

computel 2026

**Ninth Workshop on the Use of Computational Methods in the  
Study of Endangered Languages**

**Proceedings of the Workshop**

July 4, 2026

The computer organizers gratefully acknowledge the support from the following sponsors.

**Gold**



**Silver**



©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)  
317 Sidney Baker St. S  
Suite 400 - 134  
Kerrville, TX 78028  
USA  
Tel: +1-855-225-1962  
[acl@aclweb.org](mailto:acl@aclweb.org)

ISBN 979-8-89176-422-4

## Introduction

These proceedings contain the papers presented at the 9th Workshop on the Use of Computational Methods in the Study of Endangered Languages (ComputEL-9), held on July 4, 2026, in San Diego, California, USA. The workshop is co-located with the 64th Annual Meeting of the Association for Computational Linguistics (ACL 2026) and offers hybrid attendance options, allowing participants to attend in person or remotely.

As the name suggests, this is the ninth workshop dedicated to the intersection of computational tools and endangered-language research. The inaugural event took place at the Association for Computational Linguistics (ACL) main conference in Baltimore, Maryland in 2014. Subsequent workshops have been co-located with the International Conference on Language Documentation & Conservation at the University of Hawai‘i at Mānoa (2017, 2019, 2021, 2023, 2025) or ACL-related venues (2022 in Dublin, Ireland; 2024 in St. Julians, Malta). We are delighted to continue this tradition by co-locating with ACL, marking the third time the workshop has been co-located with ACL. This time, we are co-ordinating our activities with Americas-NLP, held on the previous day.

The primary aim of ComputEL-9 is to bring together computational researchers, documentary linguists, and community language practitioners. By uniting these diverse groups, the workshop fosters a collaborative environment for exchanging ideas, methods, and resources that support the documentation and revitalization of endangered languages. The organizers are gratified by the variety of contributions, which reflect the importance of collaborative efforts across disciplines and communities.

This year, we received 33 submissions in the form of short or long papers. Following a thorough review process, 19 were accepted.

We extend our appreciation to all authors for their submissions and to the Program Committee for the thoughtful review of each proposal. We also thank the 64th ACL organizers for their assistance in hosting this workshop. We hope that ComputEL-9 sparks discussions and partnerships that continue to enrich the field of endangered language research, ultimately contributing to more robust support for language communities worldwide.

# Organizing Committee

## Proceedings Editor

Godfred Agyapong, University of Florida

## Program Chair

Sarah Moeller, University of Florida

Ali Marashian, University of Colorado Boulder

Daisy Rosenblum, University of British Columbia

Antti Arppe, University of Alberta

## Program Committee

### Chairs

Sarah Moeller, University of Florida  
Godfred Agyapong, University of Florida  
Antti Arppe, University of Alberta  
Ali Marashian, University of Colorado Boulder

### Program Committee

Steven Abney, University of Michigan  
Antonios Anastasopoulos, George Mason University  
Pt Anderson, Revitalization Technology  
Martin Benjamin, Kamusi Project International  
Francisca Benyarku, University of Florida  
Blaine Billings, University of Hawaii at Manoa  
Steven Bird, Charles Darwin University  
Christopher Cox, Carleton University  
Abteen Ebrahimi, University of Colorado, Boulder  
Mengzhe Geng, National Research Council Canada  
Luke Gessler, Indiana University Bloomington  
Michael Ginn, University of Colorado  
Jeff Good, University at Buffalo  
Christopher Hammerly, University of British Columbia  
Atticus Harrigan, Independent Scholar  
Ryan Henke, University of Wisconsin–Madison  
Gary Holton, University of Hawaii  
David H u g g i n s - D a i n e s, Independent Researcher  
Anna Kazantseva, National Research Council Canada  
Frantisek Kratochvil, Palacky University Olomouc  
Roland Kuhn, National Research Council of Canada  
Ngoc Tan Le, Universite du Quebec a Montreal  
Éric Le Ferrand, Boston College  
G i n a - A n n e Levow, University of Washington  
Zoey Liu, Department of Linguistics, University of Florida  
Olga Lovick, University of Saskatchewan  
Amogh Mannekote, University of Florida  
Bradley Mcdonnell, University of Hawai'i at Mānoa  
Alexis Michaud, CNRS - LACITO  
Steven Moran, University of Neuchâtel  
Saliha Muradoglu, The Australian National University  
Remo Nitschke, University of Arizona  
Chibuzor Okocha, University of Florida  
Alexis Palmer, University of Colorado Boulder  
Aidan Pine, National Research Council Canada  
Flammie Pirinen, UiT Norgga Árkatalaš universitehta  
Emily Prud'hommeaux, Boston College  
Karthick Narayanan Ramakrishnan, Krea University  
Daisy Rosenblum, UBC

Elizabeth Salesky, Johns Hopkins University  
Emmanuel Schang, Université d'Orléans  
Yves Scherrer, University of Oslo  
Lane Schwartz, University of Alaska Fairbanks  
Mandana Seyfeddinipur, Endangered Languages Archive  
Bhargav Shandilya, University of Colorado Boulder  
Gary Simons, SIL Global  
Divyansh Singh, University of Florida  
Sonal Sinha, Indian Institute of Technology, Johdpur, India.  
Richard Sproat, Sakana.ai  
Meiraba Takhellambam, Manipur University  
Nick Thieberger, University of Melbourne  
Paul Trilsbeek, Max Planck Institute for Psycholinguistics  
Daan Van Esch, Google Research  
Linda Wiechetek, UiT Norgga árktálaš universitehta  
Kweku Yamoah, University of Florida  
Borui Zhang, University of Florida

## Table of Contents

<i>Morphological Parsing for Media Lengua: When Accessibility Matters More Than State-of-the-Art</i> Jesse Stewart and Olga Kriukova.....	1
<i>Speech Recognition and Synthesis Technologies Applied to Preservation and Revitalization of the Ainu Language</i> Tatsuya Kawahara and Kohei Matsuura.....	10
<i>Choosing an ASR model for Dënë Sùtné: Navigating polysynthesis and unstandardized orthography</i> Olga Kriukova, Antti Arppe and Olga Lovick.....	15
<i>An Interactive System for Generating Revisable Grammar Lessons for Extremely Low-Resource Languages Without Expert Annotation</i> Sebastien Christian.....	26
<i>Voices from the Margins: Modeling Linguistic Diversity in Spontaneous Speech for Low-Resource Languages</i> Vithal Bhandari, Tiya Kumar and Katharine Mulhern.....	37
<i>Digital posters: Publishing Gurindji plant and animal poster content as websites using an open-source template-based RO-Crate preview tool</i> Ben Foley, Abigail Davis and Felicity Meakins.....	55
<i>AvarLab: An Integrated Digital Ecosystem for Avar, a Morphologically Rich Low-Resource Language</i> Kebed Zagidov and Thomas Brochhagen.....	62
<i>Revitalising Endangered Languages and Cultural Heritage through Language Technology: A Pilot Study for Dzardzongke</i> Hannah Claus, Songbo Hu, Emre Isik, Anna Korhonen, Kitty Liu and Marieke Meelen.....	72
<i>Annotation Tools for Language Documentation: A Survey of Capabilities, Gaps, and Morphological Support</i> Changbing Yang, Pt Anderson, Godfred Agyapong and Sarah Moeller.....	80
<i>Addressing Domain Mismatch in ASR for Akuzipik Language Documentation</i> Summer Chambers, Sylvia Woodrose Schwartz, Matthew Kelley and Lane Woodrose Schwartz.....	93
<i>Low-Resource Methods for Hawaiian Machine Translation</i> Nolan Brophy and Winston Wu.....	104
<i>Child Support: Leveraging Lexifiers Resources to Support Creoles ASR</i> Éric Le Ferrand and Fabiola Henri.....	111
<i>Indigenous Writing Systems Matter: Rethinking NLP beyond Alphabetic Bias through Script-Aware Modeling</i> Ngoc Tan Le, Mamady Traore, Cristian Ahumada Oliva and Fatiha Sadat.....	118
<i>CoRSAL-OCR: Evaluating Zero-Shot OCR for Language Archive Materials</i> Luke Gessler and Andrew Haynes.....	125
<i>The Missing Middle: Language Documentation Needs Better Infrastructure, Not Better Models</i> Luke Gessler, Antonios Anastasopoulos, Sandra Auderset, Timotheus Bodt, Shobhana Chelliah, Sebastien Christian, Maxime Fily, Santiago Herrera, Eva Huber, Sharid Loaiciga, Marieke Meelen, Robert Östling, Alexis Palmer and Eline Visser.....	136

<i>Aspects of Selecting the Right ASR Training Languages for Under-Resourced Languages</i>	
J. Elizabeth Liebl, Summer Chambers, Matthew Kelley and Géraldine Walther .....	148
<i>Bottlenecks of In-Context Learning for Fieldwork ASR: A Case-study of Panāra</i>	
Siyu Liang, Myriam Lapierre and G i n a - A n n e Levow .....	157
<i>Developing A Hawaiian Corpus Toolkit for Data-Driven Language Learning</i>	
Joseph Winkie, Michol Miller and Winston Wu .....	167
<i>Voice Activation Detection for Transcription of Indigenous Languages</i>	
Rolando C o t o - S o l a n o, Mikaela Browning, Thomas Corrado and Sally Akevai Nicholas	177

# Program

**Saturday, July 4, 2026**

09:00 - 09:30     *Opening Remarks*

09:30 - 10:30     *Session 1*

*CoRSAL-OCR: Evaluating Zero-Shot OCR for Language Archive Materials*

Luke Gessler and Andrew Haynes

*Developing A Hawaiian Corpus Toolkit for Data-Driven Language Learning*

Joseph Winkie, Michol Miller and Winston Wu

10:30 - 11:00     *Break*

11:00 - 12:00     *Session 2*

*AvarLab: An Integrated Digital Ecosystem for Avar, a Morphologically Rich Low-Resource Language*

Kebed Zagidov and Thomas Brochhagen

*Morphological Parsing for Media Lengua: When Accessibility Matters More Than State-of-the-Art*

Jesse Stewart and Olga Kriukova

12:00 - 12:30     *Lunch*

12:30 - 13:30     *Poster Session*

*Low-Resource Methods for Hawaiian Machine Translation*

Nolan Brophy and Winston Wu

*Child Support: Leveraging Lexifiers Resources to Support Creoles ASR*

Éric Le Ferrand and Fabiola Henri

*Voice Activation Detection for Transcription of Indigenous Languages*

Rolando C o t o - S o l a n o, Mikaela Browning, Thomas Corrado and Sally Akevai Nicholas

**Saturday, July 4, 2026 (continued)**

*Aspects of Selecting the Right ASR Training Languages for Under-Resourced Languages*

J. Elizabeth Liebl, Summer Chambers, Matthew Kelley and Géraldine Walther

*Revitalising Endangered Languages and Cultural Heritage through Language Technology: A Pilot Study for Dzardzongke*

Hannah Claus, Songbo Hu, Emre Isik, Anna Korhonen, Kitty Liu and Marieke Meelen

*Digital posters: Publishing Gurindji plant and animal poster content as websites using an open-source template-based RO-Crate preview tool*

Ben Foley, Abigail Davis and Felicity Meakins

*Voices from the Margins: Modeling Linguistic Diversity in Spontaneous Speech for Low-Resource Languages*

Vitthal Bhandari, Tiya Kumar and Katharine Mulhern

*Indigenous Writing Systems Matter: Rethinking NLP beyond Alphabetic Bias through Script-Aware Modeling*

Ngoc Tan Le, Mamady Traore, Cristian Ahumada Oliva and Fatiha Sadat

13:30 - 15:30 *Session 3*

*Choosing an ASR model for Dënë Sųłné: Navigating polysynthesis and unstandardized orthography*

Olga Kriukova, Antti Arppe and Olga Lovick

*Bottlenecks of In-Context Learning for Fieldwork ASR: A Case-study of Panāra*

Siyu Liang, Myriam Lapierre and G i n a - A n n e Levow

*Addressing Domain Mismatch in ASR for Akuzipik Language Documentation*

Summer Chambers, Sylvia Woodrose Schwartz, Matthew Kelley and Lane Woodrose Schwartz

*Speech Recognition and Synthesis Technologies Applied to Preservation and Revitalization of the Ainu Language*

Tatsuya Kawahara and Kohei Matsuura

15:30 - 16:00 *Break*

16:00 - 17:30 *Session 4*

**Saturday, July 4, 2026 (continued)**

*An Interactive System for Generating Revisable Grammar Lessons for Extremely Low-Resource Languages Without Expert Annotation*

Sebastien Christian

*The Missing Middle: Language Documentation Needs Better Infrastructure, Not Better Models*

Luke Gessler, Antonios Anastasopoulos, Sandra Auderset, Timotheus Bodt, Shobhana Chelliah, Sebastien Christian, Maxime Fily, Santiago Herrera, Eva Huber, Sharid Loaiciga, Marieke Meelen, Robert Östling, Alexis Palmer and Eline Visser

*Annotation Tools for Language Documentation: A Survey of Capabilities, Gaps, and Morphological Support*

Changbing Yang, Pt Anderson, Godfred Agyapong and Sarah Moeller

17:30 - 18:00     *Closing Remarks*