

# **1st Workshop on Computational Approaches to Code Switching 2014**

**Held at the Conference on Empirical Methods in Natural  
Language Processing (EMNLP 2014)**

**Doha, Qatar  
25 October 2014**

**ISBN: 978-1-63439-481-9**

**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© (2014) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2014)

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: 845-758-2634  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

## Table of Contents

<i>Foreign Words and the Automatic Processing of Arabic Social Media Text Written in Roman Script</i> Ramy Eskander, Mohamed Al-Badrashiny, Nizar Habash and Owen Rambow . . . . .	1
<i>Code Mixing: A Challenge for Language Identification in the Language of Social Media</i> Utsab Barman, Amitava Das, Joachim Wagner and Jennifer Foster . . . . .	13
<i>Detecting Code-Switching in a Multilingual Alpine Heritage Corpus</i> Martin Volk and Simon Clematide . . . . .	24
<i>Exploration of the Impact of Maximum Entropy in Recurrent Neural Network Language Models for Code-Switching Speech</i> Ngoc Thang Vu and Tanja Schultz . . . . .	34
<i>Predicting Code-switching in Multilingual Communication for Immigrant Communities</i> Evangelos Papalexakis, Dong Nguyen and A. Seza Dođruöz . . . . .	42
<i>Twitter Users #CodeSwitch Hashtags! #MoltoImportante #wow</i> David Jurgens, Stefan Dimitrov and Derek Ruths . . . . .	51
<i>Overview for the First Shared Task on Language Identification in Code-Switched Data</i> Thamar Solorio, Elizabeth Blair, Suraj Maharjan, Steven Bethard, Mona Diab, Mahmoud Ghoneim, Abdelati Hawwari, Fahad AlGhamdi, Julia Hirschberg, Alison Chang and Pascale Fung . . . . .	62
<i>Word-level Language Identification using CRF: Code-switching Shared Task Report of MSR India System</i> Gokul Chittaranjan, Yogarshi Vyas, Kalika Bali and Monojit Choudhury . . . . .	73
<i>The CMU Submission for the Shared Task on Language Identification in Code-Switched Data</i> Chu-Cheng Lin, Waleed Ammar, Lori Levin and Chris Dyer . . . . .	80
<i>Language Identification in Code-Switching Scenario</i> Naman Jain and Riyaz Ahmad Bhat . . . . .	87
<i>AIDA: Identifying Code Switching in Informal Arabic Text</i> Heba Elfardy, Mohamed Al-Badrashiny and Mona Diab . . . . .	94
<i>The IUCL+ System: Word-Level Language Identification via Extended Markov Models</i> Levi King, Eric Baucom, Timur Gilmanov, Sandra Kübler, Dan Whyatt, Wolfgang Maier and Paul Rodrigues . . . . .	102
<i>Mixed Language and Code-Switching in the Canadian Hansard</i> Marine Carpuat . . . . .	107
<i>“I am borrowing ya mixing ?” An Analysis of English-Hindi Code Mixing in Facebook</i> Kalika Bali, Jatin Sharma, Monojit Choudhury and Yogarshi Vyas . . . . .	116
<i>DCU-UVT: Word-Level Language Classification with Code-Mixed Data</i> Utsab Barman, Joachim Wagner, Grzegorz Chrupala and Jennifer Foster . . . . .	127
<i>Incremental N-gram Approach for Language Identification in Code-Switched Text</i> Prajwol Shrestha . . . . .	133
<i>The Tel Aviv University System for the Code-Switching Workshop Shared Task</i> Kfir Bar and Nachum Dershowitz . . . . .	139