



LREC 2026

**The 7th Workshop on Open-Source Arabic Corpora and
Processing Tools (OSACT7) with 5 Shared Tasks**

Workshop Proceedings

Editors
Hend Al-Khalifa
Mo El-Haj
Saad Ezzini

May 11, 2026

Proceedings of The 7th Workshop on Open-Source Arabic Corpora and Processing Tools
(OSACT7) with 5 Shared Tasks

©ELRA Language Resources Association (ELRA), 2026
These proceedings are licensed under a Creative Commons Attribution-
NonCommercial 4.0 International License (CC BY-NC 4.0)

ISBN 978-2-493814-52-4
EAN 9782493814524

Preface

Welcome to the 7th Workshop on Open-Source Arabic Corpora and Processing Tools (OSACT2026), co-located with LREC 2026. This workshop brings together researchers and practitioners working on Arabic Natural Language Processing (NLP), with a focus on open-source resources, tools, and systems that advance the state of the art in Arabic language technologies.

Arabic NLP continues to present unique and exciting challenges, driven by the rich morphological complexity of the language, the diversity of its dialects, and the growing demand for high-quality tools and resources across a wide range of applications. At the same time, the rapid advancement of Large Language Models (LLMs) has opened new research directions, raising important questions about their effectiveness, limitations, and adaptability for Arabic.

For this edition, we received 35 submissions covering a broad range of topics. After a thorough double-blind review process, we accepted 16 papers, yielding an acceptance rate of approximately 46%. The accepted papers are organized into two main thematic sessions: the first focusing on Sentiment Analysis, Information Retrieval, and Language Resources, and the second on Machine Translation, Dialects, and LLM Applications.

In addition, the workshop hosted five shared tasks spanning Arabic politeness detection, Islamic inheritance reasoning, speech dictation with automatic diacritization, humor generation, and sentiment analysis. These shared tasks attracted a total of 33 system paper submissions, reflecting the strong community engagement and growing interest in building shared benchmarks and evaluation frameworks for Arabic NLP.

We thank the members of the program committee for their thorough and thoughtful reviews. We also thank the organizers of LREC 2026 for their continued support. Most importantly, we thank all the authors and participants whose contributions and enthusiasm made this workshop a vibrant and valuable venue for the Arabic NLP community.

Hend Al-Khalifa
Mo El-Haj
Saad Ezzini

Organizing Committee

Hend Al-Khalifa, King Saud University (KSU), Saudi Arabia
Mo El-Haj, Lancaster University, United Kingdom; VinUniversity, Vietnam
Saad Ezzini, King Fahd University of Petroleum and Minerals (KFUPM), Saudi Arabia

Program Committee

Mohammed Alliheedi, Al-Baha University, Saudi Arabia
Mamoun Abu Helou, Al-Istiqlal University, Palestine
Maria Bounnit, Cadi Ayyad University, Morocco
Hani M. Iwida, Al-Istiqlal University, Palestine
Noorhan Abbas, University of Leeds, United Kingdom
Sultan Alrowili, IBM Research, Saudi Arabia
Abdulaziz Alhamadani, Florida Polytechnic University, United States
Amal Haddad Haddad, University of Granada, Spain
Dima Taji, Charles University, Czech Republic
Hamada Nayel, Prince Sattam Bin Abdulaziz University, Saudi Arabia
Salmane Chafik, Mohammed VI Polytechnic University, Morocco
Abdessalam Boucekif, Hamad Bin Khalifa University, Qatar
Sharefah Al-Ghamdi, King Saud University (KSU), Saudi Arabia
Akram Mohammed Ahmed Al-Rumaim, Independent Researcher, Netherlands
Ashraf Elnagar, University of Sharjah, United Arab Emirates
Ameera Almasoud, King Saud University (KSU), Saudi Arabia
Imed Zitouni, Meta, United States
Mohammad Abuoudeh, Al-Hussein Bin Talal University, Jordan
Wajdi Zaghouani, Northwestern University in Qatar, Qatar
Rana Malhas, Qatar University, Qatar
Ehsan Lotfi, University of Antwerp, Belgium
Abdelkader El Mahdaoui, Mohammed VI Polytechnic University, Morocco
Khalid Choukri, ELRA/ELDA, France
Khalid Al-Khatib, University of Groningen, Netherlands
Amany Fashwan, Alexandria University, Egypt
Salima Harrat, ENS Bouzaréah, Algeria
Manal Albahlal, King Saud University (KSU), Saudi Arabia
Wafa Aissa, UCLouvain, Belgium
Bassam Haddad, University of Petra, Jordan
Noof Alfear, King Saud University (KSU), Saudi Arabia
Amr Keleg, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI), United Arab Emirates
Sara Nabhani, University of Groningen, Netherlands
Ali Al-Laith, University of Copenhagen, Denmark
Fethi Bougares, Avignon University, France
Amal Htair, Aston University, United Kingdom
Nourah Alangari, King Saud University (KSU), Saudi Arabia
Nadia Ghezaiel, University of Ha'il, Saudi Arabia
Khlood Al Jallad, Syrian Society for Startups and Research, Syria
Paul Rayson, Lancaster University, United Kingdom

Hamza Alami, Sidi Mohamed Ben Abdellah University (USMBA), Morocco
Abdulhamid Abubakar, Nasarawa State University (NSU), Nigeria
Kamel Gaanoun, National Institute of Statistics and Applied Economics (INSEA), Morocco
Mariam M. Biltawi, Al Hussein Technical University, Jordan
Abdullah I. Alharbi, King Salman Global Academy for Arabic Language, Saudi Arabia
Marwah Alian, World Islamic Sciences and Education University, Jordan
Sarah Alnefaie, King Abdulaziz University, Saudi Arabia
Hatim Derrouz, Ibn Tofail University, Morocco
Nada Ghneim, Damascus University, Syria
Taha Zerrouki, University of Bouira, Algeria
Fouzi Takelait, University of North Dakota, United States
Serry Sibae, Prince Sultan University, Saudi Arabia
Yassine Saoudi, Al-Ahliyya Amman University, Jordan
Hadeel Saadany, Birmingham City University, United Kingdom
Salima Mdhaffar, Avignon University, France
Nouran Khallaf, University of Leeds, United Kingdom
Amir Hazem, The University of Tokyo, Japan
Osama Hamed, Palestine Technical University, Palestine

Table of Contents

<i>Hidden Sentiments: The Impact of Low-level Adversarial Perturbations on Arabic Sentiment Analysis Services</i>	
Abdelrahman Hamada Hefny Abdelkader	1
<i>LLM-Based Financial Sentiment Analysis in Arabic: Evidence from Saudi Markets</i>	
Mona H. Albaqawi, Eman M. Albalkhi, Joud A. Albaiti and Enrico Lopedoto	14
<i>Does Translation Preserve Sentiment? An Analysis of Arabic-English Cross-Lingual Classification</i>	
Nour Aldin Al Mubarak and Noura Al Moubayed	25
<i>When Bigger Isn't Better: Evaluating LLMs for Arabic Sentiment Analysis</i>	
Mohamed Ibrahim, Abdullah Makki, Youssef Barakat, Nour Samy and Sarah AlHumoud	35
<i>GATE-Reranker: A Strong Arabic Cross-Encoder for Document Reranking</i>	
Omer Nacar, Omar Elshehy, Mohamed Zaytoon and Khlood Al Jallad	40
<i>How Foundation Models Behave for Arabic Image Captioning?</i>	
Khaoula Dahimi, Amel Belabbaci, Hadda Cherroun and Abdelhamid Haouhat	49
<i>AlignAR: Generative Sentence Alignment for Arabic–English Parallel Corpora of Legal and Literary Texts</i>	
Baorong Huang and Ali Asiri	59
<i>Helpful or Harmful? The Dual Role of Linguistic Features in LLM-Based Dialectal Machine Translation</i>	
Abdelhalim Hafedh Dahou and Mohamed Amine Cheragui	66
<i>ASCAT: Arabic Scientific Benchmark for Advanced Translation Evaluation</i>	
Serry Sibae, Khlood Al Jallad, Zineb Yousfi, Israa Elhosiny, Yousra Yousra El-Ghawi, Batool Balah and Omer Nacar	76
<i>SHEINfer: Implicit Product Category Inference from Arabic E-commerce Reviews</i>	
Hend Al-Khalifa	81
<i>NAJD-MT: High-Fidelity Saudi Najdi–English Training Data for Bidirectional Neural Machine Translation</i>	
Nour Qandos, Samar Essa Ahmed, Omer Nacar, ahmad alrabghi, Rahaf Saeed Al Hallay, Aya Hamod and Shaden Alsuhaim	88
<i>Parsing Arabic Dialects Revisited: New Benchmarks, Models, and Insights</i>	
Ahmed Farouk Zakaria Elshabrawy, Go Inoue, Muhammed AbuOdeh and Nizar Habash	94
<i>On LLM Prompting Techniques for Arabic Language Arithmetic Reasoning</i>	
Reem Alenezi and Ayed Atallah Salman	106
<i>DIA2 - a Comprehensive and Diverse Diacritized Arabic Corpus for NLP Research</i>	
Fatima Dekmak, Shady Elbassuoni, Khaled Shaban, Hazem Hajj, Wassim El-Hajj, Yasmine Abu Adla and Buthaina Alabrash	115

<i>CV-18 NER: Augmented Common Voice for Named Entity Recognition from Arabic Speech</i> youssef saidi, Haroun Elleuch and Fethi Bougares	131
<i>ARHAHA 2026: The Shared Task on Arabic Humor Automatic Generation</i> Ameera Masoud Almasoud, Hend Al-Khalifa, Reem Fahad Alqifari, Nourah Alangari and Manal M. Albahlal	141
<i>The AdabEval 2026 Shared Task on Arabic Politeness Detection</i> Reem Fahad Alqifari, Hend Al-Khalifa, Nadia Ghezaiel, Maria Bounnit, Hend Hamed Al- hazmi, Ameera Masoud Almasoud, Sharefah Ahmed Al-Ghamdi and Noof Abdullah Alfear	151
<i>GHAD NLP at AdabEval2026: Transformer-Based Approach for Arabic Politeness and Prag- matic Category Classification</i> Ghada Alfattni and ghader kurdi	157
<i>MOSKA-NLP at AdabEval 2026: Feature-Enriched Ensembling for Arabic Politeness Detection</i> Nina A. Andriyanova-Almaamary	165
<i>SHU at AdabEval 2026: Category-Aware Fine-Tuning of MARBERT for Arabic Politeness and Pragmatic Function Classification</i> Alla Zwawi and Stephen Wu	174
<i>Comparative Study of Machine Learning and Transformer-Based Approaches for Arabic Polite- ness Detection at AdabEval 2026</i> mariem ben arbia, Ghada Ben Amor and Omar Trigui	179
<i>AdabEval 2026 Task B : Multi-Label Classification of Arabic Politeness Criteria in Social Media Media</i> Rand Abdullah Alturki	185
<i>QIAS 2026: Overview of the Shared Task on Islamic Inheritance Reasoning</i> Abdessalam Boucekif, Somaya Eltanbouly, Shahd Gaben, Mohammed Ghaly, Samer Rashwani, MOHAMED Emad and Heba Sbahi	191
<i>QU-NLP at QIAS 2026: Multi-Stage QLoRA Fine-Tuning for Arabic Islamic Inheritance Reason- ing</i> Mohammad ALSmadi	199
<i>PSL at QIAS 2026: Which Models Perform Better in Arabic Inheritance Reasoning?</i> amine Mohammed and Chahinez Boucekif	204
<i>AGS-KSU at QIAS 2026: A Comparative Study of Prompting and LLM Approaches for Structured Islamic Inheritance Reasoning</i> Hicham Ghazi Sidaoui	209
<i>Silah at QIAS 2026: Fine-Tuning vs. Retrieval-Augmented Generation for Islamic Inheritance Reasoning</i> Ghader Kurdi, Hanan Justanieah and Hala Justanieah	213

KSAA-2026 Shared Task on Arabic Speech Dictation with Automatic Diacritization

Asma Ali Al Wazrah, Waad Alshammari, Rawan Almatham, Raghad Al-rasheed, Afrah Abdulaziz Altamimi, Rufael Marew, Sawsan Alqahtani, Hanan Aldarmaki, Abdullah I. Alharbi, Abdulrahman Saeed Alshehri, Mohamed Assar, Amal Almazrui and Abdulrahman Alosaimy
220

Thaka at KSAA-2026 Task 2: Regularized Fine-Tuning for Arabic Speech Diacritization

Meshal Abdullah Alamr, Hassan Rshed Alqaeri and Abdullah Aldahlawi 225

TantaArabNLP at KSAA-2026 Task 2: Adapting CATT-Whisper for Arabic Speech Dictation with Automatic Diacritization

Nada Adel Esmaeil, Reda M. Elbasiony and Mohamed T. Faheem 229

Fine-Tashkeel at KSAA-2026: A Comprehensive Evaluation of Seq2Seq and Multimodal Approaches for Automatic Diacritization of Arabic Speech Dictation

Hassan Barmandah, Fatimah Emad Eldin, Omer Nacar and Wareef Alzubaidi 234

Eraserhead at OSACT7 Shared Task: ASR Consistency Filtering and Speaker-Adaptive Post-Processing for Arabic Speech Diacritization

Muhammad Abu Horaira and Nahian Chowdhury 247

Abjad AI at KSAA-2026 Shared Task 2: Grouped Speech Conditioning for Arabic Diacritization

Naif Saad Alharthi, Ahmad Ghannam, Faris Alasmay, Kholood Al Tabash, Shouq Sadah and Lahouari Ghouti 252

AraSentEval 2026: A Shared Task on Sentiment Analysis and Swapping in Arabic

Saad Ezzini, Shadi Abudalfa, Maram I. Alharbi, Salmane Chafik, Hamzah Luqman, Mo El-Haj, Paul Rayson and Reem Alotaibi 256

TTLab at AraSentEval: SARF Sentiment Analysis via Root-based Fusion for Multi-Dialectal Arabic

Ali Abusaleh, Bhuvanesh Verma and Alexander Mehler 262

A Comparative Study of Arabic Sentiment Swap Models for AraSentEval 2026

Yumna Hamdy, Mohab ElDamhougy, Yomna Eid and Ensaf Hussein 269

L3IA at AraSentEval 2026 Subtask 2: LLM-Based Multi-Step Pipeline for Arabic Sentiment Swap

Abdessamad Benlahbib, Hamza Alami, Mohamed M'haouach and Kaouthar Elyoussoufi
274

CasbAI at AraSentEval 2026: Robust Dialectal Arabic Sentiment Classification via Multi-Seed Ensembling and Data Augmentation.

chaima abdelaziz, KahinaHouda Saadaoui, Faiza BELBACHIR and Lynda Said Lhadj 278

BDSI at AraSentEval Shared Task : A Multi-Transformer Contrastive Learning for Arabic Dialect Sentiment Analysis

Mohamed M'haouach, Kaouthar Elyoussoufi, Abdessamad Benlahbib and Hamza Alami
284

Codezone Research Group at AraSentEval Shared Task: Arabic Sentiment Swap beyond Negation Prepending, Benchmarking Multilingual T5 against Large Language Models on the MA'AKS Corpus

Abdulkadir Shehu Bichi and Sarah Yassine 288

L3IA-Subtask 1 at AraSentEval Shared Task: Multi-Dialect Arabic Sentiment Classification via a Transformer-Based Approach

Mohamed M'haouach, Kaouthar Elyoussoufi, Hamza Alami and Abdessamad Benlahbib
292

University of Tripoli at AraSentEval: Fine-Tuning MARBERTv2 and CAMELBERT for Multi-Dialect Arabic Sentiment Analysis

Abdusalam F. Ahmad Nwesri, Amani Bahlul Sharif and Sarah Farag S. Hmeid 296

LinguArabic at AraSentEval 2026: MARBERT for Multi-Dialect Arabic Sentiment Analysis

Norah Saud Alshahrani, Elham Abdullah Al-Qarni and Shatha Hussan Alshomrani . . . 302

Workshop Program

09:00–09:05 Opening Remarks

Session 1: Sentiment Analysis, Information Retrieval & Language Resources

09:05–10:25

- | | |
|-------------|---|
| 09:05–09:15 | Hidden Sentiments: The Impact of Low-level Adversarial Perturbations on Arabic Sentiment Analysis Services |
| 09:15–09:25 | LLM-Based Financial Sentiment Analysis in Arabic: Evidence from Saudi Markets |
| 09:25–09:35 | When Bigger Isn't Better: Evaluating LLMs for Arabic Sentiment Analysis |
| 09:35–09:45 | Does Translation Preserve Sentiment? An Analysis of Arabic-English Cross-Lingual Classification |
| 09:45–09:55 | GATE-Reranker: A Strong Arabic Cross-Encoder for Document Reranking |
| 09:55–10:05 | How Foundation Models Behave for Arabic Image Captioning? |
| 10:05–10:15 | AlignAR: Generative Sentence Alignment for Arabic–English Parallel Corpora of Legal and Literary Texts |
| 10:15–10:25 | Helpful or Harmful? The Dual Role of Linguistic Features in LLM-Based Dialectal Machine Translation |

10:30–11:00 Coffee Break

Session 2: Machine Translation, Dialects & LLM Applications

11:00–12:15

- | | |
|-------------|--|
| 11:00–11:12 | ASCAT: Arabic Scientific Benchmark for Advanced Translation Evaluation |
| 11:12–11:24 | NAJD-MT: High-Fidelity Saudi Najdi–English Training Data for Bidirectional Neural Machine Translation |
| 11:24–11:36 | On LLM Prompting Techniques for Arabic Language Arithmetic Reasoning |
| 11:36–11:48 | Parsing Arabic Dialects Revisited: New Benchmarks, Models, and Insights |
| 11:48–12:00 | DIA2 – a Comprehensive and Diverse Diacritized Arabic Corpus for NLP Research |
| 12:00–12:12 | CV-18 NER: Augmented Common Voice for Named Entity Recognition from Arabic Speech |

Workshop Program (Cont.)

Session 3: Arabic NLP Shared Tasks

12:15–12:55

12:15–12:23 **ARHAHA 2026: The Shared Task on Arabic Humor Automatic Generation**

12:23–12:35 **The AdabEval 2026 Shared Task on Arabic Politeness Detection**
• GHAD NLP at AdabEval 2026 (lightning talk)

12:35–12:43 **QIAS 2026: Overview of the Shared Task on Islamic Inheritance Reasoning**
• QU-NLP at QIAS 2026 (lightning talk)

12:43–12:50 **KSAA-2026 Shared Task on Arabic Speech Dictation with Automatic Diacritization**
• Thaka at KSAA-2026 (lightning talk)

12:50–12:55 **AraSentEval 2026: A Shared Task on Sentiment Analysis and Swapping in Arabic**
• TTLab at AraSentEval 2026 (lightning talk)

12:55–13:00 Closing Remarks