NAACL 2025

The Fifth Workshop on NLP for Indigenous Languages of the Americas

Proceedings of the Workshop

May 4, 2025

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

Sponsor

©2025 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

ISBN 979-8-89176-236-7

Introduction

We would like to welcome you to AmericasNLP 2025, the Fifth Workshop on Natural Language Processing for Indigenous Languages of the Americas!

The main goals of the workshop are to:

- encourage research on NLP, computational linguistics, corpus linguistics, and speech around the globe to work on Indigenous American languages.
- promote research on both neural and non-neural machine learning approaches suitable for low-resource languages.
- connect researchers and professionals from underrepresented communities and native speakers of endangered languages with the machine learning and NLP communities.

In 2025, AmericasNLP will be held in Albuquerque, USA, on May 4th. Prior to the workshop three shared tasks were hosted: (1) the Shared Task on Machine Translation into Indigenous Languages, (2) the Shared Task on the Creation of Educational Materials for Indigenous Languages, and new for 2025 (3) the Shared Task on Machine Translation Metrics for Indigenous Languages. During the workshop, there will be two invited talks, a panel, poster session, as well as multiple paper and shared task submission presentations.

We received a total of 22 submissions: 12 research papers, 1 extended abstract, and 8 shared task system description papers (across all shared tasks). 8 archival papers were accepted (acceptance rate: 66%) – in addition to the previously published and system description papers.

We would like to acknowledge all the time and effort put into the reviewing process, and thank for program committee members for helping us create a high-quality program in a short amount of time. AmericasNLP would not have been possible without the help our sponsor for 2025: Google. Finally, we would also thank all the authors who submitted their work to the workshop, the participants of the shared tasks, and everyone who will be at the workshop, both in-person and remote, to exchange and discuss their ideas for improving natural language technologies for Indigenous languages of the Americas!

Manuel Mager, Abteen Ebrahimi, Shruti Rijhwani, Robert Pugh, Arturo Oncevay, Luis Chiruzzo, Rolando Coto-Solano, and Katharina von der Wense

AmericasNLP 2025 Organizing Committee

Program Committee

Program Committee

Eduardo Blanco, University of Arizona
Ruixiang Cui, University of Copenhagen
Cristina España-Bonet, DFKI GmbH
Silvia Fernandez Sabido, CentroGeo
Luke Gessler, Indiana University Bloomington
Santiago Góngora, Universidad de la República
Éric Le Ferrand, Boston College
Zoey Liu, Department of Linguistics, University of Florida
Daniela Moctezuma, Centrogeo
Sarah Moeller, University of Florida
Alejandro Molina-Villegas, SECIHTI
Remo Nitschke, University of Arizona
John E. Ortega, Northeastern University
Tanmay Parekh, University of California Los Angeles
Nathaniel Robinson, Johns Hopkins University
Atnafu Lambebo Tonja, Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC)
Daan Van Esch, Google Research
Si Wu, Northeastern University

Keynote Talk

Alexis Palmer
University of Colorado Boulder

Bio: Dr. Palmer is an expert in computational discourse and semantics; computational linguistics for low-resource languages and language documentation; discourse structure and coherence, and modes of discourse and social analytics, including automated detection of offensive language in social media. She received her PhD from UT Austin in 2009, has held a number of prestigious post docs and research positions in Germany (including positions at the Institute for Computational Linguistics in Heidelberg and the Institut für Deutsche Sprache in Mannheim). Until her move to CU in 2021, she was an assistant professor at the University of North Texas, Denton. Dr. Palmer brings a prestigious National Science Foundation CAREER grant with her to CU. In this project, she is working on cross-linguistic methods for better development of language processing tools for low-resource languages. The project is called FOLTA (From One Language to Another). She has also recently become interested in the question of how we can make the outcomes of linguistic documentation more useable and accessible, particularly to support development of pedagogical materials for a language.

Table of Contents

Text-to-speech system for low-resource languages: A case study in Shipibo-Konibo (a Panoan language from Peru)
Daniel Menendez and Hector Gomez 1

Does a code-switching dialogue system help users learn conversational fluency in Choctaw?
Jacqueline Brixey and David Traum 8

A hybrid Approach to low-resource machine translation for Ojibwe verbs
Minh Nguyen, Christopher Hammerly and Miikka Slifverberg 18

Advancing Uto-Aztecan Language Technologies: A Case Study on the Endangered Comanche Language
Jesus Alvarez C, Daua Karajeanes, Ashley Prado, John Ruttan, Ivory Yang, Sean O'brien, Vasu Sharma and Kevin Zhu 27

Py-Elotl: A Python NLP package for the languages of Mexico
Ximena Gutierrez-Vasques, Robert Pugh, Victor Mijangos, Diego Barriga Martínez, Paul Aguilar, Mikel Segura, Paola Innes, Javier Santillan, Cynthia Montaño and Francis Tyers 38

Analyzing and generating English phrases with finite-state methods to match and translate inflected Plains Cree word-forms
Antti Arppe 48

Unsupervised, Semi-Supervised and LLM-Based Morphological Segmentation for Bribri
Carter Anderson, Mien Nguyen and Rolando Coto-Solano 63

FUSE : A Ridge and Random Forest-Based Metric for Evaluating MT in Indigenous Languages
Rahul Raja and Arpita Vats 77

UCSP Submission to the AmericasNLP 2025 Shared Task
Jorge Asillo Congora, Julio Santisteban and Ricardo Lazo Vasquez 84

Machine Translation Using Grammar Materials for LLM Post-Correction
Jonathan Hus, Antonios Anastasopoulos and Nathaniel Krasner 92

Machine Translation Metrics for Indigenous Languages Using Fine-tuned Semantic Embeddings
Nathaniel Krasner, Justin Vasselli, Belu Ticona, Antonios Anastasopoulos and Chi-Kiu Lo . . 100

JHU's Submission to the AmericasNLP 2025 Shared Task on the Creation of Educational Materials for Indigenous Languages
Tom Lupicki, Lavanya Shankar, Kaavya Chaparala and David Yarowsky 105

Leveraging Dictionaries and Grammar Rules for the Creation of Educational Materials for Indigenous Languages
Justin Vasselli, Haruki Sakajo, Arturo Martínez Peguero, Frederikus Hudi and Taro Watanabe 112

Harnessing NLP for Indigenous Language Education: Fine-Tuning Large Language Models for Sentence Transformation
Mahshar Yahan and Dr. Mohammad Islam 119

Leveraging Large Language Models for Spanish-Indigenous Language Machine Translation at AmericasNLP 2025
Mahshar Yahan and Dr. Mohammad Islam 126

Findings of the AmericasNLP 2025 Shared Tasks on Machine Translation, Creation of Educational Material, and Translation Metrics for Indigenous Languages of the Americas

Ona De Gibert, Robert Pugh, Ali Marashian, Raul Vazquez, Abteen Ebrahimi, Pavel Denisov, Enora Rice, Edward Gow-Smith, Juan Prieto, Melissa Robles, Rubén Manrique, Oscar Moreno, Angel Lino, Rolando Coto-Solano, Aldo Alvarez, Marvin Agüero-Torales, John E. Ortega, Luis Chiruzzo, Arturo Oncevay, Shruti Rijhwani, Katharina Von Der Wense and Manuel Mager 134