

AmericasNLP 2026

**Sixth Workshop on NLP for Indigenous Languages of the
Americas**

Proceedings of the Workshop

July 3, 2026

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

Sponsor



©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

ISBN 979-8-89176-415-6

Introduction

We would like to welcome you to AmericasNLP 2026, the Sixth 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 2026, AmericasNLP will be held in San Diego, USA, on July 4th. Prior to the workshop, we introduced a new shared task: The AmericasNLP 2026 Shared Task on Cultural Image Captioning for Indigenous Languages, the first shared task dedicated to generating captions for images depicting Indigenous cultures of the Americas, written in the Indigenous languages themselves. To support this, we introduced and publicly released a newly constructed dataset spanning five cultures and their dominant languages: Bribri, Guaraní, Yucatec Maya, Orizaba Nahuatl, and Wixárika. Eight teams participated, submitting 27 systems in total.

We received a total of 39 submissions: 28 research papers, 3 non-archival works, and 8 shared task system description papers (across all shared tasks). 18 archival papers were accepted (acceptance rate: 64%) – 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 the 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 of our sponsor Google. Finally, we would also like to thank all the authors who submitted their work to the workshop, the participants of the shared task, and everyone who will be at the workshop to exchange and discuss their ideas for improving natural language technologies for Indigenous languages of the Americas!

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

AmericasNLP 2026 Organizing Committee

Program Committee

Program Committee

Heriberto Avelino, National Institute for Anthropology and History
Eduardo Blanco, University of Arizona
Ona De Gibert, University of Helsinki
Cristina España-Bonet, BSC/DFKI GmbH
Silvia Fernandez Sabido, CentroGeo
Andrew Fisher, University of New Brunswick
Luke Gessler, Indiana University Bloomington
Santiago Góngora, Universidad de la República
Lewis Howe, University of Georgia
Elwin Huaman, University of Cambridge
Simran Khanuja, Carnegie Mellon University
Éric Le Ferrand, Boston College
Zoey Liu, Department of Linguistics, University of Florida
Ali Marashian, University of Colorado Boulder
Daniela Moctezuma, Centrogeo
Remo Nitschke, University of Arizona
Angeles Belem Priego Sanchez, Universidad Autónoma Metropolitana
Nathaniel Robinson, Johns Hopkins University
Hossain Shaikh Saadi, Johannes Gutenberg University Mainz
Mario Sanz-Guerrero, Johannes Gutenberg University Mainz
Daan Van Esch, Google Research
Raul Vazquez, University of Helsinki

Keynote Talk

Jacqueline Brixey

University of Wisconsin–Madison



Bio: Jacqueline [Lina] Brixey is a postdoctoral researcher at the University of Wisconsin-Madison. She completed her PhD in Computer Science at the University of Southern California in December 2024. Her research focuses on Indigenous and endangered languages, dialogue systems, and bilingualism.

Table of Contents

| | |
|--|-----|
| <i>Neural Text-to-Speech for Myaamia: Speech Synthesis for an Indigenous Algonquian Language</i> Anita Baral, John Femiani, Hunter Lockwood, Daniela Inclezan and Balam Bhandari | 1 |
| <i>Evaluating Frontier LLM Translation Capability for Lakota</i> Lance Robertson | 11 |
| <i>Bridging Digital Tools for Linguistic Documentation and Revitalization</i> Christopher Haberland, Carly Crowther, Jingnong Qu and Anuk Centellas | 22 |
| <i>A Systematic Comparison of Parameter-Efficient Fine-Tuning Techniques for Low-Resource Neural Machine Translation: Evidence from Indigenous Languages of the Americas</i> Drew Stackhouse and Justin Debenedetto | 33 |
| <i>Linguistic Feature Tagging for Automatic Classification of 27 Closely-Related Quechua Varieties</i> Claire Post and Alexis Palmer | 46 |
| <i>What Resources Matter for Interlinear Glossing? Using LLMs and RAG for the Low-Resource Mapudungun Language</i> Anaís Almendra, Arianna Bisazza, Claudio Gutierrez and Felipe Hasler | 64 |
| <i>Deer, Deities, and Dancing: Culturally Biased LLM Hallucination in Low-Resource Wixárika Translation</i> Henry Gagnier and Ashwin Kirubakaran | 74 |
| <i>IndigiEval: Evaluating LLMs in North American Indigenous Languages</i> Julia Mainzinger and Jacqueline Brixey | 82 |
| <i>A data-centric approach to performance improvement in under-resourced ASR: The case of Dëně Sųtné</i> Olga Kriukova, Olga Lovick and Antti Arppe | 95 |
| <i>Towards a Community-accessible Cahuilla corpus: Developing HTR for J.P. Harrington’s handwritten fieldnotes on Mountain Cahuilla</i> Ray Huaute and Jacqueline Brixey | 107 |
| <i>Corpora duplication for NLP in low-resource languages: A case study of Nahuatl</i> Juan Jose Guzman Landa, Juan-Manuel Torres-Moreno, Luis Moreno Jimenez, Elvys Linhares Pontes, Miguel Figueroa-Saavedra, Graham Ranger and Martha Lorena Avendaño Garrido | 115 |
| <i>On the Robustness of Morphosyntactic Transformation with Large Language Models: The Case of Quechua Collao</i> Pool Pocco and Arturo Oncevay | 128 |
| <i>Building Community-Centred NLP Resources for Puno Quechua</i> Elwin Huaman, Adrian Gamarra Lafuente, Johanna Cordova and Anna Korhonen | 147 |
| <i>The Power of Simplicity: N-Grams and Transformers in Nahuatl Language Identification</i> Luis Mercado Campos, Robert Pugh and Alexis Palmer | 153 |
| <i>RAN: Resource Abundance Notation for Languages in NLP</i> Jared Coleman, Tainã Coleman and Bhaskar Krishnmachari | 168 |
| <i>Bringing Mapudungun into the Modern MT Ecosystem: Morphology-Aware Tokenization for NLLB-200 Fine-Tuning</i> Isaac Thompson, Brandon Rogers and Eric Ringger | 173 |

| | |
|--|-----|
| <i>QomL’aqtaqa: A Qom–Spanish Parallel Corpus for Natural Language Processing with Machine Translation Evaluation</i> | |
| Viviana Cotik, Aleksei Korablev, Paola Cúneo and Pablo Laciana | 186 |
| <i>Toward a Coarse-Labeled Spoken Language Identification Dataset for Central Alaskan Yup’ik and Samoan from US Broadcast Archives</i> | |
| Yangyang Chen, Kyeongmin Rim and James Pustejovsky | 203 |
| <i>Retrieval-Augmented Long-Context Translation for Cultural Image Captioning: Gators submission for AmericasNLP 2026 shared task</i> | |
| Aashish Dhawan, Christopher Driggers-Ellis, Dzmitry Kasinets, Christan Grant and Zhe Wang | 212 |
| <i>From Machine Translation to Image Captioning: Training Vision-Language Models for Indigenous Languages of the Americas</i> | |
| Luis Lara and Param Raval | 224 |
| <i>Culturally-Aware Image Captioning for Guaraní with Multimodal Prompting: IUHoosiers at AmericasNLP 2026</i> | |
| Wenchen Shi, Phakphum Artkaew and Luke Gessler | 236 |
| <i>6fanle Submission to the AmericasNLP 2026 Shared Task on Wixarika Image Captioning</i> | |
| Ji Wang and Hanqi Yang | 243 |
| <i>Culturally Grounded Image Captioning in Indigenous Languages with Vision-Language Models: Cascaded and Single-Stage Approaches</i> | |
| Mirelle Bueno and Sushil Garg | 248 |
| <i>Schema-Constrained Image Captioning for Five Low-Resource Indigenous Languages</i> | |
| Diego Cuadros, Nicholas Leeds, Amanda Avalos, Azul Alpizar-Velazquez, Jared Coleman, Faezeh Dehghan Tarzjani and Bhaskar Krishnamachari | 257 |
| <i>USP at AmericasNLP 2026 Shared Task: Culturally-Aware Image Captioning for Indigenous Languages via Vision-Language Models and Fine-Tuned Neural Machine Translation</i> | |
| Rafael Fernandes | 264 |
| <i>Nearest-Neighbor Retrieval for Indigenous Image Captioning</i> | |
| Justin Vasselli, Arturo Martínez Peguero, Shintaro Ozaki, Frederikus Hudi, Haruki Sakajo and Taro Watanabe | 272 |
| <i>Findings of the AmericasNLP 2026 Shared Task on Cultural Image Captioning for Indigenous Languages</i> | |
| Minh Duc Bui, David Guzmán, Abteen Ebrahimi, Franklin Morales, Marvin Agüero-Torales, Raquel Insfrán, Cecilia González, Ramón Araujo, Luca Cernuzzi, Carlos Raul Noh Chi, Carlos Eduardo Tec Cahun, Sindi Estrella Poot Cohuo, Daniel Ricardo Benítez Chi, Santos Natanael Palomo Arévalo, Jessica Elizabeth Canul Canche, Deysi Aracely Poot Poot, Wendy Marleny Dzib Dzib, Eduardo José Ake Pool, Reynaldo Alexander Couoh Martin, Silvia Fernandez Sabido, Luis Samuel Santiago Melchor, Sotero Silverio, Robert Pugh, Raúl Vázquez, John E. Ortega, Arturo Oncevay, Rubén Manrique, Luis Chiruzzo, Rolando Coto-Solano, Elisabeth Mager, Shruti Rijhwani, David Ifeoluwa Adelani, Manuel Mager and Katharina von der Wense | 279 |