

# **Sixth Workshop on the Use of Computational Methods in the Study of Endangered Languages (computel 2023)**

Online  
5-6 March 2023

ISBN: 978-1-7138-9029-4

**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© (2023) by the Association for Computational Linguistics  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2024)

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)

## Table of Contents

<i>Leveraging supplementary text data to kick-start automatic speech recognition system development with limited transcriptions</i>	
Nay San, Martijn Bartelds, Blaine Billings, Ella de Falco, Hendi Feriza, Johan Safri, Wawan Sahrozi, Ben Foley, Bradley McDonnell and Dan Jurafsky	1
<i>Applications of classification trees for endangered language description Finite verb morphology in Kolyma Yukaghir</i>	
Albert Ventayol-Boada	7
<i>Using LARA to rescue a legacy Pitjantjatjara course</i>	
Manny Rayner and Sasha Wilmoth	13
<i>User-Centric Evaluation of OCR Systems for Kwak’wala</i>	
Shruti Rijhwani, Daisy Rosenblum, Michayla King, Antonios Anastasopoulos and Graham Neubig	19
<i>Towards a finite-state morphological analyser for San Mateo Huave</i>	
Francis M. Tyers and Samuel Herrera Castro	30
<i>A Survey of Computational Infrastructure to Help Preserve and Revitalize Bodwéwadmimwen</i>	
Robert Lewis	38
<i>Morphological Data Generation from FLEx</i>	
Shengyu Liao, Beth Bryson and Sarah Moeller	45
<i>A text-to-speech synthesis system for Border Lakes Ojibwe</i>	
Christopher Hammerly, Sonja Fougère, Giancarlo Sierra, Scott Parkhill, Harrison Porteous and Chad Quinn	54
<i>From Raw Data to Acoustic Analysis A Roadmap for Acquaviva Collecroce</i>	
Simon Gonzalez	60
<i>Studying the impact of language model size for low-resource ASR</i>	
Zoey Liu, Justin Spence and Emily Prud’Hommeaux	71
<i>FileLingR An R Script validation tool for depositors and users of digital language collections</i>	
Irene Yi and Claire Bowerman	78
<i>Challenges and Issue of Gender Bias in Under-Represented Languages An Empirical Study on Inuktitut-English NMT</i>	
Ngoc Tan Le, Oussama Hansal and Fatiha Sadat	83
<i>Text normalization for low-resource languages the case of Ligurian</i>	
Stefano Lusito, Edoardo Ferrante and Jean Maillard	92
<i>LSDT a Dependency Treebank of Lombard Sinti</i>	
Marco Forlano and Luca Brigada Villa	98
<i>A morphological analyzer for Huasteca Nahuatl</i>	
Ana Tona, Guillaume Thomas and Ewan Dunbar	106
<i>Speech-to-text recognition for multilingual spoken data in language documentation</i>	
Lorena Martín Rodríguez and Christopher Cox	111
<i>Towards Universal Dependencies in Cook Islands Māori</i>	
Sarah Karnes, Rolando Coto-Solano and Sally Akevai Nicholas	118
<i>Investigating Speaker Diarization of Endangered Language Data</i>	
Gina-Anne Levow	124