

# **Joint Workshop on Language Technology for Closely Related Languages, Varieties and Dialects 2015**

International Conference on Recent Advances  
in Natural Language Processing 2015

Hissar, Bulgaria  
10 September 2015

ISBN: 978-1-5108-1784-5

**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© (2015) by the Association for Computational Linguistics  
All rights reserved.

Printed by Curran Associates, Inc. (2015)

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>Overview of the DSL Shared Task 2015</i>	
Marcos Zampieri, Liling Tan, Nikola Ljubešić, Jörg Tiedemann and Preslav Nakov . . . . .	1
<i>***INVITED TALK***: Handling and Mining Linguistic Variation in UGC</i>	
Leon Derczynski . . . . .	10
<i>Distributed Representations of Words and Documents for Discriminating Similar Languages</i>	
Marc Franco-Salvador, Paolo Rosso and Francisco Rangel . . . . .	11
<i>Joint Bayesian Morphology Learning for Dravidian Languages</i>	
Arun Kumar, Lluís Padró and Antoni Oliver . . . . .	17
<i>Use of Transformation-Based Learning in Annotation Pipeline of Igbo, an African Language</i>	
Ikechukwu Onyenwe, Mark Hepple, Chinedu Uchechukwu and Ignatius Ezeani . . . . .	24
<i>***INVITED TALK*** WikiTrans: Swedish-Danish Machine Translation in a Constraint Grammar Framework</i>	
Eckhard Bick . . . . .	34
<i>Language Identification using Classifier Ensembles</i>	
Shervin Malmasi and Mark Dras . . . . .	35
<i>Discriminating Similar Languages with Token-Based Backoff</i>	
Tommi Jauhiainen, Heidi Jauhiainen and Krister Lindén . . . . .	44
<i>NLEL UPV Autoritas Participation at Discrimination between Similar Languages (DSL) 2015 Shared Task</i>	
Raül Fabra Boluda, Francisco Rangel and Paolo Rosso . . . . .	52
<i>Discriminating between Similar Languages Using PPM</i>	
Victoria Bobicev . . . . .	59
<i>Comparing Approaches to the Identification of Similar Languages</i>	
Marcos Zampieri, Binyam Gebrekidan Gebre, Hernani Costa and Josef van Genabith . . . . .	66
<i>A Two-level Classifier for Discriminating Similar Languages</i>	
Judit Ács, László Grad-Gyenge and Thiago Bruno Rodrigues de Rezende Oliveira . . . . .	73
<i>Experiments in Discriminating Similar Languages</i>	
Cyril Goutte and Serge Leger . . . . .	78
<i>Building Monolingual Word Alignment Corpus for the Greater China Region</i>	
fan xu, Xiongfei Xu, Mingwen Wang and Maoxi Li . . . . .	85