European Journal of Philosophical Research. 2026. 13(1)
E-ISSN: 2413-7286
Volume-13/Issue-1/2026

Deep Learning Algorithms for Arabic Machine Translation

Yaakoub DALI
Hassiba Ben Bouali University, Chlef (Algeria)
Y.dali90@univ-chlef.dz
Safa MEDJAHED
University of Tahri Mohamed, Bechar (Algeria)
medjahed.safa@univ-bechar.dz
Mohamed GUEBLI
University of Ahmed Zabana, Relizane (Algeria)
Mohamedguebli79@gmail.com
Hayet KASSI
Jilali Bounaama University, Khemis Miliana (Algeria)
hayetkassi02@gmail.com
Siham MOUSSAOUI
Hassiba Ben Bouali University, Chlef (Algeria)
ltala.lab@hotmail.fr

Soumission : 08/12/2025 Acceptation :01/06/2026 Publication : 16/06/2026

Abstract:
Arabic machine translation has advanced significantly with deep learning, despite the language's complexity. Arabic's rich morphology and variation between Modern Standard Arabic and dialects present major challenges. This paper reviews the evolution of machine translation from rule-based and statistical methods to neural models, focusing on Seq2Seq, attention mechanisms, and Transformers. It highlights the importance of multilingual and joint training to address data scarcity, especially for dialects. Key preprocessing steps such as normalization, tokenization, and morphological analysis are discussed for improving quality. The study also examines training strategies like pretraining and transfer learning, along with evaluation metrics such as BLEU and TER. It identifies challenges including limited resources and ambiguity, and explores applications, ethical issues, and future research directions.
Key words: Arabic Machine Translation; Deep Learning; Arabic Dialects; Transformer Models

Introduction

The Arabic language is spoken by more than 400 million people across a range of dialects and Modern Standard Arabic (MSA), the official language in 29 countries.

Arabic text is written from right to left using the Arabic script, which lacks capital letters and uses a diacritic system to indicate short vowels and other phonetic properties.

Machine Translation (MT) aims to automatically translate a text from one natural language to another. Its origin dates back to the late fifties, motivated by the rapid developments in both automatic digital computing and linguistics in western countries (Almahairi et al., 2016). The first attempt at MT was done using a Rule-Based Machine Translation (RBMT) in 1954 Harris et al. for Russian to English. For all linguistic components translators had to write large amount of rules covering almost all the grammatical styles of Russian and English (Almansor & Al-Ani, 2017).

1. Background and Context

The Arabic language is one of the oldest and most widely spoken languages globally. It has a high linguistic complexity due to its rich morphology, dialectal variation, and different writing forms. The most widely used variety of Arabic is Modern Standard Arabic (MSA), which is used in formal situations, media, literature, and education and is understood by speakers of all dialects. An estimated 511 million people use Arabic as their first and second language. Dialectal Arabic (DA), the spoken form of Arabic, is used in informal situations (Almansor & Al-Ani, 2017). Worldwide, Arabic is the fifth-most spoken language as a native language and the world's sixth-most spoken language overall.

Machine Translation (MT) refers to the use of a computer or software to automatically translate text or speech from one language to another, without the human intervention and received substantial attention in the last decades (Almahairi et al., 2016). MT aims to remove the language barrier between people and improve communication rapidly. Many of the early MT models were defined in the 1950s and were based on rule-based methods. Statistical-based MT models such as IBM models and phrase-based models emerged in the early 1990s and gained more popularity. In the early 2010s, deep-learning neural networks began to attract attention in many machine learning applications and showed better performance than traditional models. Since then, a few neural machine translation models based on deep-learning technologies also have been proposed for Arabic MT and gained substantial results.

1.1. Characteristics of the Arabic Language

Arabic is a Semitic language with a long history of written documentation spanning almost two millennia. Written Arabic comes in two standard varieties characterized according to its register. Modern Standard Arabic, used today by the media, government and academic institutions, is closely based on the Classical Arabic of the 7th century Qur'an and has changed little since that time. Contemporary, everyday Arabic is more diverse and is represented by many spoken dialects. These, although originating from a common source, are sufficiently different to be frequently mutually unintelligible across half the Arabic-speaking world – although they are all spoken in the same region, such as

Morocco to Qatar. The variety represented in news broadcast media or other consistent forms within a specific region is often referred to as that region's "standard Arabic." Arabic is written from right to left, has no capital letters, and the alphabet consists of twenty-eight letters without explicit marking of long vowels; in addition a series of optional diacritics are available to mark short vowels, gemination (doubling) or any of a few other features of pronunciation.

A considerable amount of Arabic language parallel texts are freely available for educational and interest purposes on the Internet but they tend to be either old and in Classical Arabic or very contemporary and in one or another of the regional dialects. The deep learning (DL) technologies observed widely in most languages remain in their infancy for dialectal Arabic and the topic has not been widely addressed by the MT community in recent years (Almansor & Al-Ani, 2017).

1.2. Evolution of Machine Translation

Automatic communication through language is of paramount importance for humans and machines alike, and to this end machine translation (MT) has been a focal point for several decades. MT has evolved through three phases since the 1950s: rule-based MT, statistical MT, and more recently neural MT. Explicit encoding of grammatical rules constituted the early vision of MT, which drew heavily on the work done in logic. Statistical methods established themselves rapidly following the advent of the Internet, as vast monolingual and parallel corpora became available; significant results were obtained through bilingual and multilingual phrase-based MT systems based on the IBM Models and N-gram models. Once again, the emergence of large datasets brought about a shift towards neural formalisms (Almahairi et al., 2016).

2. Core Deep Learning Architectures for Arabic MT

The Arabic language exhibits characteristics that pose challenges for MT (Almahairi et al., 2016). The short video presents a solution based on end-to-end multilingual NMT. The multilingual architecture of the model Kolcov et al. (2022) allows a single model to learn multiple languages simultaneously. For the Arabic-English direction, the model achieves competitive scores with the state-of-the-art supervised approach using competitive monolingual resources. Copying source tokens when translating from Arabic to English or switching source languages induces minimal interference from other languages. The sequence-to-sequence MT model known as Long Short-Term Memory shows moderate effectiveness. The Arabic-English translation task is also approached with copy mechanisms and joint training except in the multilingual setup. The Arabic preposition task confirms the expected performance drop when using the pre-trained multilingual model with the Arabic cue word.

The architecture of the leading Arabic-English NMT system features additional components that enhance its accuracy. The Iranian model has approximately 350 million parameters and utilizes a 4-layer, 4-head architecture. The English side adopts a shared

BERT vocabulary and a subword tokenization method that includes independent pre-processing BPE encoding. Arabic text normalization addresses several widely cited orthographic issues that hinder deep learning-based language processing. Three datasets are created to evaluate MT performance: an English corpus of mixed domains, a general-purpose news corpus for English-to-Arabic translation, and a manually validated speech transcription dataset. A shared vocabulary generated over both text sources successfully reduces Arabic out-of-vocabulary tokens.

2.1. Sequence-to-Sequence Models

Recently proposed sequence-to-sequence models provide novel approaches to machine translation and other text applications. They represent input sequences in a continuous format amenable to transformation and can generate sequences of arbitrary length. The approach directly models the mapping between source and target sentences as a single optimization problem, thus avoiding the multiple, matching-coordinate-mapping problems of older models.

The architecture consists of two recurrent neural networks (RNNs): an encoder, which computes a representation of the source sentence, and a decoder, which generates a target sentence from that representation. The two networks are typically trained together to maximize the likelihood of the training target sentences. In Arabic machine translation, the first application of the replenished approach of attention-based sequence-to-sequence models was to standard Arabic–dialectal Arabic translation (Almansor & Al-Ani, 2017). The dataset employed was extracted from penny-press articles, which appear in merged form on news websites and often contain dialectal expressions. The preprocessed parallel data contained 84 632 sentence pairs in total. As evident from this application, the approach accommodates the use of various architectures in the encoder and the decoder. Attention-based sequence-to-sequence models have also been successfully applied to Arabic medical text, improving performance when transferring a system trained on standard Arabic.

2.2. Transformer Models

The introduction of the Transformer architecture in 2017 has greatly impacted the field of MT and NLP in general (Wang et al., 2019). Unlike previous architectures that rely on RNNs or CNNs, Transformers use self-attention as the core mechanism for building the model. When introduced as an encoder-decoder model (the paradigm used in this section), it does not require reordering of the input. The authors report state-of-the-art performance on both English-Arabic and Arabic-English translations using the WIT3 dataset for evaluation.

2.3. Multilingual and Jointly Trained Architectures

The introduction of neural machine translation (NMT) has led to significant advances in translating distinct languages separately, even with limited parallel data. Nevertheless, challenges remain in processing low-resource dialectal Arabic (DA).

Jointly trained MT systems for high- and low-resource languages, such as Modern Standard Arabic (MSA) and Dialectal Arabic (DA), aim to facilitate adaptation across different modalities and integrate wide-domain linguistic features with low-resource data (Almansor & Al-Ani, 2017).

Multilingual and joint training provides a solution for low-resource languages and allows the incorporation of language-agnostic features into NMT (Moatez Billah Nagoudi et al., 2021). Combining language pairs for auxiliary training can improve the translation quality of a target language, given that multilingual models share a single set of parameters across all languages (Almahairi et al., 2016). A multilingual Arabic-English NMT architecture was developed for the translation of Arabic-English code-mixing across two distinct varieties: Modern Standard Arabic (MSA) and Egyptian Arabic (EA).

3. Data Resources and Preprocessing

Arabic has different varieties (or dialects)—such as Maghrebi, Egyptian, Gulf, Levantine, and Modern Standard Arabic (MSA)—that exhibit large intralanguage divergence in script, phonetics, syntax, morphology, semantics, and style, complicating machine translation (MT) (Almansor & Al-Ani, 2017). Transfer of semantic knowledge between close dialects can facilitate translation, and pre-training models on wiretap transcription vastly and rapidly boosts Egyptian Ar→MSA translation. Two-step translation systems are viable approaches for dialectal Arabic in MT (Albalawi et al., 2021). Characteristics of the Arabic language, evolution of MT, core deep-learning architectures for Arabic MT, data resources and preprocessing, training techniques and optimization, challenges and mitigation strategies, evaluation and benchmarking studies, applications and use cases, ethical-legal-social implications, future directions, and conclusion highlight the particularities, difficulties, and perspectives of Arabic Machine Translation and the advances achieved to date.

3.1. Corpus Types and Sources

Numerous Arabic corpora for various research domains are actively generated, and extensive linguistic data sources have been developed. These corpora encompass diverse topics such as morphology, discourse processing, and language variability issues like code-switching and dialectal mixing (Alkahtani, 2015). Arabic parallel corpora for machine translation cover topics including media, finance, law, Islamic literature, patents, and localization.

Corpus preparation for Arabic translation is facilitated through open-source data processing tools like OpenSubtitles. Alhaija et al. constructed a parallel corpus from English-to-Arabic Wikipedia articles, comprising 275,542 sentence pairs and versioned on GitHub. Almahairi et al. established a CADIM-compliant corpus of Arabic dialectal data from YouTube to foster the study of dialectal translation, featuring 10,000 sentence pairs in the dialectal-to-standard Arabic direction. Following guidelines from the African Language Technology (ALT) conference, they created an Arabic-English corpus based on TEDx video transcripts, containing aligned subtitles (en→ar and fr→ar) for 325 TEDx talks, to support the exploration of African languages in machine translation. Furthermore, Alshahrani et al. compiled a domain-specific corpus for law documents operating between Arabic and English, also providing the preliminary technical report on the Arabic-to-English translation published at the 1st Arabic and English Machine Translation Shared Task.

3.2. Tokenization and Normalization for Arabic

Arabic text exhibits complex linguistic phenomena that pose challenges to tokenization. A typical Arabic natural language processing pipeline involves steps such as the cleaning and removing of diacritics, tokenization, and normalization (Zarnoufi et al., 2022). Considerable ambiguity surrounds Arabic tokenization because Arabic consists of a base stem and many affixes, yet not all affixes require whitespace separation under the Arabic language rules. Moreover, some writings tend to preserve distinct semantic segments; however, adding whitespace in this case does not comply with Arabic tokenization rules. The lack of standardized definition of tokenization further exacerbates the problem. Implementing normalisation before tokenisation frequently helps automatic tokenization, minimizing the likelihood of erroneous whitespace insertion.

3.3. Handling Morphology and Orthography

Handling morphology and orthography involves addressing variations in word forms and spelling conventions in Arabic. Techniques such as word embeddings can capture semantic features between dialects and standard Arabic without rules. These methods have been effective even on small datasets and are expected to perform well on larger ones (Almansor & Al-Ani, 2017).

Morphologically rich languages tend to have more fully inflected words, with many morphemes representing various features, increasing sparsity and ambiguity. Semitic languages like Arabic and Hebrew are highly ambiguous, especially due to the presence or absence of short vowels. Arabic has about 12 analyses per word on average. Morphological dictionaries and analyzers encode all potential word inflections and analyses, which can be disambiguated through predicting morphological features and ranking analyses accordingly. Dialectal Arabic (DA) lacks a standard orthography, increasing ambiguity and sparseness, which is addressed by normalization using the CODA convention. Extensive studies on Arabic morphological tagging and disambiguation have been conducted, with recent advances utilizing deep learning models, often modeling features separately or focusing on limited subsets. Diacritization and lemmatization are important tasks with many recent deep learning approaches, such as LSTMs and sequence-to-sequence models. Early lemmatization methods used finite state machines, while recent models employed sequence-to-sequence architectures with attention. Joint modeling of morphological features has been explored with models like MADAMIRA, using SVMs and neural networks, often relying on external analyzers, with surface form normalization serving as a byproduct rather than an explicit target (Zalmout & Habash, 2019).

4. Training Techniques and Optimization

Deep learning frameworks have introduced new training and optimization techniques which improve performance of Arabic translation systems. While the target languages may vary in each case, multilingual models can be built on the many parallel resources that exist for Arabic. In addition, adapting large pretrained language models through fine-tuning or few-shot prompting has become prevalent, and research on Arabic illustrates this growth. The most resource-efficient and effective approach often consists of parameter-efficient transfer learning methods, where an Arabic checkpoint is initialized with either an English or multilingual set of weights.

Research has pushed forward the exploration of deep learning approaches to machine translation for Arabic language pairs, enhancing the creative process of authors and industries in educational, social, political, commercial, and governmental domains. Several core developments remain important for the future evolution of such systems. Pretraining large multilingual checkpoints on web-scale corpora containing Arabic text, in addition to text from other languages, represents the first major contribution and indicates the growing importance of transfer learning in the field. Various architectures remain competitive with state-of-the-art methods, and standalone models still outperform multilingual ones. Significant progress has led to tuning pretrained models through standard datasets in zero-shot and few-shot configurations, coupled with adaptations to address a range of Arabic-specific challenges.

4.1. Pretraining and Fine-tuning Paradigms

Pretraining consists of training a model on a large corpus of unlabeled data before performing any specific task, while fine-tuning uses supervised data to adapt a model that has already been pretrained to a task of interest. In the context of deep learning for Arabic MT, several pretraining and fine-tuning techniques and paradigms have been investigated, especially using encoder-based models. (Ghaddar et al., 2022) conducted a systematic study of pretrained language models (PLMs) for Arabic understanding that focused exclusively on pretraining objectives, masking strategies, and evaluation benchmarks by analyzing existing PLMs. They introduced five new models based on transformer architectures and BERT- or T5-style pretraining objectives. Pretraining on Arabic text led to gains on downstream tasks. In MT, non-autoregressive and other pretraining strategies enhance performance. (Almahairi et al., 2016) showed that proper preprocessing, including normalization and morphology-aware tokenization, enhances Arabic translation quality and facilitates adaptation to out-of-domain data. Finally, a recent pretraining strategy suitable for various Arabic NLP tasks was proposed. A multilingual PLM was converted into an Arabic LM by training on Arabic data and adapting BERT objectives. Results demonstrate that Arabic pretraining improves downstream task performance.

4.2. Transfer Learning and Domain Adaptation

Domain adaptation refers to the task of adapting a model trained in a general or irrelevant domain to a specific, relevant target domain (Zhang, 2017). In the context of Arabic NMT, domain adaptation aims to enhance performance on the target domain by adapting the model from either a more general Arabic dataset or a different but thematically related domain. Arabic NMT provides a crucial balance between a standard and a low-resource system. Finally, a mid-resource or closely linked language can be employed to aid university. A few-domain adaptation approach configures a homespun multi-source NMT model for low-resource applications in Arabic for three Arabic dialects and geolocated Modern Standard Arabic (Saunders, 2021).

4.3. Evaluation Protocols and Metrics for Arabic MT

Machine Translation (MT) is the task of automatically translating text from a source language (SL) to a target language (TL). Arabic MT systems must cope with several challenges, including Arabic morphology, dialectology, and script variations that hinder Arabic corpus acquisition (Ghaddar et al., 2022).

Human evaluation is the most reliable method for assessing the quality of MT systems. Automatic approaches cannot replace humans entirely but can provide complementary assessments that speed up the evaluation process. While BLEU is the most popular automatic evaluation metric for Arabic MT (Joshi et al., 2013) , it has several weaknesses.

In order to alleviate the difficulties of Arabic MT evaluation, the following metrics are recommended: BLEU, TER (Translation Edit Rate), and human evaluation for a comprehensive assessment of MT outputs.

5. Challenges and Mitigation Strategies

Research in machine translation (MT) has largely been driven by English and other major languages of the global internet. However, the Arab world today represents a growing market of over 400 million speakers of Arabic across 22 countries, many of whom are more comfortable expressing themselves in Arabic than in English, French, or other languages of wider use (Y. Habash et al., 2006). The Gulf Cooperation Council (GCC) and North African countries uniting Algeria, Libya, Mauritania, Morocco, and Tunisia, with Morocco ranked first among Arabic-speakers, additionally present new opportunities for Arabic-English MT, as many companies aspire to develop an Arabic-English MT system. Arabic dialects remain an issue, as they differ significantly from Modern Standard Arabic (MSA), the official language of 22 Arab countries and the primary language utilized in government, cross-border communications, schools, media, and most literary works. Other dialects, such as Algerian Arabic, Egyptian Arabic, Emirati Arabic, Iraqi Arabic, Jordanian Arabic, Libyan Arabic, Maghreb Arabic, Mohammadia Arabic, Tunisian Arabic, and others with regional or national variants, add further complexity. In certain countries, MSA is taught but not spoken; instead, students freely use the local variant. As MT systems trained on MSA fail entirely when vernaculars were input, solutions remain a priority (Almahairi et al., 2016).

5.1. Data Scarcity and Domain Mismatch

Low-resource languages, such as Arabic dialects and varieties, often suffer from data scarcity and domain mismatch (Almansor & Al-Ani, 2017). This means collecting a substantial amount of modern parallel data for training deep neural models is complicated; hence, researchers often focus on only a few dialects and resort to indirect and less effective methods of translation. Such languages lack large, suitable parallel corpora across multi-domain and multi-genre datasets, making it impossible to implement an end-to-end pipeline.

5.2. Arabic Morphology and Dialectal Variation

Arabic morphology is complex due to its extensive inflectional and derivational processes, which operate on a system of root consonants and patterns (Erdmann et al., 2017). This kind of morphological variation is an inherent characteristic of the language, and yet it is much more prominent in dialectal Arabic. Residential Arabic speakers employ words or phrases from dialectal Arabic in formal public data posted from others. This evidence demonstrates Arabic dialectal variation, and highlights the motivational gap between two language variants: dialetal Arabic in a multi-domain parallel corpus and Modern Standard Arabic (MSA) (Eldesouki et al., 2017).

The dialect continues to be a research-rich region of Natural Language Processing (NLP); and its segmentation is specifically a topic of considerable interest. When co-locating Arabic dialectal variation, three dialects are deemed important: Egyptian, Levantine, and Gulf. Arabic is the world’s 5th most-spoken language, and it remains essential to most English speakers; and there are many followers of Public Data, as evidently shown by the uses of other dialects in the 4-billion-parallel Corpus. Given minimal language resources of one million sentences, data-aided analyses combining either bi-LSTM-CRF or SVM are attractive; yet momentum should be built independently around specific sectorial prefixes, for instance: “ال” for an indefinite article.

5.3. Script Variants and Diacritization

Several variants of the Arabic script coexist. The most commonly used scripts are the Arabic and Persian scripts, which differ only in a few letters. However, other widely used scripts such as the Naskh, Naskhi, Maghribi, and Ruqʿa scripts differ in the shape and form of letters, diacritical marks, connected letters, ligatures, and other aspects. Furthermore, horseʿat script, a style of handwriting that uses only significant letters, poses additional challenges for Arabic character recognition. The use of the Arabic script for other languages, such as Uighur, Pashto, and Gawar̄, has led to the formation of additional character sets.

Diacritics are typically not included when writing Arabic unless needed for different interpretations of words or for emphasizing correct pronunciation. In Arabic literature, for example, diacritics are added to clarify ambiguity (Abbad & Xiong, 2020). The use of written diacritics is therefore not constant and varies based on the author, target audience, type of document, and text genre. Text without diacritics incurs a high-risk solution, especially for non-native speakers (Abdelali et al., 2018). As a result, diacritization is a challenging and complex task when dealing with Arabic texts within natural language processing (NLP) systems.

6. Evaluation and Benchmarking Studies

Deep-learning models require careful evaluation because of architecture, framework, and hyperparameter selection. Although many such models are publicly available for Arabic and other languages, satisfactory releases for pre-trained and fine-tuned models remain limited. Common evaluation criteria include robustness and usability. Quantitative evaluation using benchmarks should thus accompany qualitative appraisal using error typologies. Willingness to share platforms and resources decreases redundancy, boosts accessibility, and aids adoption.

Four Arabic machine-translation assemblies are among eight competitive multi-language systems in the WMT 2022 shared task. Robustness experiments assess the effect of domain shift and noisy data. The top encoder–decoder model, T2T, integrates pre-training (Ghaddar et al., 2022) by minimizing cross-entropy loss with multiple maskings. A dominant transformer-base encoder–decoder structure relies solely on self-attention. Another multi-lingual model, mBART, with pre-training for text denoising shares various characteristics but remains less effective than T2T.

6.1. Standard Benchmarks and Results

Neural Machine Translation (NMT) models transform the Arabic-English MT landscape, surmount earlier challenges and receiving extensive scholarly attention. The first publicly available Arabic-English NMT system, based on the simple neural probabilistic model, is introduced (Almahairi et al., 2016). A sequence-to-sequence NMT system trained on the Agence France-Presse (AFP) corpus is presented. Arabic normalization techniques for MT are revisited, showing that both phrase-based and neural models benefit from normalization and morphology-aware tokenization, with BLEU score gains exceeding two points. Generative pretraining is leveraged for NMT by jointly training a denoising autoencoder on a large extracted corpus. This transfer-learning strategy enhances BLEU scores when finetuning models on the MT05 task.

The fully attention-based Transformer architecture is applied to Arabic-English and French-English translation, demonstrating that the architecture equals or exceeds the performance of state-of-the-art recurrent models while improving training speed

(Ghaddar et al., 2022). Graph Transformer and multilingual models using pretrained cross-lingual embeddings yield competitive results through simple integration. A hybrid Arabic-English system employing a pretrained Transformer model initializes training, alleviating the data scarcity problem and boosting cross-lingual zero-shot capabilities. Adding the two languages to both Seq2Seq and Transformer frameworks sufficient to address the data limitation constraint is shown.

6.2. Error Analysis and Typology

Error analysis of Arabic Machine Translation systems, focusing on two research directions aligned with the specificity of the datasets, domains, or specific characteristics of the Arabic language. (Afli et al., 2016) review the literature on OCR error analysis in Arabic user-generated content and propose a methodology where the type of errors made by OCR systems are mapped and analysed to improve further research. Over 30 sources are reviewed and categorized by distortion type, error characterisation, and proposed corrective technique. Mistranslation of certain dialect words is a major issue for (Saadany & Orasan, 2020) applying neural machine translation to Arabic reviews. Their survey characterises the error types, which include not preserving the semantics of the reviews.

7. Applications and Use Cases

Automated translation of news and social media content has seen rapid growth in countries where Arabic is the official language. At the institutional level, the UAE Ministry of Economy and Malaysia's Bank Muamalat are both using Arabic MT systems to analyse foreign media outlets and security reports for economic and financial information relevant to national interests. Arabic MT is also aiding linguistic development, especially for educational content aimed at younger audiences. The deployment of educational web sites offering language-teaching materials, translation of children's literature, and even Arabic-language e-books and audiobooks is gaining popularity in several Middle Eastern countries and among wealthy Arab expatriates. UNICEF recently distributed more than 10,000 MT-translation units of children and youth literature and didactic material aimed at beneficiaries involved in informal Ŝtudies and leisure learning (Almansor & Al-Ani, 2017).

Building, compiling and maintaining an up-to-date corpus of highly-frequent Arabic-language terms to support formalisation of preliminary MT contracts is the initial problem underpinning MT development for humanitarian orgs. Most documents dealing with injury claims from natural disasters do not follow formal technical rules and only use the audio transcription of announcement ceremonies, leading to heavy dialectical influences on vocabulary, morphology and syntax. Since Sep 2020, a combined Arabic-to-English and English-to-Arabic MT system has been deployed for use in humanitarian contexts (Almahairi et al., 2016).

In the field of public and universal hygiene, a project entitled "Arabic Hygiene Motivation" seeks to encourage families afflicted by misfortune to protect their homes and children against epidemics. The French version of the Aiming-to-School approach to public hygiene and education in families has been translated into Arabic. Public-hygiene teaching widely disseminates 1-minute audiophonic and pictorial records. An Arabic-to-English and English-to-Arabic MT system has been set up to support preparation of audiovisual pedagogy material.

7.1. Media Translation and Localization

Arabic Media Translation (MT) remains an area of active investigation because of the importance of Arabic Media for the International community. Media Translation MT seeks to translate news reports, television news/video reports, and other forms of Media. Arabic Media Translation also aims to cover localization of news portals and websites.

Arabic Media covers all forms of audiovisual material, including film, radio, television, and the internet. It encompasses the full range of audiovisual texts, from videos for full-length feature films and independent short films to the multitude of film trailers, television series, advertisements, documentaries, and audiovisual materials for governmental and social-sector institutions. It also includes applications for text-to-speech systems as well as subtitling and voice-over translation.

The blossoming of audiovisual production in the past few years has given rise to significant levels of demand for Arabic Media MT. The large and rapidly-growing number of social media and other websites offering audiovisual content has also significantly heightened the demand. Arabic Media MT in video applications can be further classified into two categories of text-based localization and voice-based localization. The former deals with the translation of audiovisual material that features only written text while the latter deals with news reports and films that also have voice items that need to be conveyed (Almahairi et al., 2016).

7.2. Educational and humanitarian contexts

The Arab world is experiencing a constant stream of humanitarian crises that have received increasing attention from numerous organisations, ranging from governments to international NGO’s and charitable foundations. The aid thus rendered is not equally distributed in every country of that region. In some countries like Yemen, Libya, and Syrian refugees in neighbouring countries, more attention (and translation activity) is given; while in others, like many West African Arab-speaking countries or even Morocco or Algeria, still important crises are ongoing without being fully reported and thus, other information within these specific areas is not as translated.

The United Nations Development Program (UNDP), with the support of two charitable foundations, has initiated the “Arab Development Portal” pocasting project to collect video stories and news from these specific countries in order to highlight the dire situations, which have been kept silent at the international level. Through the release of these videos, it is also hoped to request for vital assistance to help these inhabitants Further translating those contents into French or any other languages of the donors or further spreading them would assist these specific peoples, which have considerably received less information and assistance. (Almansor & Al-Ani, 2017)

8. Ethical, Legal, and Social Implications

The rise of machine translation (MT) systems, and especially deep learning algorithms, has promoted wider access to knowledge in all its dimensions. This development must be viewed in light of the growing use of MT in Arabic, a predominantly spoken language that has historically lagged behind other languages in the availability of resources required adc0e86a-574c-466b-82fe-2cf42988f3e6 the development of such technologies (Almachkour et al., 2023).

The Permanent Bureau of the Hague Conference on Private International Law published a paper titled "Legal Translation and Technological Change" (McLellan, 2023), which outlines the legal, ethical, and social implications of using online neural machine translation (NMT) services, such as Google Translate. In view of the questions raised by the paper in the translation of legally binding documents or other documents where a high degree of accuracy is required, translating legal documents into or from Arabic using such services is risky. Relying solely on automated MT in any context requires a consideration of its legal, ethical, and social implications (for Translation & Literary Studies & Riyadh Rahim, 2024).

8.1. Bias, Fairness, and Accessibility

Machine translation (MT) and natural language processing (NLP) systems trained on human-generated data often learn and reproduce social biases related to sensitive attributes like gender or race. Such biases can negatively impact translation quality, particularly in sequence-to-sequence systems that utilize attention mechanisms. Various studies have demonstrated that given the same input, biased MT systems—trained on publicly available data—exhibit unequal translation quality for sentences containing male or female entities. Frameworks allowing for the straightforward investigation of biases in MT have been proposed, targeting gender stereotypes in both English-to-French and English-to-German translation.

They reveal that even state-of-the-art architecture transformers amplify pre-existing biases pertaining to occupations, style, and stereotypes, and that debiasing methods informed by word embeddings do not mitigate those biases to any meaningful degree (Escudé Font, 2019). Promising alternatives using adversarial learning have also been examined. Adversarial training alongside standard system optimization significantly reduces gender bias affecting translation quality disparity between male and female sentences, while the overall quality of translation remains unaffected (Fleisig & Fellbaum, 2022).

8.2. Privacy and Data Governance

The protection of personal data gathered during the operation of an Arabic MT system is crucial for ethical and legal practices. Privacy and Data Governance encompasses the technical and institutional actions taken to protect a user's data and guarantee governance involves the establishment of rules and procedures for an organizational entity's information assets in order to achieve the data protection objective (Almansor & Al-Ani, 2017).

Compliance with legislation such as the General Data Protection Regulation (GDPR) is one of the key reasons for implementation of Privacy and Data Governance norms (Psychoula et al., 2018). The GDPR (Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016) creates an obligation to comply for any organization that collects, stores or otherwise processes personal data of data subjects within the EU, regardless of whether it operates within the EU or not. Its provisions pose new and complex compliance challenges and risks for controller and processor organizations. Deep-learning algorithms have been increasingly utilized to improve machine learning for Natural Language Processing (Sousa & Kern, 2022).

8.3. Responsible Deployment Practices

The adoption of deep learning models into real-world systems necessitates a parallel commitment to safety and fairness from their designers. Responsible deployment practices integrate both expert knowledge and stakeholder perspectives into the development process. The implementation of Neural Machine Translation (NMT) models is further complicated by the additional challenge of dialectal variety, which has been approached by several researchers through careful attention to, and essentially independent tuning of, each Arabic dialect (Almansor & Al-Ani, 2017).

9. Future Directions

Arabic machine translation must be re-evaluated as additional perspectives, tools, methodologies, and tasks emerge (Almansor & Al-Ani, 2017). The rapid growth of Arabic-English word embeddings datastores, the emergence of transformer-based models, and the increased interest in Arabic dialects indicate substantial developments affecting the future of the field. Although only limited experimentation has been undertaken with Arabic dialects to date, the availability of parallel and directional corpora for various dialects provides clear opportunities. At the same time, parallels to material in other low-resourced languages suggest automatic conversion to even more limited settings (Alam & ul Hussain, 2017). The emergence of large multilingual models offers further avenues for exploration with Arabic; such models have been found to facilitate significant transfer possibilities, and preliminary experiments with a language-adapted version have yielded positive results (Almahairi et al., 2016).

Automated approaches to the diacritisation of Modern Standard Arabic, currently a fully rule-based task for most datasets, have already been successfully prototyped using commercial datasets as sources. The emergence of large pretrained language models has renewed interest in Arabic processing, but tasks requiring strong support from upstream modelling have been less investigated. Detailed experimentation to establish the extent to which transfer is possible and the features that facilitate or inhibit it may restore Arabic to the forefront of attention.

Arabic has been a focus area for machine translation experimentation and evaluation since the earliest days of both rule-based and data-driven work. However, as new perspectives on translation and the Arabic language have gained traction, Arabic has fallen from prominence, with an increasing number of studies examining low-resource languages and dialects that lack comparable datasets or tooling. Despite the range of Arabic studies, the lack of wide-ranging surveys suggests that a summary of existing Arabic machine translation research can provide a valuable resource for both Arabic and broader machine translation communities.

Conclusion

Although Arabic machine translation (MT) remains challenging and is generally considered a low-resource language, recent advances based on deep-learning algorithms offer promising results. Existing publications reveal important trends in Arabic MT research. A considerable amount of Arabic MT work focuses on French-Arabic and English-Arabic translation. Research on Arabic-English translation has also emerged recently. Concerning architecture, several authors employ the sequence-to-sequence (S2S) framework, primarily enhanced with attention mechanisms. Application of far more recent models based on transformers, such as the BERT pre-training paradigm and other multilingual models, is still rare but appears to be a sound direction for further experimentation given their success in other languages and tasks. Although domain adaptation and the use of large pre-trained multilingual models do not seem to benefit Arabic significantly, they remain active research fields. The use of additional monolingual data and pre-training on other related languages continue to be an integral part of the research effort on Arabic MT

References:

Almahairi, A., Cho, K., Habash, N., & Courville, A. (2016). First Result on Arabic Neural Machine Translation.

Almansor, E. H. & Al-Ani, A. (2017). Translating dialectal Arabic as low resource language using word embedding.

Wang, Q., Li, B., Xiao, T., Zhu, J., Li, C., F. Wong, D., & S. Chao, L. (2019). Learning Deep Transformer Models for Machine Translation

Moatez Billah Nagoudi, E., Elmadany, A. R., & Abdul-Mageed, M. (2021). Investigating Code-Mixed Modern Standard Arabic-Egyptian to English Machine Translation.

Albalawi, Y., Buckley, J., & S. Nikolov, N. (2021). Investigating the impact of pre-processing techniques and pre-trained word embeddings in detecting Arabic health information on social media.

Alkahtani, S. (2015). Building and verifying parallel corpora between Arabic and English.

Zarnoufi, R., Bachri, W., Jaafar, H., & Abik, M. (2022). MANorm: A Normalization Dictionary for Moroccan Arabic Dialect Written in Latin Script.

Zalmout, N. & Habash, N. (2019). Joint Diacritization, Lemmatization, Normalization, and Fine-Grained Morphological Tagging

Ghaddar, A., Wu, Y., Bagga, S., Rashid, A., Bibi, K., Rezagholizadeh, M., Xing, C., Wang, Y., Xinyu, D., Wang, Z., Huai, B., Jiang, X., Liu, Q., & Langlais, P. (2022). Revisiting Pre-trained Language Models and their Evaluation for Arabic Natural Language Understanding.

Zhang, J. (2017). Domain adaptation for statistical machine translation and neural machine translation.

Saunders, D. (2021). Domain Adaptation and Multi-Domain Adaptation for Neural Machine Translation: A Survey.

Joshi, N., Mathur, I., Darbari, H., & Kumar, A. (2013). HEVAL: Yet Another Human Evaluation Metric.

Y. Habash, N., Dorr, B., & Monz, C. (2006). Challenges in Building an Arabic-English GHMT System with SMT Components.

Erdmann, A., Habash, N., Taji, D., & Bouamor, H. (2017). Low Resourced Machine Translation via Morpho-syntactic Modeling: The Case of Dialectal Arabic.

Eldesouki, M., Samih, Y., Abdelali, A., Attia, M., Mubarak, H., Darwish, K., & Laura, K. (2017). Arabic Multi-Dialect Segmentation: bi-LSTM-CRF vs. SVM

Abbad, H. & Xiong, S. (2020). Multi-components System for Automatic Arabic Diacritization.

Abdelali, A., Attia, M., Samih, Y., Darwish, K., & Mubarak, H. (2018). Diacritization of Maghrebi Arabic Sub-Dialects.

Afli, H., Aransa, W., Lohar, P., & Way, A. (2016). From Arabic user-generated content to machine translation: integrating automatic error correction.

Saadany, H. & Orasan, C. (2020). Is it Great or Terrible? Preserving Sentiment in Neural Machine Translation of Arabic Reviews.

for Translation & Literary Studies, A. W. E. J. & Riyadh Rahim, N. (2024). Google and Legal Translation: The Case Study of Contracts.

Escudé Font, J. (2019). Determining Bias in Machine Translation with Deep Learning Techniques

Fleisig, E. & Fellbaum, C. (2022). Mitigating Gender Bias in Machine Translation through Adversarial Learning.

Psychoula, I., Merdivan, E., Singh, D., Chen, L., Chen, F., Hanke, S., Kropf, J., Holzinger, A., & Geist, M. (2018). A Deep Learning Approach for Privacy Preservation in Assisted Living.

Sousa, S. & Kern, R. (2022). How to keep text private? A systematic review of deep learning methods for privacy-preserving natural language processing

Alam, M. & ul Hussain, S. (2017). Sequence to Sequence Networks for Roman-Urdu to Urdu Transliteration.

Elmadany, A., & Abdul-Mageed, M. (2023, December). Octopus: A multitask model and toolkit for Arabic natural language generation. In Proceedings of ArabicNLP 2023 (pp. 232-243).

Kadaoui, K., Magdy, S., Waheed, A., Khondaker, M. T. I., El-Shangiti, A., Nagoudi, E. M. B., & Abdul-Mageed, M. (2023, December). Tarjamat: Evaluation of bard and chatgpt on machine translation of ten arabic varieties. In Proceedings of ArabicNLP 2023 (pp. 52-75).

Alahmari, S., Atwell, E., & Saadany, H. (2024, May). Sirius_Translators at OSACT6 2024 Shared Task: Fin-tuning Ara-T5 Models for Translating Arabic Dialectal Text to Modern Standard Arabic. 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 (pp. 117-123). ELRA and ICCL.

Al-Khalifa, H., Al-Khalefah, K., & Haroon, H. (2024). Error analysis of pretrained language models (PLMs) in English-to-Arabic machine translation. Human-Centric Intelligent Systems.

Hennara, K., Hreden, M., Hamed, M. M., Aldallal, Z., Chrouf, S., & AlModhayan, S. (2025). Mutarjim: Advancing Bidirectional Arabic-English Translation with a Small Language Model. arXiv preprint arXiv:2505.17894.

Hammouda, T., Jarrar, M., & Khalilia, M. (2024). SinaTools: Open Source Toolkit for Arabic Natural Language Processing. Procedia Computer Science.

Elmadany, A., & Abdul-Mageed, M. (2022, June). TURJUMAN: A public toolkit for neural Arabic machine translation. In Proceedinsg of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection (pp. 1-11).

Elhamayed, S. A., & Nour, M. (2025). Overview of deep learning and large language models in machine translation: a special perspective on the Arabic language. Journal of Electrical Systems and Information Technology, 12(1), 27.

Al-Ghawanmeh, F., Jensenius, A., & Smaili, K. (2023, June). How can machine translation help generate Arab melodic improvisation?. In Proceedings of the 24th Annual Conference of the European Association for Machine Translation (pp. 385-392).

Elmadany, A., Adebara, I., & Abdul-Mageed, M. (2024). Toucan: Many-to-many translation for 150 african language pairs. arXiv preprint arXiv:2407.04796.