International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 04 | Apr 2026 www.irjet.net p-ISSN: 2395-0072

An Enhanced Arabic Flickr8K Dataset for Image Captioning and Sentence Correction

Hadeel Abdulrahman[1], Mohammed Fadel Sukkar[2]

[1]*PhD Student, Department of Artificial Intelligence and Natural Languages, Faculty of Informatics Engineering, University of Aleppo*

[2]*Professor, Department of Artificial Intelligence and Natural Languages, Faculty of Informatics Engineering, University of Aleppo*

Abstract – *Arabic Natural Language Processing (NLP) applications suffer from a scarcity of high-quality data, especially in tasks involving linguistic error correction and Arabic image captioning. In this research, we present a framework designed to enhance the quality of Arabic texts by constructing a dataset that has been precisely corrected for linguistic, grammatical, and orthographic errors. This was achieved by using the Arabic-translated version of the Flickr8k dataset that originally used in image captioning task, where the data underwent manual auditing and correction, followed by a subsequent review by an expert in the Arabic language to ensure accuracy and reliability. Furthermore, we provide a comprehensive statistical analysis of all errors identified in the original dataset, classifying them into five categories: grammatical, orthographic, linguistic, translation errors and captions redundancy. Utilizing this data, we fine-tuned the AraT5 and mT5 models to address the task of Arabic text correction within a text-to-text transformer. Additionally, a data augmentation was performed on the main dataset to enhance the models' generalization capabilities. To evaluate the models' performance, we employed the BLEU, CIDEr, METEOR, and AraBERT metrics. The results demonstrated that the utilization of the corrected and augmented data led to a significant improvement in model performance. Furthermore, we observed a slight improvement in the results of the AraT5 model over the mT5 model. This study contributes a high-quality Arabic linguistic resource applicable to various NLP tasks while simultaneously underscoring the critical importance of data quality in enhancing the performance of Arabic language models.*

***Key Words*: Arabic Image Captioning, Flickr8k, Arabic Text correction, Data Augmentation, AraT5, mT5.**

1. INTRODUCTION

The quality of textual data plays a main role in the efficacy of NLP models, particularly with the Arabic language, which is distinguished by its rich morphology and structural complexity. The complexity of the language is evident in a variety of aspects, including grammar, orthography and diacritics. These elements have a direct impact on the semantic meaning of the language. Additionally, the existence of multiple forms for a single word and the wide variation in dialects increase the challenge. All these features make the process of generating a high-quality dataset extremely important as any problem in data integrity can negatively impact model performance [1]. Many Arabic NLP fields, such as text generation, machine translation, and image captioning, rely on datasets translated from other languages, especially English. For instance, in the domain of Arabic image captioning, the most widely used datasets include Flickr8k [2], MSCOCO [3], and Multi30k [4]; however, all of these have been translated into Arabic either automatically or semi-automatically. This process frequently results in various linguistic errors [5], such as unnatural syntactic structures, context-insensitive literal translations, and spelling mistakes. Additionally, semantic issues often arise due to the loss or alteration of the original meaning during translation, as a result, the generated texts are of lower quality and less representative of natural Arabic. These challenges directly affect the training of NLP models [6]. Modern models, particularly transformer-based architecture, such as T5 models heavily rely on the quality of input data. In the event of the training data being noisy or inaccurate, the model will learn weak representations, resulting in outputs that are incorrect or even linguistically unnatural. This finding emphasizes the importance of verifying the data prior to its utilization in training models, rather than focusing simply on augmenting the size of the dataset. This study aims to address these challenges by constructing a manually verified Arabic dataset that has been manually corrected for linguistic, grammatical, and orthographic levels. The first 2,001 Arabic captions from the Flickr8k dataset were selected and refined through multiple stages. This involved correcting grammatical and orthographic errors, translation errors, eliminating redundancies, and standardizing sentence structures to ensure style consistency. Additionally, we enhanced the text by adding appropriate diacritical marks (Tashkeel) at the end of each word in each sentence, which reduces semantic ambiguity and enables models to learn grammatical relationships with high precision. Following the dataset construction phase, the refined data were used to fine-tune the AraT5[7] and mT5 [8] models, which are based on the T5 [9] architecture, to correct input Arabic sentences and generate grammatically correct output (Arabic text correction). The significance of this approach lies not only in improving the performance of existing models but also in generalizing to new Arabic texts containing similar types of errors. This work highlights the importance of improving

Arabic linguistic data. So, our main contributions can be outlined as follows:

1. The creation of an Arabic dataset of high quality, with linguistic and grammatical corrections, that is suitable for use in multiple Arabic natural language processing (NLP) applications.

2. A statistical analysis of errors within the original Arabic Flickr8k data, categorized by type into grammatical, orthographic, linguistic, and translation errors.

3. The utilization of the enhanced dataset to fine-tune the AraT5 and mT5 models for the automated correction of Arabic texts.

4. An evaluation of model performance using the BLEU, CIDEr, METEOR, and AraBERT to assess textual integrity and semantic accuracy. This evaluation was supported by a qualitative analysis of the resulting sentences.

2. Related Works

In this section, we review the datasets used for Arabic image captioning, as well as the T5 models that support the Arabic language. Arabic image captioning is still till now an active area of research; indeed, the process of creating a large-scale and high-quality annotated dataset is still ongoing. For instance, the aim of ImageEval2025 competition [10] was divided into two primary objectives: first, the creation of a large dataset for Arabic image captioning, and second, the development of the most effective model for this task. In addition, three datasets are currently used as benchmarks for the task of Arabic image captioning: The Arabic Flickr8k dataset is based on the English dataset Flickr8k [11], which consists of 8,000 images of a variety of scenes involving people and animals, each image paired with five English captions. Based on it, Eljundi created the Arabic version by translating three of the image captions into Arabic using the Google Translate API. This stage was followed by a manual review of the translated captions by a specialist to correct any contextual errors that may have occurred during translation. After that, the best three translated captions were selected from the original set of five. MSCOCO, translated from [12], this dataset comprises 123,287 images, each image paired with five English captions. The captions were translated automatically, without any human review. Multi30K, an extension of the Flickr30k dataset [13], this dataset consists of 31,014 images, each paired with five English captions. One of these captions was translated by human experts into German, French, Czech, and Arabic. Despite the importance of these resources, several studies have shown that translated datasets often suffer from poor semantic and linguistic quality, particularly when converting texts into languages with complex structural characteristics, such as Arabic. In many cases, the translated data lacks rigorous linguistic and grammatical accuracy, which reduces its effectiveness for tasks like text correction or Arabic image captioning, this can lead to the accumulation of errors, whether originating from inaccurate translations or from models trained on such noisy data [14].

Among the most used models for text generation is T5, a pre-trained language model designed for text-to-text transformation tasks. T5 is employed in various language generation tasks such as machine translation and text summarization, where both the model's input and output consist of text. T5 architecture is based on the encoder-decoder Transformer. Building on the T5 architecture, the mT5 model was introduced as a multilingual model that is trained on over 100 languages including Arabic, whereas the AraT5 model is trained exclusively on the Arabic language.

3. Types of Sentences and Errors in the Arabic Language

Arabic sentences can be classified into two categories [15]: Nominal sentences and Verbal sentences. A nominal sentence generally consists of two main components: the subject (Mubtada مبتدأ) and the predicate (Khabar خبر). The subject is usually a noun, while the predicate may take the form of a noun or a verbal sentence. Nominal sentences may also begin with particles such as Inna and its sister (إن وأخواتها) or Kan and its sister (كان وأخواتها). On the other hand, verbal sentence, consists of a verb and a subject (Fa'il فاعل), and verb can be transitive, requiring one or more direct objects. Simple sentences can evolve into compound sentences when two or more simple sentences are linked together using conjunctions.
Errors in Arabic sentences can be categorized into five primary types [16]:

- Orthographic Errors: These refer to the incorrect orthography of words whether due to confusion between similar letters like, errors involving the Hamza [ء, ؤ, ئ], the letter Ha versus the Ta' Marbuta [ه, ة], or between Alif [ا] versus the Alif Maqsura [ى], as well as errors distinguishing between the letter Nun [ن] and Tanwin [ً] nunation. Other types include incorrect letter order, or the addition or omission of letters.
- Morphological Errors: These relate to changes in word structure, such as the addition or omission of letters, which can affect meaning or syntax. Morphological errors also include mistakes in gender (masculine/feminine) and definiteness (definite/indefinite articles), verb forms, or incorrect inflections for nouns and pronouns [17].
- Syntactic Errors: These occur when words are incorrectly formed or written due to violations of grammar rules. Examples include errors in case diacritics, errors in genitive, accusative, and nominative, and errors in grammatical number agreement (e.g., تسعة طالبات instead of تسع طالبات) [17].
- Semantic Errors: occur when a word is inappropriate in meaning or context, such as

selecting an unsuitable word or misusing a conjunction, which can affect the correct meaning of the sentence.

- Punctuation Errors: These involve the insertion of unnecessary marks, omission of necessary marks, or the substitution of one mark for another, all of which affect the clarity of the sentence.

4. The Enhanced Arabic Flickr8k Dataset

This section will address the problems of the original Flickr8k database and the proposed enhanced database. Dataset available at: (https://www.kaggle.com/datasets/hadeelabdr/enhanced-arabic-flickr8k).

4.1 The Arabic Flickr8k issues:

The Arabic version of the Flickr8k dataset has several issues, including grammatical, orthographic, translation, and semantic problems. These issues have the following consequences:

- Low data quality: Grammatical and orthographic errors alter the original meaning of captions, making them inconsistent with the image and difficult for readers to understand correctly. This has a negative impact on the automated image captioning process.
- Inaccurate Image captions: Many image captions in Flickr8k suffer from machine translation of English captions, sometimes resulting in inconsistency between the image and its caption.
- Duplicate captions for the Same Image: These captions do not offer any new features that could be used to train the image caption model.

4.2 Development and Refinement of the Arabic Flickr8k Dataset

Within the scope of this research, we developed an enhanced Arabic image captioning dataset from the Arabic Flickr8k, adhering to precise grammatical, orthographic, and semantic standards. The descriptions for the first 667 images, comprising the 2001 captions, were manually reviewed by the researcher to identify and correct errors. This was followed by a second proofreading process by an expert in the Arabic language from the Faculty of Arts and Humanities at the University of Aleppo, thus enhancing its reliability as a tool to support related research and applications. The review and proofreading process revealed the following problems:

Translation errors: Many captions contained translation errors, totaling 643. These errors are divided into two types, the first is an error in a word or term, and the second is the possibility of using another synonym to improve the meaning of the sentence in Arabic

- "A white and brown dog" translated as "a white and a white dog كلب أبيض و أبيض"
- "A multiracial couple posing for a picture" translated as "a multiracial couple representing for a picture زوجين متعددي الأعراق يمثلان لصورة"
- "A large white dog lying on the floor" translated as "a large white dog thrown on the ground كلب أبيض كبير ملقى على الأرض"
- "A dog runs on the green grass near a wooden fence" translated as "a dog stretches on the green grass near a wooden fence كلب يمتد على العشب الأخضر بالقرب من سياج خشبي"

Grammatical errors, 195 errors were identified, including grammatical errors or ambiguity in the grammatical error. Such as:

- Two dogs play together on the beach. " اثنين من الكلاب يلعبون معا على الشاطئ", incorrect agreement in number and verb form
- A little girl. "فتاة الصغيرة", incomplete sentence.
- Two girls are skiing. "فتاتان تتزحلق", incorrect verb agreement.
- A woman is talking. "امرأة يتحدث"
- brown dogs are crossing the water. " كلاب بنيتين تمر عبر الماء", incorrect dual/plural agreement

Linguistic errors, which contain non-Arabic words, totaled 64.

- Bungee cords. "حبال البانجي"
- A woman writing a note "امرأة تكتب نوتة"

Spelling errors:

- منطقة حرجية (misspelled form)
- رجل بدون قميص يسير على لكثير من الصخور(incorrect spelling لكثير)
- صبي ينظف الفقاعات من وجهة (using Ta' Marbuta instead of ha)

Caption redundancy, a total of 11 duplicated captions as in image "1193116658_c0161c35b5", the caption " فتاة تركب الدراجة الهوائية الأحادية العجلة" is repeated twice.

To address these issues in the Flicker8k database, it was necessary to restore the Arabic text (image captions) to its correct Arabic linguistic form, following these steps:

1.Proofread and correct these captions to eliminate grammatical, orthographic, and redundancy-related errors.

2. Standardize the sentences by placing the verb at the beginning of the sentence to reduce syntactic ambiguity. "فتاة صغيرة مغطاة بالطلاء تجلس أمام قوس قزح" became "تجلس فتاة صغيرة مغطاة بالطلاء أمام قوس قزح"

3. Adding diacritics to the ends of words in each sentence. This stage was followed by a proofreading process by an Arabic language specialist. To illustrate the impact of the proposed refinement process, (Table 1) presents examples of image captions before and after correction. The refined captions demonstrate improvements in grammatical accuracy, lexical choice, and syntactic clarity.

Table- 1: Caption before and after modification process

Original Image caption	Image caption in English	Modified Image caption

طفلة صغيرة تتسلق إلى مسرح خشبي	A little girl climbing into a wooden playhouse	تصعدُ طفلةٌ صغيرةٌ إلى بيتِ اللعبِ الخشبي
طفلة صغيرة تتسلق الدرج إلى منزلها	A little girl climbing the stairs to her playhouse	تصعدُ طفلةٌ صغيرةٌ الدرجَ إلى بيتِ اللعبِ
فتاة صغيرة في ثوب وردي تذهب إلى المقصورة الخشبية	A little girl in a pink dress going into a wooden cabin	ترتدي فتاةٌ صغيرةٌ ثوبا ورديا وتذهبُ إلى المقصورةِ الخشبيةِ

4.3 Data Preparation and Augmentation

To achieve optimal results and enhance the models' generalizability, we implemented data augmentation to increase the dataset size, which consists of 2001 Arabic sentences. This augmentation was applied only to the original sentences before proofreading, following these steps:

1. Creating synonym groups for some verbs found in the sentences using Arabic WordNet [18], in addition to synonyms found in the Arabic captions. This resulted in 74 groups, separating the masculine singular synonyms for each verb into separate groups, as well as the feminine singular, masculine dual, feminine dual, and masculine and feminine plurals. Examples of these groups include:
   - تلبس, ترتدي
   - ينظر, يشاهد, يراقب

2. Creating noun synonym groups, which includes 27 synonym groups for nouns, as in the examples:
   - العائلة, الأسرة
   - الشاب, الرجل, الشخص

3. Creating adjective synonym groups, which results in 8 synonym groups for adjectives:
   - الساخن, الحار
   - كبيرة, ضخمة

Finding permutations of verbs, nouns, and adjectives in the sentences of the original pre-proofread sentences, therefore, the number of Arabic sentences increased to 6103.

For the sentence "طفلة صغيرة تتسلق إلى مسرح خشبي," we obtained three sentences: "فتاة صغيرة تتسلق إلى مسرح خشبي," "طفلة صغيرة تتسلق إلى مسرح خشبي," and "طفلة صغيرة تصعد إلى مسرح خشبي".

To reduce potential ambiguity for the model, explicit instructions were included with each input sentence, using a directive that guide the model to rewrite the sentence to ensure agreement between the verb and subject in terms of gender and number.

5. Our Experiment

5.1 Fine-Tuning process

During the experiment, we performed fine-tuning of the AraT5 and mT5 models on our enhanced dataset, with the objective of developing a model capable of automatically correcting input Arabic sentences. This was achieved by fine-tuning both models on the dataset before and after the proofreading and correction process. A set of evaluation criteria was used to measure the performance, including BLEU, METEOR, CIDEr, and AraBERT, in addition to qualitative analysis comparing the generated sentences with reference sentences to evaluate grammatical correctness and semantic accuracy

5.2 Experiment Setup

The experiment was conducted on the Kaggle platform using a GPU P100. Initially, we used each model's tokenizer to obtain the numerical representation of the text. To ensure consistency, sequence lengths were set to 64 tokens, corresponding to the caption lengths in the dataset. Padding was applied to captions shorter than 64 tokens. The dataset was divided into two parts: 80% for training and 20% for testing. The models were tested on the test set for evaluation. Sentences were generated sequentially using beam search with a beam size of five, selecting the optimal sentence each time. Additional constraints were applied to improve the quality of the text and reduce repetition of n-grams within each generated sentence.

5. Results and discussion

We conducted six separate experiments to evaluate the performance of the models, using a set of metrics to assess the results: BLEU, CIDEr, METEOR, and AraBERT, in addition to qualitative analysis. In the first three experiments, the mT5 model was used. The first experiment involved evaluating the mT5 model without fine-tuning, while the second and third experiments involved fine-tuning the model using the dataset of 2001 sentence pairs and 6103 sentence pairs, which were obtained after data augmentation. The same experiments were repeated for the AraT5 model, and the results are shown in (Table 2).

The results show that the models without fine-tuning are insufficient for performing Arabic sentence correction, with generally poor performance, especially in the AraT5 model. In contrast, fine-tuning led to a significant improvement in all metrics. Additionally, the dataset augmentation from 2001 to 6103 sentence pairs further enhanced performance, reflecting the importance of data size in enhancing the ability of models to correct errors and generate correct Arabic sentences.

After training models on the augmented data, the performance of the two models was very similar, with a slight advantage observed for the AraT5 over the mT5 model. Furthermore, an analysis of the evaluation criteria showed a significant increase in the CIDEr value, reaching 8.8420, indicating a high degree of similarity between the generated and reference texts. Meanwhile, the METEOR value was 0.9330, demonstrating strong semantic similarity between the generated and reference texts. The BLEU scores were also high with BLEU1, BLEU2, BLEU3, BLEU4 reaching 0.9387,

0.9197, 0.9021, 0.8754, respectively, shows a strong n-gram overlap with reference texts. However, the AraBert was a little bit higher value in mT5 than AraT5.

Table -2: mT5 and AraT5 Results

mT5 Results			
Metric	**Baseline**	**Size_2001**	**Size_6103**
B1	0.1459	0.7931	0.9348
B2	0.0761	0.7247	0.9114
B3	0.0512	0.6499	0.8879
B4	0.0423	0.5742	0.8550
METEOR	0.1314	0.7916	0.9274
CIDEr	0.4690	6.1086	8.6779
AraBERT	0.4710	0.9419	**0.9827**
AraT5 Results			
Metric	**Baseline**	**Size_2001**	**Size_6103**
B1	0	0.7180	**0.9387**
B2	0	0.6369	**0.9197**
B3	0	0.5523	**0.9021**
B4	0	0.4828	**0.8754**
METEOR	0	0.7054	**0.9330**
CIDEr	0	5.0555	**8.8420**
AraBERT	0.3708	0.9078	0.9820

Some examples are shown in (Table 3), that generated from each experiment. For the mT5 Base model, the first and third generated sentences were incomplete, while the second sentence was completely different from the input sentence, and the model failed to remove the special token<extra_id_0> because the training data was not processed correctly. In contrast, the AraT5 Base model completely failed in generating sentences and displaying random Arabic and English words and tokens, unrelated to the input sentence.

After fine-tuning the mT5 model on 2001 sentence pairs, a clear improvement was observed in word order and the overall grammatical, orthographic, and morphological quality of the sentences. However, some minor problems remained, such as in the third sentence, where semantic reduction occurred due to the omission of the verb "put the bag, يضع الحقيبة," in addition to the incorrect choice of the preposition "on public transport, على وسيلة النقل العام" which would have been better grammatically as "in public transport, في وسيلة النقل العام". In contrast, AraT5's results after training on 2001 sentences were observed to be of lower quality than mT5. Some morphological errors appeared, such as the omission of the definite article in the first sentence, "The dogs are gray, الكلاب رمادية", in addition to a semantic error in the third sentence where the verb "to sip, ترشف" was replaced with "to sprinkle, يرش" and a gender agreement error where "يقرأ" was used instead of "تقرأ".

After training on 6103 pairs of Arabic sentences, mT5's results were mostly good, with some problems, such as in the first sentence, where "يقضم" would have been semantically better than "يأكل", and the dog's color was changed from white to brown. The results for the second and third sentences were grammatically correct. As for AraT5, the model successfully transformed "to hit a toy, تضرب على لعبة" into "to fight for a toy, تتقاتل من أجل لعبة". While the second sentence was correct, it seemed more accurate to use "standing within, يقف ضمن" than "a dust storm appears, تظهر عاصفة ترابية".

It is important to mention that the qualitative examples presented in this study vary for each experiment. This can be assigned to the implementation of random data splitting, which resulted in the generation of different test samples through experiments. Therefore, this comparison illustrated the differences through models.

Table -3: Examples for the six experiments

	Input	**Prediction**	**Reference**
mT5 Baseline	صبي صغير ينزلق إلى أسفل شريحة ماء صفراء	<extra_id_0> صبي صغير	ينزلق صبي صغير إلى أسفل الحوض المائي الأصفر
	اثنين من السحالي تتقاتل في الماء	<extra_id_0> في التذكير	تتقاتل اثنتان من السحالي في الماء
	كلب في سترة حمراء يركض في الميدان	<extra_id_0> كلب في ...	يركض كلب في سترة حمراء في الميدان
mT5 Size_2001	رجل يرتدي سترة صفراء مع العديد من الأساور الملونة ويبتسم	يرتدي رجل سترة صفراء ويحمل العديد من الأساور الملونة ويبتسم	يرتدي رجل سترة صفراء وأساور ملونة ويبتسم
	اثنين من الجراء يمشيان على الحجارة	يمشي اثنان من الجراء على الحجارة	يمشي اثنان من الجراء على الحجارة
	رجل ينام مع حقيبة في حضنه على وسيلة النقل العام	ينام رجل مع حقيبة في حضنه على وسيلة النقل العام	ينام الرجل ويضع الحقيبة في حضنه في وسيلة النقل العام
mT5 Size_6103	كلب أبيض ورمادي يأكل أذن كلب أبيض وبني	يأكل كلب أبيض ورمادي أذن كلبا أحمر وبني	يقضم كلب أبيض ورمادي أذن كلب أبيض وبني
	كلبان يقفزان في حمام السباحة	يقفز كلبان في حمام السباحة	يقفز كلبان إلى حمام السباحة
	كلب أبيض وأسود القفز من خلال	يقفز كلب أبيض وأسود خلال حلقة	يقفز كلب أبيض وأسود عبر حلقة

	طارة سوداء وبيضاء	سوداء وبيضاء	سوداء وبيضاء
AraT5 Baseline	صبي مع بندقية لعبة	solar solar Malayطعنة Malay ⃠ جنهpierre	يحمل صبي بندقية لعبة
AraT5 Size_2001	اثنين من الكلاب الرمادية وكلب أسود يلعبان في بركة تحت جسر علوي	يلعب اثنان من الكلاب رمادية وكلب أسود في بركة تحت جسر	يلعب كلبان رماديان وكلب أسود في بركة تحت جسر
	رجل يركب دراجته الترابية أسفل درب صخري	يركب رجل دراجته الترابية أسفل درب صخري	يركب رجل دراجته أسفل درب صخري
	امرأة شابة ترشف القهوة وتقرأ كتابا	يرش امرأة شابة القهوة ويقرأ كتابا	ترشف امرأة شابة القهوة وتقرأ كتابا
AraT5 Size_6103	الكلاب تضرب في العشب على لعبة	تتقاتل الكلاب على العشب من أجل اللعبة	تتقاتل الكلاب على العشب من أجل اللعبة
	كلب وعاصفة ترابية	يظهر كلب وعواصف ترابية	يقف كلب ضمن عاصفة ترابية
	الكلب البني على وشك عض كرة مطبوعة	يعض الكلب البني كرة ملونة	يعض الكلب البني كرة بنقوش زرقاء وصفراء

6. CONCLUSIONS and Future Work

In this study, we presented an enhanced version of the Arabic Flickr8k dataset, along with a detailed analysis of its linguistic errors. The enhanced Arabic Flickr8k was utilized to fine-tune the mT5 and AraT5 models. The results showed that these models, particularly AraT5, produced incorrect and grammatically weak sentences without training. Fine-tuning the models on 2001 sentence pairs resulted in a significant improvement in sentence quality in the grammatical, orthographical, semantic, and morphological quality of the generated sentences, despite some minor issues. Therefore, we expanded the dataset using a data augmentation approach based on verbs, subjects, and adjective synonyms. This resulted in a dataset of 6103 Arabic sentence pairs, which helped to improve the correction of the input sentences, resulting in grammatically correct sentences with only minor differences from the reference sentences.

The results confirm that both fine-tuning and data quality enhance the ability of Arabic models to produce correct Arabic sentences.

In future work, we plan to expand the dataset by increasing the number of corrected sentences and enriching each sentence with its proper grammatical annotation (I'rab). It is also possible to integrate with other datasets to increase the diversity of Arabic sentences.

REFERENCES

[1] A. Farghaly and K. Shaalan, "Arabic Natural Language Processing: Challenges and Solutions," ACM Transactions on Asian Language Information Processing, vol. 8, no. 4, p. 14:1-14:22, Dec. 2009, doi: 10.1145/1644879.1644881.

[2] O. ElJundi, M. Dhaybi, K. Mokadam, H. Hajj, and D. Asmar, "Resources and End-to-End Neural Network Models for Arabic Image Captioning:" in Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Valletta, Malta: SCITEPRESS - Science and Technology Publications, 2020, pp. 233–241. doi: 10.5220/0008881202330241.

[3] H. A. Al-muzaini, T. N., and H. Benhidour, "Automatic Arabic Image Captioning using RNN-LSTM-Based Language Model and CNN," ijacsa, vol. 9, no. 6, 2018, doi: 10.14569/IJACSA.2018.090610.

[4] "ArEnMulti30K." Accessed: Apr. 06, 2026. [Online]. Available: https://sites.google.com/view/arenmulti30k

[5] H. Nassar, "Challenges of Post-Editing in English to Arabic Machine Translation of Technical Texts: A Study of Technological and Linguistic Barriers," International Journal of Linguistics, Literature and Translation, vol. 8, pp. 01–15, Mar. 2025, doi: 10.32996/ijllt.2025.8.4.1.

[6] T. Sarwar, A. J. J. Yepes, and L. Cavedon, "Assessing the Impact of the Quality of Textual Data on Feature Representation and Machine Learning Models: Quantitative Study Using Large Language Models," J Med Internet Res, vol. 27, p. e73325, Dec. 2025, doi: 10.2196/73325.

[7] E. M. B. Nagoudi, A. Elmadany, and M. Abdul-Mageed, "AraT5: Text-to-Text Transformers for Arabic Language Generation," in Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), S. Muresan, P. Nakov, and A. Villavicencio, Eds., Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 628–647. doi: 10.18653/v1/2022.acl-long.47

[8] L. Xue et al., "mT5: A Massively Multilingual Pre-trained Text-to-Text Transformer," in Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, and Y. Zhou, Eds., Online: Association for Computational Linguistics, Jun. 2021, pp. 483–498. doi: 10.18653/v1/2021.naacl-main.41.

[9] C. Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer".

[10] A. Bashiti et al., "ImageEval 2025: The First Arabic Image Captioning Shared Task," in Proceedings of The Third Arabic Natural Language Processing Conference: Shared Tasks, K. Darwish, A. Ali, I. Abu Farha, S. Touileb, I. Zitouni, A. Abdelali, S. Al-Ghamdi, S. Alkhereyf, W. Zaghouani, S. Khalifa, B. AlKhamissi, R. Almatham, I. Hamed, Z. Alyafeai, A. Alowisheq, G. Inoue, K. Mrini, and W. Alshammari, Eds., Suzhou, China: Association for Computational Linguistics, Nov. 2025, pp. 376–389. doi: 10.18653/v1/2025.arabicnlp-sharedtasks.52.

[11] M. Hodosh, P. Young, and J. Hockenmaier, "Framing Image Description as a Ranking Task Data, Models and Evaluation Metrics Extended Abstract".

[12] X. Chen et al., "Microsoft COCO Captions: Data Collection and Evaluation Server," Apr. 03, 2015, arXiv: arXiv:1504.00325. doi: 10.48550/arXiv.1504.00325.

[13] P. Young, A. Lai, M. Hodosh, and J. Hockenmaier, "From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions," Transactions of the Association for Computational Linguistics, vol. 2, pp. 67–78, 2014, doi: 10.1162/tacl_a_00166.

[14] H. Hejazi and K. Shaalan, "Deep Learning for Arabic Image Captioning: A Comparative Study of Main Factors and Preprocessing Recommendations," IJACSA, vol. 12, no. 11, 2021, doi: 10.14569/IJACSA.2021.0121105.

[15] K. F. Shaalan, "Arabic GramCheck: a grammar checker for Arabic," Softw: Pract. Exper., vol. 35, no. 7, pp. 643–665, Jun. 2005, doi: 10.1002/spe.653.

[16] A. Y. G. Alfaifi, "Building the Arabic Learner Corpus and a System for Arabic Error Annotation," phd thesis, University of Leeds, 2015. Accessed: Apr. 06, 2026. [Online]. Available: https://etheses.whiterose.ac.uk/id/eprint/9736/

[17] حنان جنان ع. (2008). الاخطاء اللغوية، الاملائية والنحوية والصرفية. المجلة الجزائرية التربية والصحة النفسية, 2(2), 150-171. https://asjp.cerist.dz/en/article/243922

[18] A. A. Freihat, H. M. Khalilia, G. Bella, and F. Giunchiglia, "Advancing the Arabic WordNet: Elevating Content Quality," in Proceedings of the 6th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT) with Shared Tasks on Arabic LLMs Hallucination and Dialect to MSA Machine Translation @ LREC-COLING 2024, H. Al-Khalifa, K. Darwish, H. Mubarak, M. Ali, and T. Elsayed, Eds., Torino, Italia: ELRA and ICCL, May 2024, pp. 74–83. Accessed: Apr. 06, 2026. [Online]. Available: https://aclanthology.org/2024.osact-1.9/

BIOGRAPHIES

Hadeel Abdulrahman
PhD Student, Department of Artificial Intelligence and Natural Languages Faculty of Informatics Engineering University of Aleppo

Prof. Mohammed Fadel Sukkar
Department of Artificial Intelligence and Natural Languages Faculty of Informatics Engineering University of Aleppo