

# **Celtic Language Technology Workshop 2019**

Held at Machine Translation Summit XVII

Dublin, Ireland  
19 August 2019

ISBN: 978-1-5108-9247-7

**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© (2019) by the Association for Computational Linguistics  
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-ND. (<https://creativecommons.org/licenses/by-nd/4.0/>)

Printed by Curran Associates, Inc. (2019)

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: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Contents

|   |           |
|---|-----------|
| <b>Unsupervised multi-word term recognition in Welsh</b>  | <b>1</b>  |
| <i>Irena Spasić, David Owen, Dawn Knight, Andreas Artemiou</i>  |           |
| <b>Universal dependencies for Scottish Gaelic: syntax</b>   | <b>7</b>  |
| <i>Colin Batchelor</i>  |           |
| <b>Speech technology and Argentinean Welsh</b>  | <b>16</b> |
| <i>Elise Bell</i>   |           |
| <b>Development of a Universal Dependencies treebank for Welsh</b>   | <b>21</b> |
| <i>Johannes Heinecke, Francis M. Tyers</i>  |           |
| <b>Code-switching in Irish tweets: A preliminary analysis</b>   | <b>32</b> |
| <i>Teresa Lynn, Kevin Scannell</i>  |           |
| <b>Embedding English to Welsh MT in a Private Company</b>   | <b>41</b> |
| <i>Myfyr Prys, Dewi Bryn Jones</i>  |           |
| <b>Adapting Term Recognition to an Under-Resourced Language: the Case of Irish</b>  | <b>48</b> |
| <i>John P. McCrae, Adrian Doyle</i>   |           |
| <b>Leveraging backtranslation to improve machine translation for Gaelic languages</b>   | <b>58</b> |
| <i>Meghan Dowling, Teresa Lynn, Andy Way</i>  |           |
| <b>Improving full-text search results on <i>dúchas.ie</i> using language technology</b>                                       | <b>63</b> |
| <i>Brian Ó Raghallaigh, Kevin Scannell, Meghan Dowling</i>  |           |
| <b>A Character-Level LSTM Network Model for Tokenizing the Old Irish text of the Würzburg Glosses on the Pauline Epistles</b> | <b>70</b> |
| <i>Adrian Doyle, John P. McCrae, Clodagh Downey</i>   |           |
| <b>A Green Approach for an Irish App (Refactor, reuse and keeping it real)</b>  | <b>80</b> |
| <i>Monica Ward, Maxim Mozgovoy, Marina Purgina</i>  |           |