2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages 2017 (ComputEL-2)

Honolulu, Hawaii, USA
6 – 7 March 2017

ISBN: 978-1-5108-4213-7

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© (2017) by the Association for Computational Linguistics
All rights reserved.

Printed by Curran Associates, Inc. (2017)

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

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
Web: www.proceedings.com

Table of Contents

A Morphological Parser for Odawa
Dustin Bowers, Antti Arppe, Jordan Lachler, Sjur Moshagen and Trond Trosterud 1

Creating lexical resources for polysynthetic languages—the case of Arapaho
Ghazaleh Kazeminejad, Andrew Cowell and Mans Hulden 10

From Small to Big Data: paper manuscripts to RDF triples of Australian Indigenous Vocabularies
Nick Thieberger and Conal Tuohy 19

Issues in digital text representation, on-line dissemination, sharing and re-use for African minority languages
Emmanuel Ngué Um 24

Developing collection management tools to create more robust and reliable linguistic data
Gary Holton, Kavon Hooshiar and Nick Thieberger 33

STREAMLInED Challenges: Aligning Research Interests with Shared Tasks
Gina-Anne Levow, Emily M. Bender, Patrick Littell, Kristen Howell, Shobhana Chelliah, Joshua Crowgey, Dan Garrette, Jeff Good, Sharon Hargus, David Inman, Michael Maxwell, Michael Tjalve and Fei Xia 39

Work With What You've Got
Lucy Bell and Lawrence Bell 48

Converting a comprehensive lexical database into a computational model: The case of East Cree verb inflection
Antti Arppe, Marie-Odile Junker and Delasie Torkornoo 52

Instant annotations in ELAN corpora of spoken and written Komi, an endangered language of the Barents Sea region
Ciprian Gerstenberger, Niko Partanen and Michael Rießler 57

Inferring Case Systems from IGT: Enriching the Enrichment
Kristen Howell, Emily M. Bender, Michel Lockwood, Fei Xia and Olga Zamaraeva 67

Case Studies in the Automatic Characterization of Grammars from Small Wordlists
Jordan Kodner, Spencer Kaplan, Hongzhi Xu, Mitchell P. Marcus and Charles Yang 76

Endangered Data for Endangered Languages: Digitizing Print dictionaries
Michael Maxwell and Aric Bills 85

A computationally-assisted procedure for discovering poetic organization within oral tradition
David Meyer 92

Improving Coverage of an Inuktitut Morphological Analyzer Using a Segmental Recurrent Neural Network
Jeffrey Micher 101

Click reduction in fluent speech: a semi-automated analysis of Mangetti Dune !Xung
Amanda Miller and Micha Elsner 107

DECCA Repurposed: Detecting transcription inconsistencies without an orthographic standard
C. Anton Rytting and Julie Yelle 116

Jejueo talking dictionary: A collaborative online database for language revitalization
Moira Saltzman 122

Computational Support for Finding Word Classes: A Case Study of Abui
Olga Zamaraeva, František Kratochvíl, Emily M. Bender, Fei Xia and Kristen Howell 130

Waldayu and Waldayu Mobile: Modern digital dictionary interfaces for endangered languages
Patrick Littell, Aidan Pine and Henry Davis 141

Connecting Documentation and Revitalization: A New Approach to Language Apps
Alexa N. Little 151

Developing a Suite of Mobile Applications for Collaborative Language Documentation
Mat Bettinson and Steven Bird 156

Cross-language forced alignment to assist community-based linguistics for low resource languages
Timothy Kempton 165

A case study on using speech-to-translation alignments for language documentation
Antonios Anastasopoulos and David Chiang 170