

# **Seventh Workshop on Arabic Natural Language Processing (WANLP 2022)**

Abu Dhabi, United Arab Emirates  
8 December 2022

ISBN: 978-1-7138-9152-9

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2022) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

For permission requests, please contact the Association for Computational Linguistics  
at the address below.

Association for Computational Linguistics  
209 N. Eighth Street  
Stroudsburg, Pennsylvania 18360

Phone: 1-570-476-8006  
Fax: 1-570-476-0860

[acl@aclweb.org](mailto:acl@aclweb.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>CaraNER: The COVID-19 Arabic Named Entity Corpus</i> Abdulmohsen Al-Thubaity, Sakhar Alkhereyf, Wejdan Alzahrani and Alia Bahanshal . . . . .	1
<i>Joint Coreference Resolution for Zeros and non-Zeros in Arabic</i> Abdulrahman Aloraini, Sameer Pradhan and Massimo Poesio . . . . .	11
<i>SAIDS: A Novel Approach for Sentiment Analysis Informed of Dialect and Sarcasm</i> Abdelrahman Kaseb and Mona Farouk . . . . .	22
<i>AraBART: a Pretrained Arabic Sequence-to-Sequence Model for Abstractive Summarization</i> Moussa Kamal Eddine, Nadi Tomeh, Nizar Habash, Joseph Le Roux and Michalis Vazirgiannis 31	
<i>Towards Arabic Sentence Simplification via Classification and Generative Approaches</i> Nouran Khallaf, Serge Sharoff and Rasha Soliman . . . . .	43
<i>Generating Classical Arabic Poetry using Pre-trained Models</i> Nehal Elkaref, Mervat Abu-Elkheir, Maryam ElOraby and Mohamed Abdelgaber . . . . .	53
<i>A Benchmark Study of Contrastive Learning for Arabic Social Meaning</i> Md Tawkat Islam Khondaker, El Moatez Billah Nagoudi, AbdelRahim Elmadany, Muhammad Abdul-Mageed and Laks Lakshmanan, V.S. . . . .	63
<i>Adversarial Text-to-Speech for low-resource languages</i> Ashraf Elneima and Mikołaj Bińkowski . . . . .	76
<i>NADI 2022: The Third Nuanced Arabic Dialect Identification Shared Task</i> Muhammad Abdul-Mageed, Chiyu Zhang, AbdelRahim Elmadany, Houda Bouamor and Nizar Habash . . . . .	85
<i>The Shared Task on Gender Rewriting</i> Bashar Alhafni, Nizar Habash, Houda Bouamor, Ossama Obeid, Sultan Alrowili, Dalayah AlZeer, Kawla Mohamad Shnqiti, Ahmed Elbakry, Muhammad ElNokrashy, Mohamed Gabr, Abderrahmane Issam, Abdelrahim Qaddoumi, Vijay Shanker and Mahmoud Zyate . . . . .	98
<i>Overview of the WANLP 2022 Shared Task on Propaganda Detection in Arabic</i> Firoj Alam, Hamdy Mubarak, Wajdi Zaghouni, Giovanni Da San Martino and Preslav Nakov	108
<i>ArzEn-ST: A Three-way Speech Translation Corpus for Code-Switched Egyptian Arabic-English</i> Injy Hamed, Nizar Habash, Slim Abdennadher and Ngoc Thang Vu . . . . .	119
<i>Maknuune: A Large Open Palestinian Arabic Lexicon</i> Shahd Salah Uddin Dibas, Christian Khairallah, Nizar Habash, Omar Fayeze Sadi, Tariq Sairafy, Karmel Sarabta and Abrar Ardah . . . . .	131
<i>Developing a Tag-Set and Extracting the Morphological Lexicons to Build a Morphological Analyzer for Egyptian Arabic</i> Amany Fashwan and Sameh Alansary . . . . .	142
<i>A Weak Supervised Transfer Learning Approach for Sentiment Analysis to the Kuwaiti Dialect</i> Fatemah Husain, Hana Al-Ostad and Halima Omar . . . . .	161
<i>Mawqif: A Multi-label Arabic Dataset for Target-specific Stance Detection</i> Nora Saleh Alturayef, Hamzah Abdullah Luqman and Moataz Aly Kamaleldin Ahmed . . . . .	174

<i>Assessing the Linguistic Knowledge in Arabic Pre-trained Language Models Using Minimal Pairs</i> Wafa Abdullah Alrajhi, Hend Al-Khalifa and Abdulmalik AlSalman .....	185
<i>Identifying Code-switching in Arabizi</i> Safaa Shehadi and Shuly Wintner .....	194
<i>Authorship Verification for Arabic Short Texts Using Arabic Knowledge-Base Model (AraKB)</i> Fatimah Alqahtani and Helen Yannakoudakis .....	205
<i>A Semi-supervised Approach for a Better Translation of Sentiment in Dialectal Arabic UGT</i> Hadeel Saadany, Constantin Orăsan, Emad Mohamed and Ashraf Tantawy .....	214
<i>Cross-lingual transfer for low-resource Arabic language understanding</i> Khadige Abboud, Olga Golovneva and Christopher DiPersio .....	225
<i>Improving POS Tagging for Arabic Dialects on Out-of-Domain Texts</i> Noor Abo Mokh, Daniel Dakota and Sandra Kübler .....	238
<i>Domain Adaptation for Arabic Crisis Response</i> Reem Alrashdi and Simon O’Keefe .....	249
<i>Weakly and Semi-Supervised Learning for Arabic Text Classification using Monodialectal Language Models</i> Reem AlYami and Rabah Al-Zaidy .....	260
<i>Event-Based Knowledge MLM for Arabic Event Detection</i> Asma Z Yamani, Amjad K Alsulami and Rabeah A Al-Zaidy .....	273
<i>Establishing a Baseline for Arabic Patents Classification: A Comparison of Twelve Approaches</i> Taif Omar Al-Omar, Hend Al-Khalifa and Rawan Al-Matham .....	287
<i>Towards Learning Arabic Morphophonology</i> Salam Khalifa, Jordan Kodner and Owen Rambow .....	295
<i>AraDepSu: Detecting Depression and Suicidal Ideation in Arabic Tweets Using Transformers</i> Mariam Hassib, Nancy Hossam, Jolie Sameh and Marwan Turki .....	302
<i>End-to-End Speech Translation of Arabic to English Broadcast News</i> Fethi Bougares and Salim Jouili .....	312
<i>Arabic Keyphrase Extraction: Enhancing Deep Learning Models with Pre-trained Contextual Embedding and External Features</i> Randah Alharbi and Husni Al-Muhtasab .....	320
<i>ArABIE: Joint Entity, Relation and Event Extraction for Arabic</i> Niama El Khbir, Nadi Tomeh and Thierry Charnois .....	331
<i>Emoji Sentiment Roles for Sentiment Analysis: A Case Study in Arabic Texts</i> Shatha Ali A. Hakami, Robert Hendley and Phillip Smith .....	346
<i>Gulf Arabic Diacritization: Guidelines, Initial Dataset, and Results</i> nouf alabbasi, Mohamed Al-Badrashiny, Maryam Aldahmani, Ahmed AlDhanhani, Abdullah Saleh Alhashmi, Fawaghy Ahmed Alhashmi, Khalid Al Hashemi, Rama Emad Alkhobbi, Shamma T Al Maazmi, Mohammed Ali Alyafeai, Mariam M Alzaabi, Mohamed Saqer Alzaabi, Fatma Khalid Badri, Kareem Darwish, Ehab Mansour Diab, Muhammad Morsy Elmallah, Amira Ayman Elnashar, Ashraf Hatim Elneima, MHD Tameem Kabbani, Nour Rabih, Ahmad Saad and Ammar Mamoun Sousou .	356

<i>Learning From Arabic Corpora But Not Always From Arabic Speakers: A Case Study of the Arabic Wikipedia Editions</i>	
Saied Alshahrani, Esma Wali and Jeanna Matthews	361
<i>A Pilot Study on the Collection and Computational Analysis of Linguistic Differences Amongst Men and Women in a Kuwaiti Arabic WhatsApp Dataset</i>	
Hesah Aldihan, Robert Gaizauskas and Susan Fitzmaurice	372
<i>Beyond Arabic: Software for Perso-Arabic Script Manipulation</i>	
Alexander Gutkin, Cibu Johny, Raiomond Doctor, Brian Roark and Richard Sproat	381
<i>Coreference Annotation of an Arabic Corpus using a Virtual World Game</i>	
Wateen Abdullah Aliady, Abdulrahman Aloraini, Christopher Madge, Juntao Yu, Richard Bartle and Massimo Poesio	388
<i>NatiQ: An End-to-end Text-to-Speech System for Arabic</i>	
Ahmed Abdelali, Nadir Durrani, Cenk Demiroglu, Fahim Dalvi, Hamdy Mubarak and Kareem Darwish	394
<i>The Effect of Arabic Dialect Familiarity on Data Annotation</i>	
Ibrahim Abu Farha and Walid Magdy	399
<i>Optimizing Naive Bayes for Arabic Dialect Identification</i>	
Tommi Jauhiainen, Heidi Jauhiainen and Krister Lindén	409
<i>iCompass Working Notes for the Nuanced Arabic Dialect Identification Shared task</i>	
Abir Messaoudi, Chayma Fourati, Hatem Haddad and Moez BenHajhmida	415
<i>TF-IDF or Transformers for Arabic Dialect Identification? ITFLOWS participation in the NADI 2022 Shared Task</i>	
Fouad Shammery, Yiyi Chen, Zsolt T Kardkovacs, Mehwish Alam and Haithem Afli	420
<i>Domain-Adapted BERT-based Models for Nuanced Arabic Dialect Identification and Tweet Sentiment Analysis</i>	
Giyaseddin Bayrak and ABDUL MAJEED ISSIFU	425
<i>Benchmarking transfer learning approaches for sentiment analysis of Arabic dialect</i>	
emna fsih, Sameh Kchaou, Rahma Boujelbane and Lamia Hadrich-Belguith	431
<i>SQU-CS @ NADI 2022: Dialectal Arabic Identification using One-vs-One Classification with TF-IDF Weights Computed on Character n-grams</i>	
Abdulrahman Khalifa AAIAbdulsalam	436
<i>Ahmed and Khalil at NADI 2022: Transfer Learning and Addressing Class Imbalance for Arabic Dialect Identification and Sentiment Analysis</i>	
Ahmed Oumar and Khalil Mrini	442
<i>Arabic Sentiment Analysis by Pretrained Ensemble</i>	
Abdelrahim Qaddoumi	447
<i>Dialect &amp; Sentiment Identification in Nuanced Arabic Tweets Using an Ensemble of Prompt-based, Fine-tuned, and Multitask BERT-Based Models</i>	
Reem Abdel-Salam	452
<i>On The Arabic Dialects' Identification: Overcoming Challenges of Geographical Similarities Between Arabic dialects and Imbalanced Datasets</i>	
Salma Jamal, Aly M .Kassem, Omar Mohamed and Ali Ashraf	458

<i>Arabic dialect identification using machine learning and transformer-based models: Submission to the NADI 2022 Shared Task</i>	
Nouf AlShenaifi and Aqil Azmi .....	464
<i>NLP DI at NADI Shared Task Subtask-1: Sub-word Level Convolutional Neural Models and Pre-trained Binary Classifiers for Dialect Identification</i>	
Vani Kanjirangat, Tanja Samardzic, Ljiljana Dolamic and Fabio Rinaldi .....	468
<i>Word Representation Models for Arabic Dialect Identification</i>	
Mahmoud Sobhy, Ahmed H. Abu El-Atta, Ahmed A. El-Sawy and Hamada Nayel .....	474
<i>Building an Ensemble of Transformer Models for Arabic Dialect Classification and Sentiment Analysis</i>	
Abdullah Salem Khered, Ingy Yasser Hassan Abdou Abdelhalim and Riza Batista-Navarro ..	479
<i>Arabic Dialect Identification and Sentiment Classification using Transformer-based Models</i>	
Joseph Attieh and Fadi Hassan .....	485
<i>Generative Approach for Gender-Rewriting Task with ArabicT5</i>	
Sultan Alrowili and Vijay Shanker .....	491
<i>AraProp at WANLP 2022 Shared Task: Leveraging Pre-Trained Language Models for Arabic Propaganda Detection</i>	
Gaurav Singh .....	496
<i>TUB at WANLP22 Shared Task: Using Semantic Similarity for Propaganda Detection in Arabic</i>	
Salar Mohtaj and Sebastian Möller .....	501
<i>SI2M &amp; AIOX Labs at WANLP 2022 Shared Task: Propaganda Detection in Arabic, A Data Augmentation and Name Entity Recognition Approach</i>	
Kamel Gaanoun and Imade Benelallam .....	506
<i>iCompass at WANLP 2022 Shared Task: ARBERT and MARBERT for Multilabel Propaganda Classification of Arabic Tweets</i>	
Bilel - Taboubi, Bechir Brahem and Hatem Haddad .....	511
<i>ChavanKane at WANLP 2022 Shared Task: Large Language Models for Multi-label Propaganda Detection</i>	
Tanmay Chavan and Aditya Manish Kane .....	515
<i>AraBERT Model for Propaganda Detection</i>	
Mohamad Sharara, Wissam Mohamad, Ralph Tawil, Ralph Chobok, Wolf Assi and Antonio Tannoury .....	520
<i>AraBEM at WANLP 2022 Shared Task: Propaganda Detection in Arabic Tweets</i>	
Eshrag Ali Refaee, Basem Ahmed and Motaz Saad .....	524
<i>IITD at WANLP 2022 Shared Task: Multilingual Multi-Granularity Network for Propaganda Detection</i>	
Shubham Mittal and Preslav Nakov .....	529
<i>Pythoneers at WANLP 2022 Shared Task: Monolingual AraBERT for Arabic Propaganda Detection and Span Extraction</i>	
Joseph Attieh and Fadi Hassan .....	534
<i>CNLP-NITS-PP at WANLP 2022 Shared Task: Propaganda Detection in Arabic using Data Augmentation and AraBERT Pre-trained Model</i>	
Sahinur Rahman Laskar, Rahul Singh, Abdullah Faiz Ur Rahman Khilji, Riyanka Manna, Partha Pakray and Sivaji Bandyopadhyay .....	541

*NGU CNLP atWANLP 2022 Shared Task: Propaganda Detection in Arabic*

Ahmed Samir Hussein, Abu Bakr Soliman Mohammad, Mohamed Ibrahim, Laila Hesham Afify  
and Samhaa R. El-Beltagy ..... 545