibi, Subaiki Ikhwan, and Milki Aan, *Artificial Intelligence (AI) Dan Pendidikan Bahasa Arab: Sebuah Revolusi Pembelajaran Bahasa Arab*, 1st edn (Bogor: Abdi Fama Group, 2025).

learners, this platform has even shifted into a popular medium for autonomous learning.<sup>6</sup> However, beneath its efficiency, the utilization of automatic translation technology on languages with complex morphological and syntactic structures, such as Arabic, frequently triggers misunderstandings that can compromise the context of cross-cultural communication.

In the TikTok ecosystem, the auto-caption feature works by converting audio into text in real time, which is then supported by the automatic translation feature to reach a global audience.<sup>7</sup> However, the process of digitalizing Arabic audio into text frequently encounters technical failures in capturing diverse phonetic nuances and dialects. The NMT system on TikTok tends to generalize sentence structures, thereby often producing inaccurate transcriptions or rigid translations due to algorithmic limitations in understanding cultural context and Arabic linguistic complexity.<sup>8 9 10</sup>

Previous studies have extensively explored Arabic error analysis across various domains, such as academic texts,<sup>11</sup> thesis abstracts,<sup>12</sup> as well as vocabulary usage by university students.<sup>13 14</sup> Additionally, evaluations of AI applications such as Perplexity.ai<sup>15</sup> and Google Translate<sup>16</sup> demonstrating that

---

<sup>6</sup> Muhammad Afik Aminullah, Fadilah Al Azmi, and Darul Jalal, 'Pembelajaran Bahasa Arab Mandiri Melalui Platform Aplikasi Tiktok Sebagai Tren Belajar Masa Kini', *Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab Dan KebahasaAraban*, 5.2 (2022), 283–302.

<sup>7</sup> Dirahima Berida Shafira and Raden Arief Nugroho, 'Translation Error Types Analysis on TikTok Indonesian-English Auto-Translation Content', *Atlantis Pess*, 797 (2023), 45–51.

<sup>8</sup> Dzulkahfi, Herry Sujaini, and Tursina, 'Perbandingan Hasil Penerjemahan Neural Machine Translation ( NMT ) Dengan MarianNMT Terhadap Sumber Korpus Comparison of Neural Machine Translation ( NMT ) Results with MarianNMT Against Wikimedia and QED & TED', *JURISTI (Jurnal Riset Sains Dan Teknologi Informatika)*, 01.1 (2023), 151–59.

<sup>9</sup> Muhammad Ibnu Pamungkas and others, 'Error Analysis of Consonant Pronunciation In Arabic Language Among Fifth-Grade Students', *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 8.1 (2024), 56–72.

<sup>10</sup> Ismail Fahmi, 'Kesalahan Terjemahan Otomatis Tiktok Dalam Akun @Ahmad Rojih', *Maliki Interdisciplinary Journal (MIJ)*, 2.11 (2024), 1586–95.

<sup>11</sup> Gabriel Delpiter Tafonao and others, 'Analisis Kesalahan Gramatikal Dalam Teks Akademik Mahasiswa', *TA'EHAO: Jurnal Ilmiah Pendidikan Bahasa Dan Sastra Indonesia*, 1.1 (2026), 11–24.

<sup>12</sup> Rifqi Muhammad Firdaus, Rohanda Rohanda, and Dayudin Dayudin, 'A Linguistic Analysis of Arabic Language Errors in The Thesis Abstracts of Arabic Language and Literature Students at Uin Sunan Gunung Djati 2024', *International Conference on Language Learning and Literature (ICLL3)*, 55.3 (2025), 145–57.

<sup>13</sup> Muthiah Prasong and Pahlawan Pakro, 'Analisis Kesalahan Menulis Kosa Kata Bahasa Arab Mahasiswa Pusat Studi Bahasa Asing (Psba) Stkip Muhammadiyah Kalabahi Tahun Ajaran 2023/2024', *Syntax Literate: Jurnal Ilmiah Indonesia p-ISSN:*, 9.10 (2024), 5407–15.

<sup>14</sup> Nabila Rahmafillah and Ari Khairurrijal Fahmi, 'Syntactic Error Analysis in Writing Arabic Sentences at Tsanawiyah Level and Its Implications', *El-Ibtikar Jurnal Pendidikan Bahasa Arab*, 14.1 (2025), 38–54 <<https://doi.org/10.24235/ibtikar.v14i1.19940>>.

<sup>15</sup> Ahmad Zaimul Umam and others, 'Analysis of Translation Errors in Cnn Indonesia News Texts Into Arabic Through Perplexity.Ai Application', *JILSA: Jurnal Ilmu Linguistik & Sastra Arab*, 8.1 (2024), 40–52.

<sup>16</sup> Eka Lutfiyatun, Depi Kurniati, and Nurul Fajriah, 'Pemanfaatan Artificial Intelligence ( AI ) Dalam Tarjamah Dan Muhadatsah Di Perguruan Tinggi', *Seulanga: Jurnal Pendidikan Dan Pelatihan*, 2.2 (2023), 93–105.

automatic translation systems remain vulnerable to syntactic and semantic errors. On the TikTok platform itself, research has begun to address translation errors on specific accounts<sup>17</sup> and the effectiveness of translation techniques in news captions.<sup>18</sup> However, a highly specific research gap persists: there is an absence of an in-depth, integrated study mapping Arabic grammatical errors (*Nahwu* and *Sharf*) in TikTok's auto-caption feature generated in real-time from original Arabic audio. Therefore, this study adopts a case study approach on the @syukriarabs account to test the system's accuracy on content with clear audio articulation. The selection of the @syukriarabs account as the object of this case study is justified by the characteristics of its Arabic content, which features clear audio articulation and tends to utilize standard language (*Fusha*).

This gap triggers a conflict between Arabic syntactic/morphological theories and AI algorithms that tend to simplify structures. System failures frequently occur in distinguishing syntactic structures and sentence roles, resulting from the algorithm's failure to grasp context or speech rate. If left unaddressed without linguistic verification, a "normalization" of language errors will occur among digital users, which according to Wiwik Romauli (2024) poses a major linguistic challenge in the digital era.<sup>19</sup> Excessive reliance on automated systems without critical analysis will result in the degradation of Arabic language skills and the spread of misinformation.<sup>20</sup>

This study offers a cutting-edge solution by dissecting the flaws of TikTok's translation system through the lens of pure Arabic linguistics, with a primary focus on the uploads of the @syukriarabs account. Its novelty lies in the dynamic nature of the research object (audio-to-text) and the trending utilization of editing technologies such as CapCut, which are currently integrated into educational content creation.<sup>21</sup> By mapping error analysis research trends from 2013 to 2025,<sup>22</sup> this study positions itself as a bridge between technological needs and the rigor of linguistic rules.

Theoretically, this study is grounded in the foundation of Error Analysis pioneered by S.P. Corder, who views errors not merely as failures, but as a

---

<sup>17</sup> Fahmi.

<sup>18</sup> Farrah Chaiya Mas and Muhammad Yunus Anis, 'The Effect of Applying Arabic Translation Techniques on the Translation Quality Assessment of Al Jazeera Captions on TikTok Social Media', *Arabiyatuna*, 8.2 (2024), 509–36.

<sup>19</sup> Nababan and others.

<sup>20</sup> Muhammad Azhar and others, 'Al- Revolusi Pembelajaran Bahasa Arab Di Era Digital: Analisis Peran Artificial Intelligence Dalam Pengembangan Kompetensi Berbahasa Muhammad', *Al-Ma'luma*, 3.2 (2025), 58–81.

<sup>21</sup> Nurfitriyadi, 'Desain Model Problem Based Learning Menggunakan Aplikasi Capcut Pada Pembelajaran Ski Dalam Meningkatkan Hasil Belajar Siswa Kelas X Man 1 Majene' (Institut Agama Islam Negeri (IAIN) Parepare, 2024).

<sup>22</sup> Ucu Latipah and others, 'الاتجاهات في بحوث التحليل التقابلي والتحليل الأخطائي في تعليم اللغة العربية: مراجعة منهجية' (2013-2025)', *Ta'limi: Journal of Arabic Education & Arabic Studies*, 5.1 (2026), 189–217.

window into understanding the processes of language acquisition and processing.<sup>23</sup> This theory is integrated with the concepts of Neural Machine Translation (NMT) and Natural Language Processing (NLP) to dissect how artificial intelligence decodes the systems of *i'rab* (*Nahwu*) and *wazan* (*Sharf*).<sup>24</sup> The application of this paradigm enables the researcher to classify whether the inaccuracies in TikTok's auto-captions are categorized as interlingual errors (resulting from dialectal interference or native language) or intralingual errors caused by the internal complexity of the Arabic language rules themselves.<sup>25</sup> Thus, this theory functions not only as a tool for error identification but also as an evaluative framework for the performance of AI algorithmic logic in dealing with the flexibility of Arabic linguistics.

Based on the aforementioned exposition, this study is structured to address the following fundamental questions: How does TikTok's auto-caption feature negotiate linguistically with the structures of *i'rab* and *wazan* in the content of the @syukriarabs account, and to what extent do the failures of the Neural Machine Translation algorithm in processing these syntactic and morphological categories imply distortions in digital communication meaning? This question is designed to uncover not only the types of errors (such as the phonological or orthographic errors previously investigated by M. Ibnu Pamungkas<sup>26</sup> and Muthiah Prasong<sup>27</sup>), but also the logic behind the failures of the AI system in dealing with the flexibility of the Arabic language on social media.

The purpose of this study is to evaluate the grammatical accuracy of TikTok's auto-caption feature on the @syukriarabs account and to provide linguistic recommendations for developing more adaptive AI technology. The research hypothesis posits that the most significant errors lie within syntactic aspects (*Nahwu*) due to the failure to detect word roles in long sentences, which aligns with findings of syntactic errors in students' writing.<sup>28</sup> Practically, the significance of this study is expected to enhance user awareness regarding the importance of manual verification; theoretically, it aims to enrich the corpus of digital sociolinguistics within Arabic language education.<sup>29</sup>

## Method

This study uses a qualitative approach with a descriptive analytical design through case study methods to analyze linguistic errors in artificial intelligence

---

<sup>23</sup> S P Corder, *Error Analysis and Interlanguage*, 1981.

<sup>24</sup> Moh Syaifudin and others, *Linguistik Arab Di Era Digital: Teori, Praktik, Dan Kecerdasan Buatan* (Kuningan: Goresan Pena, 2025).

<sup>25</sup> Sanju Febrio Saragih and Selviana Napitupulu, 'Language Errors in Tiktok Content Captions', *Jurnal Ilmu Pendidikan Dan Sosial (JIPSI)*, 4.4 (2026), 979–87.

<sup>26</sup> Pamungkas and others.

<sup>27</sup> Prasong and Pakro.

<sup>28</sup> Rahmafillah and Fahmi.

<sup>29</sup> Moh Aziz Arifin, 'Tren Perkembangan Pendidikan Bahasa Arab Di Era Digital : Suatu Kajian Pustaka', *Lisan An-Natiq*, 72 (2025), 403–23.

technology. The researcher acted as a key instrument, fully present in the digital observation process, from determining the focus to drawing conclusions. This design was chosen because it provides a systematic overview of grammatical errors in the Arabic auto-caption feature on TikTok, which is dynamic. The researcher's presence is crucial for providing in-depth interpretation of the automatically generated text to ensure that the source of the error is purely the algorithm's response to the speaker's audio.

The data source for this study specifically comprises the Arabic auto-caption texts from the uploads of the TikTok account @syukriarabs. The sampling technique utilizes purposive sampling, with selection criteria focusing on videos that feature clear articulation and employ standard Arabic (*Fusha*). The choice of the @syukriarabs account is based on its content characteristics, which exhibit clear audio quality, making it ideal for testing the accuracy of the AI system in transcribing standard language rules. The analyzed data consists of 10 videos with a duration of at least 30–120 seconds uploaded within the last two years. Data collection was conducted over a two-month period within the digital realm of the TikTok platform using the *simak dan catat* (observation and note-taking) method, whereby the researcher documented findings in the form of screenshots for every grammatical anomaly that emerged.

Data analysis was conducted using a tabular instrument to comprehensively dissect the linguistic quality of TikTok's auto-caption feature. The operational details of the four primary points in the analysis table are as follows:

1. Audio Text (Original): This column contains the manual transcription performed by the researcher of the original speech from the @syukriarabs videos. This data serves as the baseline phonetic standard and the primary basis for comparison. By transcribing the original audio, the researcher can verify whether the AI errors were caused by the speaker's pronunciation or by an inherent failure of the system to capture the sound.
2. TikTok Auto-caption: This column captures the raw data consisting of the text generated automatically by the TikTok algorithm. The researcher copies this text exactly as it appears, preserving all typographical errors, *harakat* (vowel marks), and structural formations on the screen. This data serves as the material object of the research, whose accuracy rate will be evaluated.
3. Correct Rule (Standard Formulation): This column presents the text reconstruction adjusted to standard Arabic grammar rules (*Fusha*). This correction is based on authoritative literature, such as classical and contemporary treatises on *Nahwu* and *Sharf*. The presence of this column is crucial for demonstrating the gap or distance between the output of the AI engine and human linguistic standardization.

4. Error Classification: This column contains the technical classification of the detected anomalies. The researcher divides the errors into two broad categories:

- Syntactic Domain (*Nahwu*): Focuses on word position errors (*I'rab*), mismatches between *fi'il* (verb) and *fa'il* (subject), or the failure to detect *jumlah ismiyyah* (nominal sentences) and *jumlah fi'liyyah* (verbal sentences).
- Morphological Domain (*Sharf*): Focuses on word formation errors, modifications of root words, or inaccuracies in determining the correct *tashrif* (inflectional paradigms).

## Result and Discussion

Based on the analysis of 10 videos from the TikTok account @syukriarabs, it is revealed that TikTok's auto-caption feature still experiences various forms of linguistic errors at the syntactic (*Nahwu*), morphological (*Sharf*), phonetic, lexical, orthographic, and word-segmentation levels. These errors demonstrate that TikTok's Artificial Intelligence (AI) system based on Neural Machine Translation (NMT) is not yet fully capable of understanding the complexity of Arabic language structures, particularly regarding the aspects of *i'rab*, *wazan*, and the phonetic sounds of Arabic letters with close articulation points (*makhraj*).

The study found that the most common errors were in the phonetic and morphological domains. The auto-caption system frequently fails to distinguish letters with similar acoustic properties, such as *ṣād* (ص) and *sīn* (س), *qāf* (ق) and *kāf* (ك), as well as *dhāl* (ذ) and *dāl* (د). Furthermore, numerous errors were identified in the form of tanwin omission, pronoun shifts, particle deletion, and morphological alterations, all of which trigger significant distortions in meaning.

In general, the research results indicate that TikTok's auto-caption feature tends to oversimplify the grammatical structures of the Arabic language. This is evident in the omission of *harakat* (vowel marks), the reduction of syntactic particles, and word adjustments based on the acoustic proximity closest to the AI system. This phenomenon proves that TikTok's automated transcription algorithm remains highly oriented toward phonetic matching rather than a holistic understanding of the linguistic context.

To facilitate the identification of error patterns, the researcher categorizes the findings into several primary categories, as illustrated in the following table.

Table 1. Classification of TikTok Auto-caption Errors

| Error Category                 | Form of Error                                                      | Sample Findings                         |
|--------------------------------|--------------------------------------------------------------------|-----------------------------------------|
| Syntactic ( <i>Nahwu</i> )     | Omission of <i>tanwin</i> , particles, and <i>i'rab</i> mismatches | كثير ما المشكلات ← كثيرا<br>من المشكلات |
| Morphological ( <i>Sharf</i> ) | Alteration of <i>wazan</i> (word patterns) and pronoun forms       | المفرد علينا ← المفروض<br>علينا         |
| Phonetic                       | Misdetection of letter acoustic sounds                             | يعكد ← يحقد                             |
| Lexical                        | Word substitution altering the semantic meaning                    | عذب عنا الحزن ← اذهب<br>عنا الحزن       |
| Orthographic ( <i>Imla'</i> )  | Inaccuracies in letter and symbol rendering                        | 3 ← ثلاثة                               |
| Word Segmentation              | Erroneous splitting or merging of words                            | كانهم لكم ← كأنه ملك                    |

Based on the table above, phonetic errors constitute the category that appears most frequently. This indicates that TikTok's AI system encounters significant difficulties in capturing the phonological characteristics of the Arabic language, which is notably rich in pharyngeal/guttural sounds (*halqiyah*), velarized/emphatic letters (*tafkhim*), and highly subtle differences in articulation points (*makhraj*).

In addition, syntactic (*Nahwu*) errors were also found in large numbers, particularly concerning the omission of *tanwin manshub*, alterations of particles, and mismatches within the *mudhaf-mudhaf ilaih* (genitive constructive) structure. For instance, the phrase “كثيراً من المشكلات” was transcribed as “كثير ما المشكلات”. This error demonstrates that the system fails to understand the function of the particle “*min*” as a crucial syntactic element within Arabic sentence construction.

In the morphological aspect, TikTok's auto-caption frequently alters word patterns (*wazan*), thereby producing word forms with different meanings. For example, the word “المفروض” was changed to “المفرد”. Linguistically, these two words originate from different roots and possess unequal semantic functions. Such alterations prove that the AI system is not yet capable of accurately processing the derivational structures of Arabic words.

Another quite dominant finding is lexical errors. The auto-caption system several times produced words whose meanings completely contradicted the original text. In the data “اذهب عنا الحزن” which changed into “عذب عنا الحزن”, the meaning of the prayer "remove sadness" ironically changed into "torture



sadness". Such semantic distortions indicate that the AI errors are not only technical in nature, but can also seriously affect the understanding of meaning.

The results of the study show that the TikTok auto-caption feature still has significant limitations in processing Arabic linguistically. This finding aligns with the Error Analysis theory proposed by S.P. Corder,<sup>30</sup> that errors can be used to identify system weaknesses in processing a language. In the context of this study, the errors do not originate from human users, but rather from the AI system that fails to understand the complexity of Arabic language structures.

The dominance of phonetic errors in this study demonstrates that TikTok's speech recognition algorithm is not yet capable of accurately distinguishing the distinctive sounds of the Arabic language. Letters such as *qāf* (ق), *ṣād* (ص), *zā'* (ظ), and *dhāl* (ذ) frequently undergo substitution into letters that share similar articulation points (*makhraj*). This phenomenon shows that the AI system relies more on sound probability than contextual linguistic analysis. These findings align with the research by Bashar Alhafni et al., which states that Arabic Grammatical Error Detection technology still experiences difficulties in processing Arabic phonology and morphological structures.<sup>31</sup> Furthermore, this acoustic misdetection aligns with the findings of Ajidin et al. (2025) in *Albariq: Jurnal Pendidikan Bahasa Arab*, which demonstrated that phonetic variations and shifts in articulation points—such as in the realization of the letter *Qaaf*—frequently occur in digital audiovisual content, posing technical challenges for phonetic detection in automated platforms.<sup>32</sup>

Phonetic errors are also closely related to the characteristics of the Arabic language as a language that is highly sensitive to sound changes.<sup>33</sup> A single letter change can produce a drastic shift in meaning. In the data “يُحَقِّد” which changed into “يُكَدِّد”, the system fails to recognize the letters *ḥā'* (ح) and *qāf* (ق), which are guttural letters with complex phonetic characteristics. This finding reinforces the view of Dzulkahfi et al. that Neural Machine Translation (NMT) systems still experience weaknesses in handling languages with high morphological flexibility.<sup>34</sup>

Within the syntactic domain (*Nahwu*), this study finds that the auto-caption system frequently omits crucial syntactic elements such as *tanwin*, particles, and pronouns (*dhamir*). These errors demonstrate that the AI is not yet capable of

---

<sup>30</sup> Corder.

<sup>31</sup> Alhafni and others.

<sup>32</sup> Salsa Hauna Ajidin and others, 'Phonetic Variation of Qaf and Jim in Egyptian and Saudi Arabic Dialects: Insights from Youtube Audiovisual Data and Popular Music', *Albariq: Jurnal Pendidikan Bahasa Arab*, 6.02 (2025), 1–11 <<https://doi.org/10.24239/albariq.v6i02.123>>.

<sup>33</sup> Novel and Charsan Hidayat, 'Kesalahan Fonologi Dalam Bahasa Arab', *INNOVATIVE: Journal Of Social Science Research Volume*, 4.1 (2024), 7719–30.

<sup>34</sup> Dzulkahfi, Sujaini, and Tursina.

comprehensively understanding the grammatical function of each sentence element. The omission of the particle “أن” in the sentence “لا أستطيع أن أعيش” causes the sentence structure to change and disrupts the syntactic relationship between elements. In Arabic grammar rules, the particle “أن” has an important function in forming the *fi'il mudhari manshub*.<sup>35</sup> This inaccuracy proves that TikTok's system still operates at the level of sound transcription, rather than syntactic analysis.

The AI's limitation in processing these small elements aligns with the findings of Zulharby and Apriliani, who emphasize that inaccuracies in understanding changes to the function and meaning of particle elements, such as *harf jar*, become the main cause of intralingual errors that obscure the semantic construction of a sentence.<sup>36</sup>

This finding aligns with the research by Rahmafillah and Fahmi, which shows that syntactic errors frequently occur when the system or user fails to understand the relationships between Arabic sentence elements.<sup>37</sup> However, this study possesses novelty because the analyzed object is not human writing, but rather the results of real-time AI-based automated transcription.

In addition to syntactic errors, this study also finds a dominance of morphological (*Sharf*) errors. The auto-caption system several times alters word forms, pronouns (*dhamir*), and word derivation patterns, thereby producing different meanings. In the data “المفروض علينا” which changed into “المفرد علينا”, the system fails to recognize the *maf'ul* pattern and instead replaces it with the *mufrad* form. This error demonstrates that the AI algorithm is not yet capable of reading the relationships between root words and Arabic word formation patterns.

This phenomenon reinforces the research by Umam et al., which states that Arabic translation AI systems are still vulnerable to morphological distortion due to the complexity of *tashrif* and changes in word patterns.<sup>38</sup> In the Arabic language, a change in a single letter or a single diacritical mark (*harakat*) can completely shift the semantic function.<sup>39</sup> Therefore, morphological errors in TikTok's auto-caption cannot be considered minor errors, but rather fundamental errors that affect the meaning.

---

<sup>35</sup> Nur Hanifansyah and Menik Mahmudah, ‘Strategi Efektif Simplifikasi Nahwu Dalam Percakapan Arab Pada Kelas Intensif Daurah Ramadhan’, *Kuras: Journal of Arabic Education, Linguistics and Literature Studies*, 3.1 (2025), 35–48.

<sup>36</sup> Puti Zulharby and Analisa Apriliani, ‘Kesalahan Penerjemahan Harf Jar Pembentuk Idiom Pada Mahasiswa Pendidikan Bahasa Arab’, *Jurnal Alfazuna: Jurnal Pembelajaran Bahasa Arab Dan Kebahasaaraban*, 6.1 (2021), 112–34 <<https://doi.org/https://doi.org/10.15642/alfazuna.v6i1.1301>>.

<sup>37</sup> Rahmafillah and Fahmi.

<sup>38</sup> Umam and others.

<sup>39</sup> Samsul Bahri, ‘Analisis Fungsi l'râb Dalam Bahasa Arab Antara Semantis Dan Estetis’, *Al Qalam: Jurnal Ilmiah Keagamaan Dan Kemasyarakatan*, 17.1 (2023), 1–10.

This study also finds that lexical errors in TikTok's auto-caption tend to produce quite serious semantic distortions. For example, in the sentence “ اذهب عنا الحزن ” which was transcribed as “ عذب عنا الحزن ”. The replacement of the verb changes a positive meaning into a negative one. This indicates that the AI system does not comprehend the pragmatic and semantic context of the sentence as a whole.

This finding supports the research by Shafira and Nugroho, which explains that auto-translation technology on TikTok still tends to apply a literal approach based on word probability.<sup>40</sup> The system is not yet capable of understanding the emotional, cultural, and semantic contexts contained within Arabic speech.

On the other hand, the research results demonstrate that auto-caption errors increasingly escalate when the audio contains a high speech rate, emotional intonation, or the use of poetic diction. In several data pieces from Video 5 and Video 7, the system experienced total failure in recognizing complete phrases, thereby producing words that were far removed from the original text. This indicates that audio quality and speech complexity significantly affect AI performance.

Theoretically, this study reinforces the concept that Arabic is a language with high morphological and syntactic complexity, thereby requiring a more adaptive NLP approach. AI systems based on NMT that rely solely on statistical sound matching are insufficient for understanding the flexibility of Arabic language structures. Therefore, the development of AI technology for the Arabic language needs to integrate classical Arabic linguistic approaches, such as *i'rab* analysis, *wazan*, and semantic context.

Practically, this study provides important implications for social media users, educators, and digital technology developers. For TikTok users, the results of this study indicate the importance of manual verification regarding the auto-caption feature, particularly in Arabic language educational content. Total dependence on automated captions has the potential to normalize language errors in the digital space.

For Arabic language educators, this study can serve as evaluation material to enhance students' digital literacy so that they become more critical toward the use of AI in language learning. Educators need to guide learners to combine the learning process with interactive digital media that is more measurable in vocabulary mastery, as emphasized by Nurchasanah and Fahmi that the utilization of adaptive and interactive digital platforms such as Wordwall is capable of effectively designing the *mufrodhat* learning process while

---

<sup>40</sup> Shafira and Nugroho.

simultaneously boosting student learning interest without getting trapped in structural errors.<sup>41</sup>

This study also reinforces the view of Nababan et al. that the development of digital technology brings major challenges to the quality of language use on social media.<sup>42</sup> If auto-caption errors are continuously left without linguistic evaluation, a phenomenon of normalizing Arabic language errors will emerge in the digital public space.

Nevertheless, this study possesses several limitations. First, the research only focuses on one TikTok account, namely @syukriarabs, so the results cannot yet be generalized to all Arabic content on TikTok. Second, this study only analyzes the auto-caption feature without comparing it with other platforms such as YouTube Auto-caption or Google Speech Recognition.

### Conclusion and Suggestion

Based on the analysis of the Artificial Intelligence (AI)-based auto-caption feature on the TikTok account @syukriarabs, this study concludes that the Neural Machine Translation (NMT) and speech recognition systems employed still exhibit significant limitations in processing standard Arabic (*Fusha*). Although the initial hypothesis predicted that the most substantial errors would reside within the syntactic domain, empirical findings demonstrate that the most dominant errors ironically occur in the phonetic domain—due to the system's failure to distinguish letters with close articulation points (*makhraj*)—as well as the morphological domain (*Sharf*), which alters word patterns (*wazan*) and pronouns (*dhamir*). Within the syntactic domain (*Nahwu*), the AI's failure materializes in the form of omitting crucial elements such as *tanwin manshub*, particles (*harf*), and inaccurate *i'rab* in long sentences. This failure logic reveals that TikTok's algorithm is oriented more toward a literal, probability-based approach and statistical sound matching rather than holistic contextual understanding, which ultimately triggers serious semantic distortions where the meaning of the generated text shifts drastically or even contradicts the original audio.

Theoretically, this study supports S.P. Corder's Error Analysis theory, which posits that errors in automated systems can reveal the limitations of technology in processing languages with high morphological flexibility. Practically, these findings carry important implications regarding the urgency of manual verification by users to stem the risk of "normalizing" Arabic language

---

<sup>41</sup> Maryam Nurchasanah and Ari Khairurrijal Fahmi, 'Pengaruh Wordwall Dalam Meningkatkan Minat Belajar Kosakata Bahasa Arab Pada Siswa Kelas VIII Di MTs Al Muddatsiriyah Jakarta', *Al-Fusha: Arabic Language Education Journal*, 6.1 (2024), 37–43  
<<https://doi.org/https://doi.org/10.62097/alfusha.v6i1.1528>>.

<sup>42</sup> Nababan and others.

errors, which could degrade digital literacy in the public sphere. However, this study has limitations due to the specific scope of its object, focusing only on a single educational account with a sample of 10 videos. Therefore, future researchers are suggested to expand the data coverage by conducting comparative studies across digital platforms such as YouTube Auto-caption or by testing the effectiveness of AI systems on different variations of Arabic dialects ('*Ammiyah*).

## References

- Ajidin, Salsa Hauna, Yanda Rismawati, Asep Ajidin, and Farhan Alfaruq Ajidin, 'Phonetic Variation of Qaf and Jim in Egyptian and Saudi Arabic Dialects: Insights from Youtube Audiovisual Data and Popular Music', *Albariq: Jurnal Pendidikan Bahasa Arab*, 6 (2025), 1–11 <<https://doi.org/10.24239/albariq.v6i02.123>>
- Alhafni, Bashar, Go Inoue, Christian Khairallah, and Nizar Habash, 'Advancements in Arabic Grammatical Error Detection and Correction: An Empirical Investigation', *Journal Arxiv*, 2 (2023), 1–19
- Amien, Mukhlis, 'Sejarah Dan Perkembangan Teknik Natural Language Processing (Nlp) Bahasa Indonesia: Tinjauan Tentang Sejarah, Perkembangan Teknologi, Dan Aplikasi Nlp Dalam Bahasa Indonesia', *Journal Arxiv*, 1 (2023), 1–7
- Aminullah, Muhammad Afiq, Fadilah Al Azmi, and Darul Jalal, 'Pembelajaran Bahasa Arab Mandiri Melalui Platform Aplikasi Tiktok Sebagai Tren Belajar Masa Kini', *Al Mi'yar: Jurnal Ilmiah Pembelajaran Bahasa Arab Dan KebahasaAraban*, 5 (2022), 283–302 <<https://doi.org/10.35931/am.v5i2.1219>>
- Arifin, Moh Aziz, 'Tren Perkembangan Pendidikan Bahasa Arab Di Era Digital : Suatu Kajian Pustaka', *Lisan An-Natiq*, 72 (2025), 403–23
- Azhar, Muhammad, Destia Yolanda, Angga Frananda, M. Ripani Saputra, Resy Mulyani, and Siti Nurdinah, 'Al- Revolusi Pembelajaran Bahasa Arab Di Era Digital: Analisis Peran Artificial Intelligence Dalam Pengembangan Kompetensi Berbahasa Muhammad', *Al-Ma'luma*, 3 (2025), 58–81
- Bahri, Samsul, 'Analisis Fungsi I'râb Dalam Bahasa Arab Antara Semantis Dan Estetis', *Al Qalam: Jurnal Ilmiah Keagamaan Dan Kemasyarakatan*, 17 (2023), 1–10
- Corder, S P, *Error Analysis and Interlanguage*, 1981
- Dzulkahfi, Herry Sujaini, and Tursina, 'Perbandingan Hasil Penerjemahan Neural Machine Translation ( NMT ) Dengan MarianNMT Terhadap Sumber Korpus Comparison of Neural Machine Translation ( NMT ) Results with MarianNMT Against Wikimedia and QED & TED', *JURISTI (Jurnal Riset Sains Dan Teknologi Informatika)*, 01 (2023), 151–59 <<https://doi.org/10.26418/juristi.v1i1.61146>>
- Fahmi, Ismail, 'Kesalahan Terjemahan Otomatis Tiktok Dalam Akun @Ahmad Rojih', *Maliki Interdisciplinary Journal (MIJ)*, 2 (2024), 1586–95
- Firdaus, Rifqi Muhammad, Rohanda Rohanda, and Dayudin Dayudin, 'A Linguistic

- Analysis of Arabic Language Errors in The Thesis Abstracts of Arabic Language and Literature Students at Uin Sunan Gunung Djati 2024', *International Conference on Language Learning and Literature (ICL3)*, 55 (2025), 145–57
- Hanifansyah, Nur, and Menik Mahmudah, 'Strategi Efektif Simplifikasi Nahwu Dalam Percakapan Arab Pada Kelas Intensif Daurah Ramadhan', *Kuras: Journal of Arabic Education, Linguistics and Literature Studies*, 3 (2025), 35–48
- Latipah, Ucu, Viqri Aditya, Ahmad Bukhori Muslim, Koderi, Zulhannan, and Amiruddin, 'الاتجاهات في بحوث التحليل التقابلي والتحليل الأخطائي في تعليم اللغة العربية: مراجعة منهجية' (2013-2025)', *Ta'limi: Journal of Arabic Education & Arabic Studies*, 5 (2026), 189–217
- Lutfiyatun, Eka, Depi Kurniati, and Nurul Fajriah, 'Pemanfaatan Artificial Intelligence (AI) Dalam Tarjamah Dan Muhadatsah Di Perguruan Tinggi', *Seulanga: Jurnal Pendidikan Dan Pelatihan*, 2 (2023), 93–105
- Mas, Farrah Chaiya, and Muhammad Yunus Anis, 'The Effect of Applying Arabic Translation Techniques on the Translation Quality Assessment of Al Jazeera Captions on TikTok Social Media', *Arabiyatuna*, 8 (2024), 509–36
- Nababan, Wiwik Romauli, Nadia Rahmadani, Wisdom Ortiz Valencia Tamba, and Taufiq Kurniawan Hidayat Nst, 'Tantangan Bahasa Di Era Digital Terhadap Kesalahan Berbahasa Dalam Komunikasi Media Sosial', *Jurnal Bahasa Daerah Indonesia*, 1 (2024), 1–9 <<https://doi.org/10.47134/jbdi.v1i3.2602>>
- Novel, and Charsan Hidayat, 'Kesalahan Fonologi Dalam Bahasa Arab', *INNOVATIVE: Journal Of Social Science Research Volume*, 4 (2024), 7719–30
- Nurchasanah, Maryam, and Ari Khairurrijal Fahmi, 'Pengaruh Wordwall Dalam Meningkatkan Minat Belajar Kosakata Bahasa Arab Pada Siswa Kelas VIII Di MTs Al Muddatsiriyah Jakarta', *Al-Fusha: Arabic Language Education Journal*, 6 (2024), 37–43 <<https://doi.org/https://doi.org/10.62097/alfusha.v6i1.1528>>
- Nurfitrayadi, 'Desain Model Problem Based Learning Menggunakan Aplikasi Capcut Pada Pembelajaran Ski Dalam Meningkatkan Hasil Belajar Siswa Kelas X Man 1 Majene' (Institut Agama Islam Negeri (IAIN) Parepare, 2024)
- Pamungkas, Muhammad Ibnu, Dede Ahmad Satibi, Alwi Zakhratunnisa, and Aas Lestari, 'Error Analysis of Consonant Pronunciation In Arabic Language Among Fifth-Grade Students', *Ta'lim Al-'Arabiyyah: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 8 (2024), 56–72
- Prasong, Muthiah, and Pahlawan Pakro, 'Analisis Kesalahan Menulis Kosa Kata Bahasa Arab Mahasiswa Pusat Studi Bahasa Asing (Psba) Stkip Muhammadiyah Kalabahi Tahun Ajaran 2023/2024', *Syntax Literate: Jurnal Ilmiah Indonesia p-ISSN:*, 9 (2024), 5407–15
- Rahmafillah, Nabila, and Ari Khairurrijal Fahmi, 'Syntactic Error Analysis in Writing Arabic Sentences at Tsanawiyah Level and Its Implications', *El-Ibtikar Jurnal Pendidikan Bahasa Arab*, 14 (2025), 38–54 <<https://doi.org/10.24235/ibtikar.v14i1.19940>>
- Saragih, Sanju Febrio, and Selviana Napitupulu, 'Language Errors in Tiktok Content Captions', *Jurnal Imu Pendidikan Dan Sosial (JIPSI)*, 4 (2026), 979–87

- <<https://doi.org/10.58540/jipsi.v4i4.1327>>
- Satibi, Subaiki Ikhwan, and Milki Aan, *Artificial Intelligence (AI) Dan Pendidikan Bahasa Arab: Sebuah Revolusi Pembelajaran Bahasa Arab*, 1st edn (Bogor: Abdi Fama Group, 2025)
- Shafira, Dirahima Berida, and Raden Arief Nugroho, 'Translation Error Types Analysis on TikTok Indonesian-English Auto-Translation Content', *Atlantis Pess*, 797 (2023), 45–51 <<https://doi.org/10.2991/978-2-38476-162-3>>
- Sukmawati, Rosita Eka, and Nanda Nini Anggalih, 'Perancangan Konten Media Sosial Tiktok Sebagai Promosi Desa Wisata Kampung Majapahit Bejijong', *Jurnal Barik*, 6 (2024), 199–213
- Syaifudin, Moh, Nafis Azmi Amrullah, M. Isroul Laili, and Kiki Cahya Muslimah, *Linguistik Arab Di Era Digital: Teori, Praktik, Dan Kecerdasan Buatan* (Kuningan: Goresan Pena, 2025)
- Tafonao, Gabriel Delpiter, Desta Nataring Ge, Sinarti Ningsih Batee, and Arozatulo Bawamenewi, 'Analisis Kesalahan Gramatikal Dalam Teks Akademik Mahasiswa', *TA'EHAO: Jurnal Ilmiah Pendidikan Bahasa Dan Sastra Indonesia*, 1 (2026), 11–24
- Umam, Ahmad Zaimul, Akhmad Ridlo, Mohamad Zaka Al Farisi, and Mad Ali, 'Analysis of Translation Errors in Cnn Indonesia News Texts Into Arabic Through Perplexity.Ai Application', *JILSA: Jurnal Ilmu Linguistik & Sastra Arab*, 8 (2024), 40–52
- Zulharby, Puti, and Analisa Apriliani, 'Kesalahan Penerjemahan Harf Jar Pembentuk Idiom Pada Mahasiswa Pendidikan Bahasa Arab', *Jurnal Alfazuna: Jurnal Pembelajaran Bahasa Arab Dan Kebahasaaraban*, 6 (2021), 112–34 <<https://doi.org/https://doi.org/10.15642/alfazuna.v6i1.1301>>