CL 2019

The Third Workshop on Arabic Corpus Linguistics
WACL-3

Proceedings of the Workshop

July 22, 2019
Cardiff, UK

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-950737-32-1

Preface

Welcome to The Third Workshop on Arabic Corpus Linguistics (WACL-3) held at CL 2019 in Cardiff, UK.

Over recent years, research into the Arabic language using corpora and corpus methods has moved from a small scale activity with isolated pockets of activity to a much larger very active field, with work advancing rapidly on many different fronts in both corpus and computational linguistics. Previously, we organised the 2011 and 2013 Workshops on Arabic Corpus Linguistics WACL-1 and WACL-2, with the second edition being held in conjunction with 2013 Corpus Linguistics Conference at Lancaster University. We now continue working on creating a venue where different activities in corpus research into Arabic can be brought together to explore progress in the field by continuing the WACL series through organising the 3rd Workshop on Arabic Corpus Linguistics (WACL-3) at the CL2019 conference at Cardiff University.
The scope of the workshop encompasses both (a) the design, construction and annotation of Arabic corpora, and (b) the use of corpora in research on the Arabic language – in areas including lexis and lexicography, syntax, collocation, stylistics, and discourse analysis, and Natural Language Processing (NLP) systems and analysis tools. The workshop will pay special attention to non-standard Arabic varieties, Arabic dialects as well as Modern Standard Arabic (MSA).
We received 19 main workshop submissions, out of which 11 were accepted, 6 were rejected, and 2 were withdrawn. All main workshop submissions were reviewed by at least three reviewers with acceptance rate of 65%.

We would like to acknowledge all the hard work of the submitting authors and thank the reviewers for the valuable feedback they provided. We hope these proceedings will serve as a valuable reference for researchers and practitioners in the field of Arabic Corpus Linguistics.

Mahmoud El-Haj, General Chair, on behalf of the organising Committee of the WACL-3 workshop.

Organizers

General chair

Dr Mahmoud El-Haj, Lancaster University, UK

Program Chairs

Prof. Eric Atwell, University of Leeds, Dr UK
Dr Paul Rayson, Lancaster University, UK

Publication Chair

Lama Alsudias, Lancaster University, UK

Publicity Chair

Dr Paul Rayson, Lancaster University, UK
Dr Mahmoud El-Haj, Lancaster University, UK

Invited Keynote Speaker

Dr Alicia González Martínez , Hamburg University, Germany

Program Committee

Ahmed Ali, Qatar Computing Research Institute (QCRI), Qatar
Ahmed Abdelali, Qatar Computing Research Institute (QCRI), Qatar
Almoataz B. Al-Said, Cairo University, Egypt
Eric Atwell, Leeds University, UK
Haithem Afli, Dublin City University, Ireland
Hazem Hajj, American University of Beirut, Lebanon
Ignatius Ezeani, Lancaster University, UK
Imed Zitouni, Microsoft Research, USA
Karim Bouzoubaa, Mohamed Vth University, Morocco
Khaled Shaban, Qatar University, Qatar
Lama Alsudias, Lancaster University, UK
Mahmoud El-Haj, Lancaster University, UK
Mariam Aboelezz, Birkbeck, University of London, UK
Nadi Tomeh, University of Paris 13, France
Nizar Habash, New York University Abu Dhabi, UAE
Nora Al-Twairesh, King Saudi University, Saudi Arabia
Paul Rayson, Lancaster University, UK
Scott Piao, Lancaster University, UK
Taha Zerrouki, Ecole Nationale Supérieure d'Informatique, Algeria
Tamer Elsayed, Qatar University, Qatar
Violetta Cavalli-Sforza, Al Akhawayn University, Morocco
Wajdi Zaghouani, Hamad Bin Khalifa University, Qatar
Wassim El-Hajj, American University of Beirut, Lebanon

Table of Contents

Workshop Program

Monday, July 22, 2019

09:15–9:20 *Opening Remarks*
Dr Mahmoud El-Haj
Lancaster University, UK

09:20–10:05 *Keynote Speaker*
Dr Alicia González Martínez
University of Hamburg, Germany

10:05–12:30 **Session 1** (Chair: Dr Paul Rayson)**:**

10:05–10:30 *Writing Styles of Salwa and Al-Qarni*
Ahmed Omer and Michael Oakes

10:30–10:55 *Classifying Information Sources in Arabic Twitter to Support Online Monitoring of Infectious Diseases*
Lama Alsudias and Paul Rayson

10:55–11:15 ***Coffee Break [VJ gallery and marquee]***

11:15–11:40 *Verbs in Egyptian Arabic: a case for register variation*
Michael Grant White and Deryle W. Lonsdale

11:40–12:05 *The Design of the SauLTC application for the English-Arabic Learner Translation Corpus*
Maha Al-Harthi and Amal Alsaif

12:05–12:30 *Computer Stylometric Comparison of Writings by Qassim Amin and Mohammed Abdu on Women's Rights*
Ahmed Omer and Michael Oakes

12:30–13:30 ***Coffee Break [VJ gallery and marquee]***

13:30–16:20 **Session 2** (Chair: Prof. Eric Atwell)**:**

13:30–13:55 *Text Segmentation Using N-grams to Annotate Hadith Corpus*
Shatha Altammami, Eric Atwell and Ammar Alsalka

13:55–14:20 *Classifying Arabic dialect text in the Social Media Arabic Dialect Corpus (SMADC)*
Areej Alshutayri and Eric Atwell

14:20–14:45 *Compiling and Analysing a Corpus of Transcribed Spoken Gulf Pidgin Arabic Based on Length of Stay in the Gulf*
Najah Albaqawi and Michael Oakes

14:45–15:10 *Crisis Detection from Arabic Tweets*
Alaa Alharbi and Mark Lee

15:10–15:30 ***Coffee Break [VJ gallery and marquee]***

15:30–15:55 *Distance-Based Authorship Verification Across Modern Standard Arabic Genres*
Hossam Ahmed

15:55–16:20 *Can Modern Standard Arabic tools be used for Arabic Dialects? Sentiment Analysis as a Case Study*
Chatrine Qwaider, Stergios Chatzikyriakidis and Simon Dobnik

16:20–16:25 *Closing Remark*
Dr Mahmoud El-Haj
Lancaster University, UK

16:50–17:50 **CL Conference:**
Welcome (large chemistry): Professor Gillian Bristow, Dean of Research for the College of Arts, Humanities and Social Sciences, Cardiff University

17:00–18:0 **CL Conference:**
Welcome (large chemistry): Plenary (Sir Martin Evans Lecture Theatre): Svenja Adolphs - Corpus Linguistics in the Digital Humanities (introduced by Dawn Knight)

18:10–20:00 **CL Conference:**
Drinks reception (marquee) [kindly sponsored by Cambridge University Press]