

# **1st Celtic Language Technology Workshop**

**(CLTW 2014)**

**Held at COLING 2014**

**Dublin, Ireland  
23 August 2014**

**ISBN: 978-1-63439-216-7**

**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© (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>Developing an Automatic Part-of-Speech Tagger for Scottish Gaelic</i> William Lamb and Samuel Danso .....	1
<i>Using Irish NLP resources in Primary School Education</i> Monica Ward .....	6
<i>Tools facilitating better use of online dictionaries: Technical aspects of Multidict, Wordlink and Clilstore</i> Caoimhin O Donnaile .....	18
<i>Processing Mutations in Breton with Finite-State Transducers</i> Thierry Poibeau .....	28
<i>Statistical models for text normalization and machine translation</i> Kevin Scannell .....	33
<i>Cross-lingual Transfer Parsing for Low-Resourced Languages: An Irish Case Study</i> Teresa Lynn, Jennifer Foster, Mark Dras and Lamia Tounsi .....	41
<i>Irish National Morphology Database: a high-accuracy open-source dataset of Irish words</i> Michal Boleslav Měchura .....	50
<i>Developing further speech recognition resources for Welsh</i> Sarah Cooper, Dewi Jones and Delyth Prys .....	55
<i>gdbank: The beginnings of a corpus of dependency structures and type-logical grammar in Scottish Gaelic</i> Colin Batchelor .....	60
<i>Developing high-end reusable tools and resources for Irish-language terminology, lexicography, onomastics (toponymy), folkloristics, and more, using modern web and database technologies</i> Brian Ó Raghallaigh and Michal Boleslav Měchura .....	66
<i>DECHE and the Welsh National Corpus Portal</i> Delyth Prys, Dewi Jones and Mared Roberts .....	71
<i>Subsegmental language detection in Celtic language text</i> Akshay Minocha and Francis Tyers .....	76