



## OPEN ACCESS

EDITED AND REVIEWED BY  
Arkaitz Zubiaga,  
Queen Mary University of London,  
United Kingdom

\*CORRESPONDENCE  
Shadi Abudalfa  
✉ shadi\_abudalfa@hotmail.com

RECEIVED 29 September 2025

ACCEPTED 17 October 2025

PUBLISHED 06 November 2025

## CITATION

Abudalfa S, Saad M and El-Beltagy S (2025)  
Editorial: Emerging techniques in Arabic  
natural language processing.  
*Front. Artif. Intell.* 8:1715520.  
doi: 10.3389/frai.2025.1715520

## COPYRIGHT

© 2025 Abudalfa, Saad and El-Beltagy. This is  
an open-access article distributed under the  
terms of the [Creative Commons Attribution  
License \(CC BY\)](#). The use, distribution or  
reproduction in other forums is permitted,  
provided the original author(s) and the  
copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Editorial: Emerging techniques in Arabic natural language processing

Shadi Abudalfa<sup>1\*</sup>, Motaz Saad<sup>2</sup> and Samhaa El-Beltagy<sup>3</sup>

<sup>1</sup>SDAIA-KFUPM JRC for Artificial Intelligence, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia, <sup>2</sup>Department of Data Science, Islamic University of Gaza, Gaza, Palestine, <sup>3</sup>School of Information Technology, Newgiza University, Giza, Egypt

## KEYWORDS

Arabic NLP, detection, recognition, opinion mining, syntactic analyzers, machine translation, LLMs

## Editorial on the Research Topic

### Emerging techniques in Arabic natural language processing

## Introduction

Arabic Natural Language Processing (NLP) is a rapidly growing field focusing on the unique computational and linguistic challenges of the Arabic language. Recent progress has been driven by deep learning approaches and the increasing use of large language models (LLMs), which have improved applications such as sentiment analysis, text processing, speech recognition, and machine translation (Haboussi et al., 2025; Abdu et al., 2025). Despite these advances, the field still faces critical obstacles, including a shortage of annotated datasets, insufficient tools for dialect handling, and the limited availability of Arabic-oriented LLMs (Mashaabi et al., 2024; Dahou et al., 2025; Abudalfa et al., 2024). This Research Topic presents studies covering various aspects of Arabic NLP, such as syntactic analysis, dialect identification, stance classification, and other tasks that contribute to practical real-world solutions.

## Key contributions

The studies featured in this Research Topic highlight advancements in Arabic NLP and introduce innovative approaches within this field. The following subsections provide a concise overview of each paper included.

## Syntactic analyzers

Syntactic analysis is a core task in NLP, particularly vital for morphologically rich languages like Arabic. Saadiyeh et al. compared a range of Arabic syntactic analyzers, from rule-based, statistical, and machine learning approaches to hybrid, neural, and transformer-based models, examining their strengths, weaknesses, and trade-offs. The complexity of Arabic morphology and syntax makes accurate parsing challenging, which they address through a detailed evaluation of algorithms and their reliance on high-quality annotated resources.

## Machine translation

Algaraady and Mahyoob conducted a study comparing Arabic translations of Google Translate after post-editing by two professional translators and ChatGPT-4o, with three experts evaluating the final output. Quality was assessed through fluency, accuracy, coherence and efficiency, and a paired *t*-test analyzed the differences. Human post-editing generally yielded superior quality, while ChatGPT-4o stood out for speed and produced fluently flowing coherent translations.

In a related line of research, Beidas et al. examine the performance of GPT-3.5, GPT-4, and Bard (Gemini) on the QADI and MADAR datasets, whereas GPT-5 was tested solely on MADAR, which covers data from more than 15 countries. The evaluation relied on several metrics, including cosine similarity, the universal similarity encoder, sentence-BERT, TER, ROUGE, and BLEU. Two prompting strategies were applied: zero-shot and few-shot.

## Opinion mining

Alkhathlan et al. presented ArabicStanceX, a large dataset for stance detection with 14,477 tweets covering 17 topics. Using the transformer-based MARBERTv2 model and a Multi-Topic Single Model approach, they achieved an F1 score of 0.74 for “favor” and “against” categories and 0.67 overall. Their results reveal strengths in stance classification but also difficulties with neutral labels and unseen topics. Additional zero-shot and few-shot learning tests show the model’s flexibility in adapting to new subjects.

Jaber et al. explored the use of ensemble-based machine learning approaches for Arabic sentiment classification. A range of homogeneous ensemble models is developed and tested on two corpora: the balanced ArTwitter dataset and the highly skewed Syria\_Tweets dataset. To address the imbalance problem, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. The experiments combine unigram features with pre-trained word embedding representations.

## Arabic poetry

Mutawa and Alrumaih presented a deep learning technique for identifying the meter of Arabic poetry using a large annotated dataset. Text was encoded at the character level to classify full and half verses without removing diacritics, ensuring that essential linguistic features were preserved. Various neural network architectures, including LSTM, GRU, and Bi-LSTM, were explored. This framework demonstrates a robust approach to Arabic meter recognition and highlights the potential of AI in NLP.

## Dialect detection

Saleh et al. presented a stacking-based technique to improve dialectal Arabic classification by combining two transformer

models, Bert-Base-Arabertv02 and Dialectal-Arabic-XLM-R-Base. The technique involves two layers: the first generates class probabilities from the transformers, which are then used by a meta-learner in the second layer. This technique was benchmarked against individual models such as LSTM, GRU, CNN, and single transformers with various embeddings. Experimental results demonstrated that the combined model outperforms single-model methods by capturing a wider range of linguistic features, improving generalization across Arabic varieties.

## Speech recognition

Al-Anzi and Thankaleela presented an Arabic speech recognition framework that begins by extracting Mel-frequency cepstral coefficients (MFCCs) from audio signals. These features are then grouped through K-means clustering, and the resulting clusters are classified using methods such as Decision Trees, Random Forests, K-Nearest Neighbors, and XGBoost. For demonstration purposes, both Euclidean Distance and Dynamic Time Warping (DTW) are employed. Additionally, the research highlights the effectiveness of Mozilla’s DeepSpeech framework in handling Arabic speech recognition.

## Cyberbullying detection

Allwaibed et al. reviewed 35 scholarly articles addressing the detection of cyberbullying in Arabic-language texts. From a methodological standpoint, traditional machine learning approaches that leverage Arabic-specific linguistic features continue to perform well on smaller datasets. However, more advanced deep learning models and transformer-based frameworks such as AraBERT achieve stronger results, especially when challenges like dialectal variation and orthographic inconsistencies are reduced.

## Conclusion

The studies gathered in this Research Topic illustrate the diversity and dynamism of ongoing efforts in Arabic NLP. Collectively, these contributions showcase how deep learning and LLMs are driving progress in Arabic NLP, while also pointing to persistent obstacles such as dialectal differences, scarcity of annotated data, and specialized domain challenges. By introducing innovative approaches, releasing new datasets, and offering comparative assessments, the featured works not only push the field forward but also stress the importance of sustained collaboration, resource creation, and tool development to enhance Arabic NLP and extend its practical impact.

## Author contributions

SA: Writing – original draft, Writing – review & editing, MS: Writing – review & editing, SE-B: Writing – review & editing.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. ChatGPT was used to rephrase some sentences to improve readability. The author(s) reviewed and take full responsibility for the content.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the

support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Abdu, F. J., Mughaus, R., Abudalfa, S., Ahmed, M., and Abdelali, A. (2025). An empirical evaluation of Arabic text formality transfer: a comparative study. *Lang. Resour. Eval.* 1–61. doi: 10.1007/s10579-025-09873-w
- Abudalfa, S. I., Abdu, F. J., and Alowaiifeer, M. M. (2024). Arabic text formality modification: a review and future research directions. *IEEE Access* 12, 185117–185148. doi: 10.1109/ACCESS.2024.3511661
- Dahou, A., Dahou, A. H., Cheragui, M. A., Abdedaïem, A., Al-qaness, M. A., Abd Elaziz, M., et al. (2025). A survey on dialect Arabic processing and analysis: recent advances and future trends. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* 24:84. doi: 10.1145/3747290
- Haboussi, S., Oukas, N., Zerrouki, T., and Djettou, H. (2025). Arabic speech recognition using neural networks: concepts, literature review and challenges. *J. Umm Al-Qura Univ. Appl. Sci.* 1–23. doi: 10.1007/s43994-025-00213-w
- Mashaabi, M., Al-Khalifa, S., and Al-Khalifa, H. (2024). A survey of large language models for Arabic language and its dialects. *arXiv [Preprint]*. arXiv:2410.20238. doi: 10.48550/arXiv.2410.20238