

# **Australasian Language Technology Association Workshop**

**(ALTA 2013)**

**Brisbane, Australia  
4 – 6 December 2013**

**Editors:**

**S. Karimi  
K. Verspoor**

**ISBN: 978-1-62993-557-7**

**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© (2013) 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)

## Contents

<b>Invited talks</b>	<b>1</b>
<i>Robust Computational Semantics</i> <b>Mark Steedman</b>	<b>2</b>
<i>Concurrent Discourse Relations</i> <b>Bonnie Webber</b>	<b>3</b>
<b>Full papers</b>	<b>4</b>
<i>Classifying English Documents by National Dialect</i> <b>Marco Lui and Paul Cook</b>	<b>5</b>
<i>Crowd-Sourcing of Human Judgments of Machine Translation Fluency</i> <b>Yvette Graham, Timothy Baldwin, Alistair Moffat and Justin Zobel</b>	<b>16</b>
<i>The Effect of the Within-speaker Sample Size on the Performance of Likelihood Ratio Based Forensic Voice Comparison: Monte Carlo Simulations</i> <b>Shunichi Ishihara</b>	<b>25</b>
<i>Noise in Speech-to-Text Voice: Analysis of Errors and Feasibility of Phonetic Similarity for Their Correction</i> <b>Hanna Suominen and Gabriela Ferraro</b>	<b>34</b>
<i>Examining the Impact of Coreference Resolution on Quote Attribution</i> <b>Tim O’Keefe, Kellie Webster, James R. Curran and Irena Koprinska</b>	<b>43</b>
<i>Multi-Objective Optimization for Clustering of Medical Publications</i> <b>Asif Ekbal, Sriparna Saha, Diego Molla and K Ravikumar</b>	<b>53</b>
<i>A Study: From Electronic Laboratory Notebooks to Generated Queries for Literature Recommendation</i> <b>Oldooz Dianat, Cecile Paris and Stephen Wan</b>	<b>62</b>
<i>A Comparative Study of Likelihood Ratio Based Forensic Text Comparison in Procedures: Multivariate Kernel Density vs. Gaussian Mixture Model-Universal Background Model</i> <b>Shunichi Ishihara</b>	<b>71</b>
<i>Finding Fires with Twitter</i> <b>Robert Power, Bella Robinson and David Ratcliffe</b>	<b>80</b>

<b>Short papers</b>	<b>90</b>
<i>Impact of Corpus Diversity and Complexity on NER Performance</i> <b>Tatyana Shmanina, Ingrid Zukerman, Antonio Jimeno Yepes, Lawrence Cave-</b> <b>don and Karin Verspoor</b>	<b>91</b>
<i>Cumulative Progress in Language Models for Information Retrieval</i> <b>Antti Puurula</b>	<b>96</b>
<i>Error Detection in Automatic Speech Recognition</i> <b>Farshid Zavareh, Ingrid Zukerman, Su Nam Kim and Thomas Kleinbauer</b>	<b>101</b>
<i>Working with Defaults in a Controlled Natural Language</i> <b>Rolf Schwitter</b>	<b>106</b>
<b>Poster papers</b>	<b>111</b>
<i>Rhythm, Metrics, and the Link to Phonology</i> <b>Jason Brown and Sam Mandal</b>	<b>112</b>
<i>Differences in Speaker Individualising Information between Case Particles and Fillers in Spoken Japanese</i> <b>Shunichi Ishihara</b>	<b>118</b>
<i>Automatic Climate Classification of Environmental Science Literature</i> <b>Jared Willett, David Martinez, J. Angus Webb and Timothy Baldwin</b>	<b>123</b>
<b>ALTA Shared Task papers</b>	<b>131</b>
<i>Overview of the 2013 ALTA Shared Task</i> <b>Diego Molla</b>	<b>132</b>
<i>Recovering Casing and Punctuation using Conditional Random Fields</i> <b>Marco Lui and Li Wang</b>	<b>137</b>
<i>e-Learning with Kaggle in Class: Adapting the ALTA Shared Task 2013 to a Class Project</i> <b>Karin Verspoor and Jeremy Nicholson</b>	<b>142</b>