ACL 2018

Computational Approaches to Linguistic Code-Switching

Proceedings of the Third Workshop

July 19, 2018
Melbourne, Australia

Workshop Sponsor:

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 978-1-948087-45-2

Introduction

Code-switching (CS) is the phenomenon by which multilingual speakers switch back and forth between their common languages in written or spoken communication. CS is pervasive in informal text communications such as news groups, tweets, blogs, and other social media of multilingual communities. Such genres are increasingly being studied as rich sources of social, commercial and political information. Apart from the informal genre challenge associated with such data within a single language processing scenario, the CS phenomenon adds another significant layer of complexity to the processing of the data. Efficiently and robustly processing CS data presents a new frontier for our NLP algorithms on all levels. The goal of this workshop is to bring together researchers interested in exploring these new frontiers, discussing state of the art research in CS, and identifying the next steps in this fascinating research area.

The workshop program includes exciting papers discussing new approaches for CS data and the development of linguistic resources needed to process and study CS. We received a total of 16 regular workshop submissions of which we accepted 11 for publication, five of them as workshop talks and six as posters.

Another component of the workshop is the First Shared Task on Named Entity Recognition (NER) of CS Data. The shared task focused on social media and included two language pairs: Modern Standard Arabic-Dialectal Arabic and English-Spanish. We had a total of 9 participants from which we received 8 submissions on English-Spanish and 6 submissions on Modern Standard Arabic-Dialectal Arabic. We received papers from all these submissions. All shared task systems will be presented during the workshop poster session and two of them will also present a talk. We would like to thank all authors who submitted their contributions to this workshop and all shared task participants for taking on the challenge of NER in code-switched data. We also thank the program committee members for their help in providing meaningful reviews. Lastly, we thank the ACL 2018 organizers for the opportunity to put together this workshop and Amazon for their generous sponsorship.

See you all in Melbourne, Australia at ACL 2018!

Workshop co-chairs,
Gustavo Aguilar
Fahad AlGhamdi
Victor Soto
Thamar Solorio
Mona Diab
Julia Hirschberg

Workshop Co-Chairs:

Gustavo Aguilar, University of Houston
Fahad AlGhamdi, George Washington University
Victor Soto, Columbia University
Thamar Solorio, University of Houston
Mona Diab, George Washington University
Julia Hirschberg, Columbia University

Shared Task Co-Chairs:

Gustavo Aguilar, University of Houston
Fahad AlGhamdi, George Washington University

Publications Chair:

Victor Soto, Columbia University

Program Committee:

Kalika Bali, Microsoft Research India
Elabbas Benmamoun, Duke University
Alan W. Black, Carnegie Mellon University
Agnes Bolonyia, NC State University
Barbara Bullock, University of Texas at Austin
Özlem Çetinoglu, Universität Stuttgart
Monojit Choudhury, Microsoft Research India
Suzanne Dikker, New York University
Björn Gambäck, Norwegian Universities of Science and Technology
Constantine Lignos, University of Southern California Information Sciences Institute
Mitchell P. Marcus, University of Pennsylvania
Cecilia Montes-Alcala, Georgia Institute of Technology
Raymond Mooney, University of Texas at Austin
Borja Navarro Colorado, Universidad de Alicante
Younes Samih, Heinrich Heine - Universität Düsseldorf
Yves Scherrer, University of Helsinki
Chilin Shih, University of Illinois at Urbana-Champaign
David Suendermann, Educational Testing Service
Jacqueline Toribio, University of Texas at Austin
David Vilares, Universidad de Coruña
Emre Yilmaz, CLS/CLST, Radboud University Nijmegen

Invited Speakers:

Pascale Fung, Hong Kong University of Science & Technology
Melinda Fricke, University of Pittsburgh

Table of Contents

Workshop Program

Thursday, July 19, 2018

09:00–10:30 Session 1 Invited Talk and Oral Presentations

9:00–9:05 *Opening Remarks*
Thamar Solorio

9:05–9:50 *Invited Talk: Learning to Codeswitch*
Pascale Fung

9:50–10:10 *Joint Part-of-Speech and Language ID Tagging for Code-Switched Data*
Victor Soto and Julia Hirschberg

10:10–10:30 *Phone Merging For Code-Switched Speech Recognition*
Sunit Sivasankaran, Brij Mohan Lal Srivastava, Sunayana Sitaram, Kalika Bali and Monojit Choudhury

10:30–11:00 Coffee Break

11:00–12:00 Session 2 Oral Presentations

11:00–11:20 *Improving Neural Network Performance by Injecting Background Knowledge: Detecting Code-switching and Borrowing in Algerian texts*
Wafia Adouane, Jean-Philippe Bernardy and Simon Dobnik

11:20–11:40 *Code-Mixed Question Answering Challenge: Crowd-sourcing Data and Techniques*
Khyathi Chandu, Ekaterina Loginova, Vishal Gupta, Josef van Genabith, Günter Neuman, Manoj Chinnakotla, Eric Nyberg and Alan W. Black

11:40–12:00 *Transliteration Better than Translation? Answering Code-mixed Questions over a Knowledge Base*
Vishal Gupta, Manoj Chinnakotla and Manish Shrivastava

Thursday, July 19, 2018 (continued)

12:00–13:30 Lunch Break

13:30–14:15 Session 3 Invited Talk

13:30–14:15 *Invited Talk: Variation in Codeswitched Language: a Psycholinguistic Approach to What, When, and Why*
Melinda Fricke

14:15–15:30 Session 4 Poster Session

Language Identification and Analysis of Code-Switched Social Media Text
Deepthi Mave, Suraj Maharjan and Thamar Solorio

Code-Switching Language Modeling using Syntax-Aware Multi-Task Learning
Genta Indra Winata, Andrea Madotto, Chien-Sheng Wu and Pascale Fung

Predicting the presence of a Matrix Language in code-switching
Barbara Bullock, Wally Guzman, Jacqueline Serigos, Vivek Sharath and Almeida Jacqueline Toribio

Automatic Detection of Code-switching Style from Acoustics
SaiKrishna Rallabandi, Sunayana Sitaram and Alan W. Black

Accommodation of Conversational Code-Choice
Anshul Bawa, Monojit Choudhury and Kalika Bali

Language Informed Modeling of Code-Switched Text
Khyathi Chandu, Thomas Manzini, Sumeet Singh and Alan W. Black

GHHT at CALCS 2018: Named Entity Recognition for Dialectal Arabic Using Neural Networks
Mohammed Attia, Younes Samih and Wolfgang Maier

Simple Features for Strong Performance on Named Entity Recognition in Code-Switched Twitter Data
Devanshu Jain, Maria Kustikova, Mayank Darbari, Rishabh Gupta and Stephen Mayhew

Thursday, July 19, 2018 (continued)

Bilingual Character Representation for Efficiently Addressing Out-of-Vocabulary Words in Code-Switching Named Entity Recognition
Genta Indra Winata, Chien-Sheng Wu, Andrea Madotto and Pascale Fung

Named Entity Recognition on Code-Switched Data Using Conditional Random Fields
Utpal Kumar Sikdar, Biswanath Barik and Björn Gambäck

The University of Texas System Submission for the Code-Switching Workshop Shared Task 2018
Florian Janke, Tongrui Li, Eric Rincón, Gualberto Guzmán, Barbara Bullock and Almeida Jacqueline Toribio

Tackling Code-Switched NER: Participation of CMU
Parvathy Geetha, Khyathi Chandu and Alan W. Black

Multilingual Named Entity Recognition on Spanish-English Code-switched Tweets using Support Vector Machines
Daniel Claeser, Samantha Kent and Dennis Felske

15:30–16:00 Coffee Break

16:00–17:00 Session 5 Shared Task Talks

16:00–16:10 *Named Entity Recognition on Code-Switched Data: Overview of the CALCS 2018 Shared Task*
Gustavo Aguilar, Fahad AlGhamdi, Victor Soto, Mona Diab, Julia Hirschberg and Thamar Solorio

16:10–16:30 *IIT (BHU) Submission for the ACL Shared Task on Named Entity Recognition on Code-switched Data*
Shashwat Trivedi, Harsh Rangwani and Anil Kumar Singh

16:30–16:50 *Code-Switched Named Entity Recognition with Embedding Attention*
Changhan Wang, Kyunghyun Cho and Douwe Kiela

16:50–17:00 *Closing Remarks*
Victor Soto