

Second Workshop on Computational Approaches to Code Switching 2016

Held at the Conference on Empirical Methods in
Natural Language Processing (EMNLP 2016)

Austin, Texas, USA
1 November 2016

ISBN: 978-1-5108-3147-6

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© (2016) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2016)

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

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
Web: www.proceedings.com

Table of Contents

<i>Challenges of Computational Processing of Code-Switching</i> Özlem Çetinoğlu, Sarah Schulz and Ngoc Thang Vu	1
<i>Simple Tools for Exploring Variation in Code-switching for Linguists</i> Gualberto A. Guzman, Jacqueline Serigos, Barbara E. Bullock and Almeida Jacqueline Toribio	12
<i>Word-Level Language Identification and Predicting Codeswitching Points in Swahili-English Language Data</i> Mario Piergallini, Rouzbeh Shirvani, Gauri S. Gautam and Mohamed Chouikha	21
<i>Part-of-speech Tagging of Code-mixed Social Media Content: Pipeline, Stacking and Joint Modelling</i> Utsab Barman, Joachim Wagner and Jennifer Foster	30
<i>Overview for the Second Shared Task on Language Identification in Code-Switched Data</i> Giovanni Molina, Fahad AlGhamdi, Mahmoud Ghoneim, Abdelati Hawwari, Nicolas Rey-Villamizar, Mona Diab and Tamar Solorio	40
<i>Multilingual Code-switching Identification via LSTM Recurrent Neural Networks</i> Younes Samih, Suraj Maharjan, Mohammed Attia, Laura Kallmeyer and Tamar Solorio	50
<i>A Neural Model for Language Identification in Code-Switched Tweets</i> Aaron Jaech, George Mulcaire, Mari Ostendorf and Noah A. Smith	60
<i>SAWT: Sequence Annotation Web Tool</i> Younes Samih, Wolfgang Maier and Laura Kallmeyer	65
<i>Accurate Pinyin-English Codeswitched Language Identification</i> Meng Xuan Xia and Jackie Chi Kit Cheung	71
<i>Unraveling the English-Bengali Code-Mixing Phenomenon</i> Arunavha Chanda, Dipankar Das and Chandan Mazumdar	80
<i>Part-of-speech Tagging of Code-Mixed Social Media Text</i> Souvick Ghosh, Satanu Ghosh and Dipankar Das	90
<i>Part of Speech Tagging for Code Switched Data</i> Fahad AlGhamdi, Giovanni Molina, Mona Diab, Tamar Solorio, Abdelati Hawwari, Victor Soto and Julia Hirschberg	98
<i>The George Washington University System for the Code-Switching Workshop Shared Task 2016</i> Mohamed Al-Badrashiny and Mona Diab	108
<i>Columbia-Jadavpur submission for EMNLP 2016 Code-Switching Workshop Shared Task: System description</i> Arunavha Chanda, Dipankar Das and Chandan Mazumdar	112

<i>The Howard University System Submission for the Shared Task in Language Identification in Spanish-English Codeswitching</i>	
Rouzbeh Shirvani, Mario Piergallini, Gauri Shankar Gautam and Mohamed Chouikha	116
<i>Codeswitching Detection via Lexical Features in Conditional Random Fields</i>	
Prajwol Shrestha	121
<i>Language Identification in Code-Switched Text Using Conditional Random Fields and Babelnet</i>	
Utpal Kumar Sikdar and Björn Gambäck	127
<i>Codeswitching language identification using Subword Information Enriched Word Vectors</i>	
Meng Xuan Xia	132