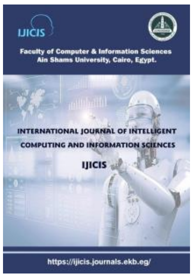
IJICIS, Vol.23, No.2, 105-115
DOI: 10.21608/ijicis.2023.210435.1270

International Journal of Intelligent
Computing and Information Sciences
https://ijicis.journals.ekb.eg/


IJICIS
Faculty of Computer & Information Sciences
Ain Shams University, Cairo, Egypt.
INTERNATIONAL JOURNAL OF INTELLIGENT COMPUTING AND INFORMATION SCIENCES
IJICIS
https://ijicis.journals.ekb.eg/

A CASE STUDY OF IMPROVING ENGLISH-ARABIC TRANSLATION USING THE TRANSFORMER MODEL

Donia Gamal*

Computer Science Department,
Faculty of Computer and
Information Sciences, Ain Shams
University,
Cairo, Egypt
donia.gamaleldin@cis.asu.edu.eg

Marco Alfonse

Computer Science
Department,
Faculty of Computer and
Information Sciences, Ain
Shams University,

aboratoie Interdisciplinaire de
l'Université Française
d'Égypte (UFEID LAB),
Université Française d'Égypte,
Cairo, Egypt
marco_alfonse@cis.asu.edu.eg

Salud María
Jiménez-Zafra

Computer Science
Department,
CEATIC,
Universidad de
Jaén, Jaén, Spain.
sjzafra@ujaen.es

Mostafa Aref

Computer Science
Department,
Faculty of Computer and
Information Sciences, Ain
Shams University
Cairo, Egypt
mostafa.aref@cis.asu.edu.eg

Received 2023-05-11; Revised 2023-05-11; Accepted 2023-06-14

Abstract: *Arabic is a language with rich morphology and few resources. Arabic is therefore recognized as one of the most challenging languages for machine translation. The study of translation into Arabic has received significantly less attention than that of European languages. Consequently, further research into Arabic machine translation quality needs more investigation. This paper proposes a translation model between Arabic and English based on Neural Machine Translation (NMT). The proposed model employs a transformer with multi-head attention. It combines a feed-forward network with a multi-head attention mechanism. The NMT proposed model has demonstrated its effectiveness in improving translation by achieving an impressive accuracy of 97.68%, a loss of 0.0778, and a near-perfect Bilingual Evaluation Understudy (BLEU) score of 99.95. Future work will focus on exploring more effective ways of addressing the evaluation and quality estimation of NMT for low-data resource languages, which are often challenging as a result of the scarcity of reference translations and human annotators.*

Keywords: *English-Arabic Translation, Neural Machine Translation, Attention Mechanism, Transformer with Multi-head Attention, Low Data Resource Languages*

1. Introduction

***Corresponding Author**: Donia Gamal

Computer Science Department, Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt

Email address: donia.gamaleldin@cis.asu.edu.eg

Translation plays a crucial role in facilitating communication and understanding between people from different linguistic and cultural backgrounds [1]. It enables the exchange of ideas, knowledge, and information across borders and helps to promote cross-cultural understanding and cooperation. In the globalized world, language barriers can make it difficult to access information. Because of the volume of information generated, it is sometimes impossible to meet the demand for translations by relying solely on professional human translators.

Machine Translation (MT) has seen significant progress in the diversity and variety of research fields and languages in the past few years[2]. The field of Natural Language Processing (NLP) has greatly benefited from advancements in MT. This has led to a growing need for fast and precise translations. MT is particularly useful in fields such as e-commerce, where it can help businesses reach new markets, and in social platforms, where it can facilitate cross-cultural communication.

Advances in MT technology have led to significant improvements in translation quality, making it an increasingly viable option for many applications [3]. Despite the existence of several MT systems, including those for the Arabic language, there is still room for improvement in the efficiency and productivity of translations produced by these systems: Sakhr[1], Bing[2], and Cambridge Dictionary[3], as well as free online MT systems like Google Translate[4].

Arabic is the primary language for 400 million people across 21 countries and is also spoken as a second language in some Islamic nations [4]. However, MT of Arabic presents unique challenges due to the language’s complex linguistic features, such as its lexicon, syntax, morphology, and textual differences from other languages. In comparison to English, Arabic has a more complex morphology. This makes it challenging to develop tools for stemming/lemmatizing and adapting tokenizers like Word-Piece and Sentence-Piece for Arabic. Additionally, context plays a crucial role in resolving lexical ambiguities common in Arabic.

MT has been categorized in several ways, with different researchers proposing various classifications. One common classification divides MT into single and hybrid approaches. The single approaches use only one method to translate between natural languages and include rule-based, direct, corpus-based, and knowledge-based methods [5]. The hybrid approaches, on the other hand, combine statistical and rule-based methods.

Deep learning, a sub-class of machine learning methods based on artificial neural networks [6], allows computational models with multiple processing layers to learn data representations at different levels of abstraction. These methods have significantly advanced language translation research. NMT is a recent approach that produces high-quality translations using enormous amounts of aligned parallel text corpora in both the source and target languages. NMT employs a large artificial neural network to predict the likelihood of a sequence of words, modeling entire sentences in a single integrated model. This approach eliminates the need and necessity for specialized systems in the statistical MT pipeline, as a single system can be trained directly on the source and target text. NMT technology is already being used by major companies.

[1] http://www.sakhr.com/
[2] http://www.bing.com/
[3] https://dictionary.cambridge.org/
[4] https://translate.google.com/

NMT has shown promising results and is now widely used for a variety of language pairs. While European languages have received much attention and have abundant resources, there is limited work on NLP for various Arabic languages. Thus, the Arabic NMT faces different and several challenges. Arabic is a morphologically rich language with complex grammar and sentence structures. It also has a wide range of dialects and variations in vocabulary and pronunciation. These factors can make it difficult for MT systems to accurately capture the meaning and context of the text. The limited availability of high-quality training data can impact the efficiency and performance of NMT systems for Arabic. The development of English-Arabic MT systems is still in its early stages due to a shortage of parallel corpora [7]. To address this challenge, this paper suggests a transformer-based approach for English-Arabic NMT system. The contents of this paper are organized into the following sections: Section 2 describes previous work on parallel Arabic corpora. Section 3 presents the used dataset, experimental setup, and tools used to train and analyze the proposed English-Arabic NMT model. The analysis continues in Section 4 with a discussion of the potential implications of the experimental results. Finally, Section 5 summarizes the research results and suggests future directions for the research community.

2. Related Work

This section provides a comprehensive overview of the NMT techniques between various languages and the Arabic Language.

Kumar, et al. [8] have translated Arabic sentences into Bangla using a NMT system with a fixed vocabulary and a Long short-term memory (LSTM) encoder decoder mechanism. This approach automatically transfers the source sequences to the target sequences on a collected dataset from ManyThings.org and it contained 90,000 record [9]. Their experiments have shown a BLEU score of 45.

Oudah, et al. [10] investigated a methodology for Arabic-English translation using an encoder-decoder and general global attention architecture. LSTM with hidden units were applied to both the encoder and decoder on a dataset (LDC2010T12 (MT04)) of 1,075 sentences. The proposed model achieved a BLEU Score of 53.54.

Al-Ibrahim and Duwairi [11] proposed a Deep Learning framework for translating Jordanian dialect into Modern Standard Arabic (MSA) using an RNN encoder-decoder model. The model performed well on their manually created Sentence2Sentence (S2S) and Word2Word (W2W) datasets. The W2W dataset, used for word-for-word translation, contains 24,200 words. The S2S dataset, used for sentence-level translation, contains 500 sentences. The experiments achieved BLEU Scores of 63.20 for the S2S dataset and 91.30 for the W2W dataset.

Kchaou, et al. [12] investigated the efficiency of Convolution Neural Network (CNN), Recurrent Neural Network (RNN), and transformer models in translating the Tunisian dialect into MSA using a dataset of 34K sentences[13]. The transformer model achieved the best translation with a BLEU Score of 60, compared to 33.36 and 53.98 for the RNN and CNN MT models, respectively.

Bensalah, et al. [14] proposed a novel Deep Learning architecture based on CNN and the transformer model to improve Arabic-English NMT results. They used a distinct preprocessing of Arabic sentences using Farasa and AraBERT on the UN dataset containing over 73,000 Arabic-English sentences. Their approach achieved a BLEU Score of 68.60.

Hamed, et al. [15] presented a parallel corpora of 6,237 Italian-Arabic statements and employed two deep learning models for Arabic-to-Italian MT, LSTM, and GRU sequence-to-sequence with attention mechanism. The LSTM-based model achieved a higher mean BLEU Score of 91 compared to the GRU-based model's score of 89.

Table 1 provides a synopsis of the related work in NMT from/to the Arabic language.

Table 1 Overview of NMT Models on Different Arabic Datasets

Authors	Translation Model	Dataset Size	Source → Target Language	BLEU Score
Kumar, et al. [9]	LSTM Encoder-Decoder Model	90,000	Arabic → Bangla	45
Oudah, et al. [10]	LSTM Encoder-Decoder Model	1,075	Arabic → English	53.54
Al-Ibrahim and Duwairi [11]	RNN Encoder-Decoder Model	24,200	Jordanian Dialect → MSA	91.30
		500		63.20
Kchaou, et al. [12]	RNN Encoder-Decoder Model	34,000	Tunisian Dialect → MSA	60
	CNN Encoder-Decoder Model			33.36
	Transformer Model			53.98
Bensalah, et al. [14]	CNN Encoder-Decoder Model	73,000	English → Arabic	68.60
Hamed, et al. [15]	LSTM Encoder-Decoder Model	6,237	Arabic → Italian	91
	GRU Encoder-Decoder Model			89

3. The Proposed Methodology

This section presents the proposed English-Arabic NMT. Figure 1 illustrates the steps of the proposed NMT framework for generating Arabic text from parallel English sentences. The process begins with loading and splitting the data into pairs, followed by a discussion of the preprocessing steps applied to the dataset. The specifics of the NMT model are then presented.

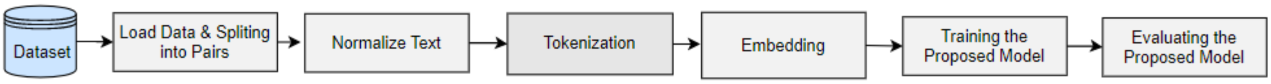

Dataset
Load Data & Spliting into Pairs
Normalize Text
Tokenization
Embedding
Training the Proposed Model
Evaluating the Proposed Model

Figure 1: The Proposed English - Arabic NMT

3.1. Dataset

The used dataset is a set of Tatoeba Project [16] sentences classified by frequency words (Arabic (front) and English (back). Sentences with high-frequency words will appear at the top, while sentences with low-frequency words will appear at the bottom.
Table 2 shows more details about the used dataset.

Table 2 Details of Tatoeba Dataset

	Target Language (Arabic)	**Source Language (English)**
Number of samples	11,433	
Size of Training Dataset	9,147	
Size of Testing	2,286	
Number of unique tokens	45	50
Max sequence length	201	209
Sample	انا اقدر ذلك	I appreciate this

Figure 2 shows the variation length of sentences in Arabic and English languages.

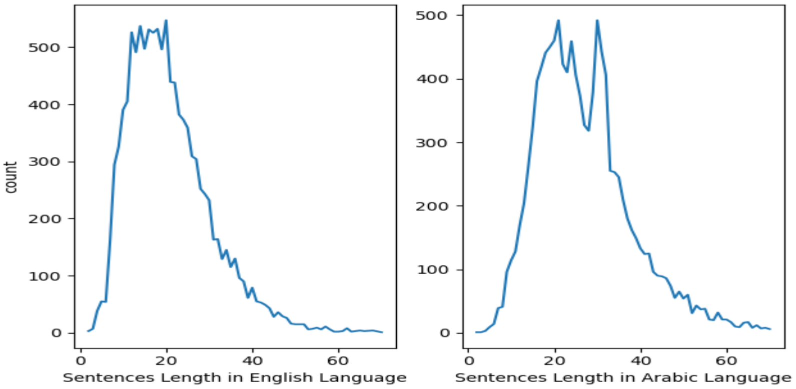


Figure 2: Sentence Length in Arabic and English Languages

Figure 3 shows the frequency of sentence length for Arabic and English languages.

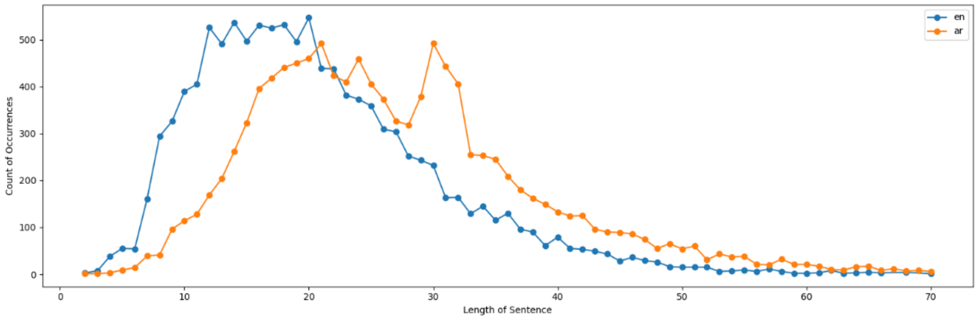


Figure 3: Frequency of Sentence Length in Arabic and English Languages

3.2. Loading Data and Splitting into Pairs

The dataset is divided into tab-delimited pairs, each of which contains an English text sequence and its equivalent Arabic text sequence. Each sentence is split into pairs as source and target languages for translation. It is important to note that a text sequence can consist of either a single sentence or a paragraph containing multiple sentences. In this proposed MT model for translating English into Arabic, in this context, English is regarded as the source language while Arabic is the target language.

3.3. Normalization

This phase performs simple normalization on texts by removing white spaces, repeating characters, and diacritics. It considers normalizing elements such as Hamza, and digits, and tokenizing punctuation and digits. It enforces consistency in various aspects of the text, including diacritics, letters, and tokenization.

3.4. Tokenization

Tokenization involves breaking down the text into smaller units called tokens, which can be words, sentences, phrases, symbols, or any other meaningful piece of text. This is a crucial step in NLP as it enables a computer to analyze and manipulate text at different levels of detail.

3.5. Embedding

A method to convert sentences into a data format that can be fed into a machine-learning model is needed. In essence, the textual input data (tokens) has to be converted into a numerical format. Using an embedding layer (Keras Embedding Layer) each input is converted into a vector of values. These vectors will be fed to positional encoding to keep the information of the position of each token and order as they are highly essential parts of any language.

3.6. Training the Proposed Model

When working with a dataset, a machine learning algorithm typically goes through two stages: training and testing. The data is usually split into training and testing subsets in an 80%-20% ratio. The transformer model, which is a type of neural network architecture that uses self-attention mechanisms to encode and decode sequences like natural language sentences for sequence-to-sequence tasks such as MT, was trained for 10 epochs. The self-attention mechanism allows the model to learn the relationships between input and output tokens without using recurrent or convolutional layers. The multi-head attention mechanism is a type of attention that enables the model to attend to information from different representations at different positions simultaneously, making it parallelizable and more efficient on hardware like GPUs and TPUs.

Transformer models, as opposed to standard recurrent or convolutional models, rely on self-attention techniques to capture the dependencies and interactions between the input tokens. The self-attention is a method of constructing a representation of each token by taking the context of the other tokens in the sequence into account. Transformer models; as represented by their creators in 2017 [17]; are made up of two major parts: an encoder and a decoder as shown in Figure 4.

The Transformer architecture utilizes an encoder-decoder structure. As depicted on the left side of Figure 4, the encoder converts an input sequence into a sequence of continuous representations. On the right side of Figure 4, the output from the encoder is combined with the output from the previous step of the decoder to produce an output sequence.

3.6.1. Transformer Encoder

The transformer encoder generates a sequence of hidden states from the input sequence, which is subsequently passed to the decoder[18]. The output sequence is generated by the decoder by paying attention to both the encoder states and its prior outputs. The transformer encoder is made up of N identical layers, in this experiment, 3 encoder layers were used. Each encoder layer is made up of two primary sub-layers as follows: The first sub-layer is a multi-head attention mechanism that accepts queries, keys, and values as input. The second sublayer is a fully connected feed-forward network. The multi-head attention layer applies a self-attention mechanism that allows the model to associate each token in the input to the other tokens in the input, for example, for the sentence “How are you”, it tries to associate the token “you” to “how” and “are” tokens. The output will be fed into a fully connected feed-forward network to create the query, key, and value vectors. The concept of key/value/query is similar to retrieval systems. For instance, when you search for videos on YouTube, the search engine compares your query (the text you enter in the search bar) to a set of keys (such as video title and description) associated with potential videos in its database. It then presents you with the most relevant videos (values). The attention operation can also be viewed as a retrieval procedure. In the translation model, the Query (Q) vector is from the target language sequence, and Key (K) and Value (V) vectors are from the source language sequence.

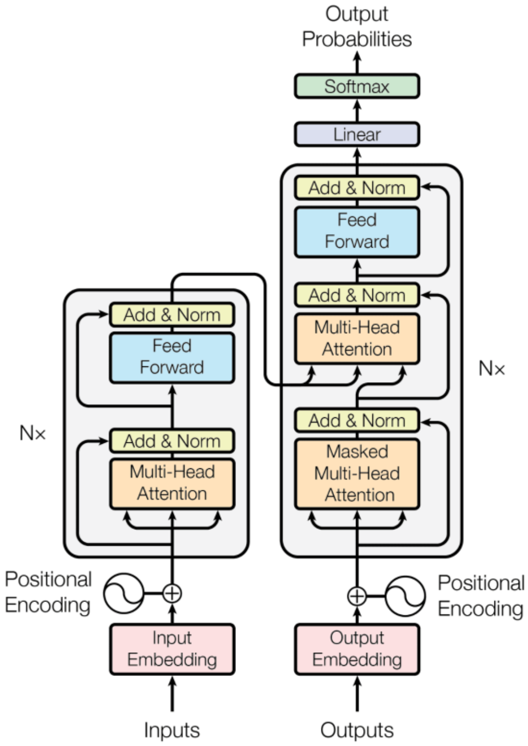


Figure 4: The Transformer architecture, as described by [17], features an encoder-decoder structure where N denotes the number of layers

3.6.2. Transformer Decoder

The transformer decoder is a component of a neural network that can generate sequences based on an input sequence and an attention-based representation of it. The transformer decoder is made up of several layers, in this experiment, 2 decoder layers were used. Each of these contains three sub-layers: a multi-head attention mechanism, a second multi-head attention mechanism, and a feed-forward network. The first multi-head attention mechanism in the decoder uses the previous decoder layer's output (or the target sequence embeddings for the first layer) as queries, keys, and values to seize the dependencies within the target sequence. The second multi-head attention mechanism in the decoder uses the first multi-head attention mechanism's output as queries and the encoder's output as keys and values to capture the dependencies between the target and input sequences. The decoder's feed-forward network is a two-layer perceptron that performs a non-linear transformation of the output of the second multi-head attention mechanism. A Layer normalization is used to normalize the output of each sub-layer in the decoder, and residual connections are added to preserve information from previous layers.

3.6.3. Multi-head attention

A multi-head attention mechanism computes attention scores between a set of queries, keys, and values by employing multiple parallel attention heads that learn different aspects of the relationships. The input token can relate (through its query vector) to each other token (depending on their key vectors matching), by using the entire collection of query-key compatibility scores. This allows the computing of a representation for each token by calculating the properly weighted sum over the other tokens.

3.7. Evaluating the Proposed Model

Two metrics for evaluating a transformer model for translation were used which are accuracy and BLEU Score. Training accuracy is how well the model predicts the proper translation on training data, and the BLEU Score is how similar the model's translation is to a human reference translation on test data. A high training accuracy does not always translate into a high BLEU Score, because the model may overfit the training data and fail to generalize to new phrases. During the training stage, the model was evaluated by loss and accuracy. For the testing stage, the model was evaluated by BLEU Score. The BLEU Score is a metric that compares a system's output against one or more reference translations to determine the quality of MT. The BLEU Score goes from 0 to 1, with 0 indicating no overlap and 1 indicating a perfect match. The BLEU Score is derived by calculating the number of n-grams (n-word sequences) shared by the system output and the reference translations and then dividing by the total number of n-grams in the system output.

4. Learning Environment and Experimental Results

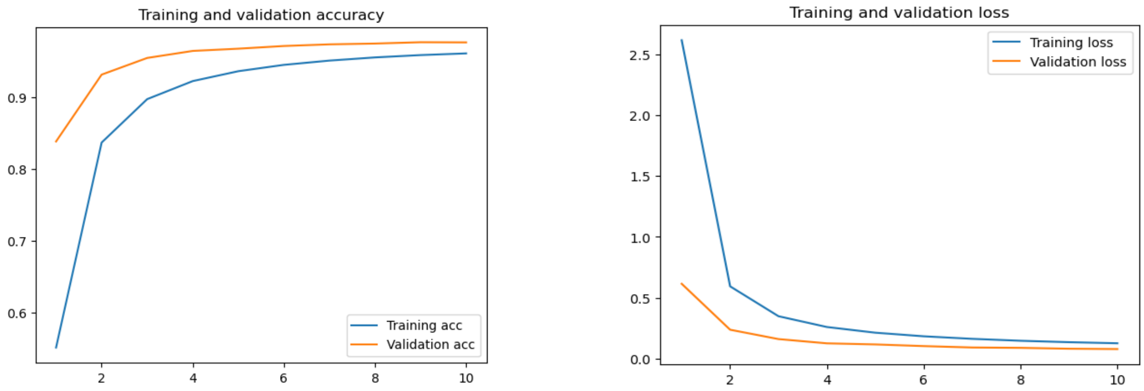
This paper offers the transformer network with multi-head attention to translate English to Arabic language. This experiment is applied using Keras Transformer [19] on Kaggle [20] environment with GPU T4 x 2 for 10 Epochs. Table 3 presents the accuracy and loss of applying the proposed model to the Tatoeba Dataset [16].

Table 3 The Experimental Results of Transformer with Multi-head Attention Translation Model

Epoch	Train loss	Validation loss	Train Accuracy	Validation Accuracy
1	2.6142	0.6145	55.16%	83.85%
2	0.5936	0.2375	83.70%	93.15%
3	0.3482	0.1606	89.74%	95.47%
4	0.2598	0.1258	92.25%	96.45%
5	0.2131	0.1168	93.63%	96.76%
6	0.1835	0.1030	94.51%	97.14%

7	0.1627	0.0916	95.10%	97.37%
8	0.1469	0.0887	95.55%	97.48%
9	0.1352	0.0816	95.87%	97.66%
10	0.1264	0.0788	96.11%	97.68%

The proposed approach for NMT has been shown to produce high-quality translations between English and Arabic languages with an impressive accuracy of 97.68% and a low loss of 0.0788 in Table 3. The training and validation progress of the transformer model can be seen in Figure 5.



(a) (b)

Figure 5: The Training and Validation of the Proposed Transformer: (a) Accuracy and (b) Loss

As demonstrated in Figure 5, the accuracy and loss curves of the proposed transformer model show how well the model performs on the training and validation data over 10 epochs. Table 4 shows a sample input and output of testing the proposed transformer model that achieves a BLEU Score of **99.90**.

Table 4 A Sample of Testing English to Arabic Translation Using Transformer Model

Input	Target	Predicted
what do you want now	ما الذي تريده الان	ماذا تريد الان
tom tried to kill me	حاول توم ان يقتلني	حاول توم قتلي
i don t like coffee	لا احب القهوه	لا احب القهوه
I don t believe it	لا يمكنني تصديق ذلك	لا يمكنني ان اصدق ذلك
I have one brother	لدي اخ واحد	لي اخ وحيد

As noted in Table 4, translation is not a word-for-word or sentence-for-sentence mapping. A translation may have more than one structure or answer depending on the situation, goal, and audience. As a result, the translation model has to be aware of the various aspects that affect the quality and appropriateness of the predicted output. A multi-head attention transformer model can consider the context of a translated paragraph. The model's multi-head attention mechanism allows it to be attentive to multiple parts of the input sequence at the same time, allowing it to capture contextual relationships between words in a phrase or paragraph.

5. Conclusion and Future Work

MT brings people from all over the world together to collaborate, share information, and form bonds. NMT has outperformed both rule-based and statistical MT techniques. While numerous ML systems support Arabic, the quality of translation still needs to be enhanced. This paper introduces a NMT model for English-Arabic translation that employs a transformer model with a multi-head attention mechanism. The experimental results demonstrate that the proposed model improves translation by gaining an accuracy of 97.68%, a loss value of 0.0788, and a BLEU Score of 99.95. NMT has achieved remarkable results, However, NMT still faces some challenges, such as handling rare words, domain adaptation, and model robustness. Also, the evaluation of NMT for low data/resource languages, which are frequently difficult to address due to a lack of reference translations and human annotators.

References

[1] Sangmin Michelle Lee, “The effectiveness of machine translation in foreign language education: a systematic review and meta-analysis,” Journal of *Computer Assisted Language Learning*, vol. 36, no. 1–2, pp. 103–125, 2021, doi: 10.1080/09588221.2021.1901745.

[2] Joss Moorkens, “Ethics and machine translation,” Journal of *Machine translation for everyone: Empowering users in the age of artificialintelligence*, vol. 18, pp. 121–140, 2022.

[3] Miguel Domingo, Mercedes García-Martínez, Alexandre Helle, Francisco Casacuberta, and Manuel Herranz, “How Much Does Tokenization Affect Neural Machine Translation?,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2023, vol. 13451 LNCS, pp. 545–554. doi: 10.1007/978-3-031-24337-0_38.

[4] “https://www.nationsonline.org/oneworld/countries_by_languages.htm.” [accessed 9 April 2023].

[5] Jaideepsinh K. and Jatinderkumar R., “Sanskrit Machine Translation Systems: A Comparative Analysis,” Journal of *International Journal of Computer Applications*, vol. 136, no. 1, pp. 1–4, 2016, doi: 10.5120/ijca2016908290.

[6] Rym Nihel Sekkal, Fethi Bereksi-Reguig, Daniel Ruiz-Fernandez, Nabil Dib, and Samira Sekkal, “Automatic sleep stage classification: From classical machine learning methods to deep learning,” Journal of *Biomedical Signal Processing and Control*, vol. 77, p. 103751, 2022, doi: 10.1016/j.bspc.2022.103751.

[7] Sanaa Kaddoura, Rowanda D. Ahmed, and D. Jude Hemanth, “A comprehensive review on Arabic word sense disambiguation for natural language processing applications,” Journal of *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 4, p. e1447, 2022, doi: 10.1002/widm.1447.

[8] Toshiba Kamruzzaman, “Arabic To Bangla Machine Translation Using Encoder Decoder Approach,” in *2020 IEEE Region 10 Symposium (TENSYMP)*, 2020, pp. 1176–1179.

[9] Rashi Kumar, Piyush Jha, and Vineet Sahula, “An augmented translation technique for low resource language pair: Sanskrit to hindi translation,” in *Proceedings of the 2nd international conference on algorithms, computing and artificial intelligence*, 2019, pp. 377–383.

[10] Mai Oudah, Amjad Almahairi, and Nizar Habash, “The Impact of Preprocessing on Arabic-English Statistical and Neural Machine Translation,” in *Machine Translation Summit XVII Volume 1: Research Track*, 2019, pp. 214–221. [Online]. Available: http://arxiv.org/abs/1906.11751

[11] Roqayah Al-Ibrahim and Rehab M Duwairi, “Neural machine translation from Jordanian Dialect to modern standard Arabic,” in *the 11th International Conference on Information and Communication Systems (ICICS)*, 2020, pp. 173–178.

[12] Saméh Kchaou, Rahma Boujelbane, and Lamia Hadrich Belguith, “Hybrid pipeline for building

Arabic Tunisian Dialect-Standard Arabic Neural machine translation model from scratch," Journal of *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2022, doi: 10.1145/3568674.

[13] Saméh Kchaou, Rahma Boujelbane, and Lamia Hadrich-Belguith, "Parallel resources for Tunisian Arabic Dialect Translation," in *the 5th Arabic Natural Language Processing Workshop*, 2020, pp. 200–206. [Online]. Available: https://aclanthology.org/2020.wanlp-1.18

[14] Nouhaila Bensalah, Habib Ayad, Abdellah Adib, and Abdelhamid Ibn El Farouk, "Transformer Model and Convolutional Neural Networks (CNNs) for Arabic to English Machine Translation," in *Lecture Notes in Networks and Systems*, 2022, vol. 489 LNNS, pp. 399–410. doi: 10.1007/978-3-031-07969-6_30.

[15] Haiam Hamed, Abdel Moneim Helmy, and Ammar Mohammed, "Holy Quran-Italian seq2seq Machine Translation with Attention Mechanism," in *the 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 2022, pp. 11–20. doi: 10.1109/MIUCC55081.2022.9781781.

[16] "https://www.manythings.org/bilingual/ara." [accessed 15 April 2023].

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," Journal of *Advances in Neural Information Processing Systems*, vol. 2017-Decem, pp. 5999–6009, 2017.

[18] Karl Hall, Victor Chang, and Chrisina Jayne, "A review on Natural Language Processing Models for COVID-19 research," Journal of *Healthcare Analytics*, vol. 2, p. 100078, 2022, doi: 10.1016/j.health.2022.100078.

[19] "https://keras.io/api/layers/attention_layers/multi_head_attention/." [accessed 21 April 2023].

[20] "https://www.kaggle.com/." [accessed 21 April 2023].