

# **Third Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2023)**

Toronto, Canada  
14 July 2023

ISBN: 978-1-7138-8182-7

**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© (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>Use of NLP in the Context of Belief states of Ethnic Minorities in Latin America</i> Olga Kellert and Mahmud Zaman .....	1
<i>Neural Machine Translation through Active Learning on low-resource languages: The case of Spanish to Mapudungun</i> Begoa Pendas, Andres Carvallo and Carlos Aspillaga .....	6
<i>Understanding Native Language Identification for Brazilian Indigenous Languages</i> Paulo Cavalin, Pedro Domingues, Julio Nogima and Claudio Pinhanez .....	12
<i>Codex to corpus: Exploring annotation and processing for an open and extensible machine-readable edition of the Florentine Codex</i> Francis Tyers, Robert Pugh and Valery Berthoud F. ....	19
<i>Developing finite-state language technology for Maya</i> Robert Pugh, Francis Tyers and Quetzil Castaeda .....	30
<i>Modelling the Reduplicating Lushootseed Morphology with an FST and LSTM</i> Jack Rueter, Mika Hmlinen and Khalid Alnajjar .....	40
<i>Fine-tuning Sentence-RoBERTa to Construct Word Embeddings for Low-resource Languages from Bilingual Dictionaries</i> Diego Bear and Paul Cook .....	47
<i>Identification of Dialect for Eastern and Southwestern Ojibwe Words Using a Small Corpus</i> Kalvin Hartwig, Evan Lucas and Timothy Havens .....	58
<i>Enriching WayunaikiSpanish Neural Machine Translation with Linguistic Information</i> Nora Graichen, Josef Van Genabith and Cristina Espaa-bonet .....	67
<i>Towards the First Named Entity Recognition of Inuktitut for an Improved Machine Translation</i> Ngoc Tan Le, Soumia Kasdi and Fatiha Sadat .....	84
<i>Parallel Corpus for Indigenous Language Translation: Spanish-Mazatec and Spanish-Mixtec</i> Atnafu Lambebo Tonja, Christian Maldonado-sifuentes, David Alejandro Mendoza Castillo, Olga Kolesnikova, No Castro-snchez, Grigori Sidorov and Alexander Gelbukh .....	94
<i>A finite-state morphological analyser for Highland Puebla Nahuatl</i> Robert Pugh and Francis Tyers .....	103
<i>Neural Machine Translation for the Indigenous Languages of the Americas: An Introduction</i> Manuel Mager, Rajat Bhatnagar, Graham Neubig, Ngoc Thang Vu and Katharina Kann .....	109
<i>Community consultation and the development of an online Akuzipik-English dictionary</i> Benjamin Hunt, Lane Schwartz, Sylvia Schreiner and Emily Chen .....	134
<i>Finding words that aren't there: Using word embeddings to improve dictionary search for low-resource languages</i> Antti Arppe, Andrew Neitsch, Daniel Dacanay, Jolene Poulin, Daniel Hieber and Atticus Harrigan	144
<i>Enhancing Spanish-Quechua Machine Translation with Pre-Trained Models and Diverse Data Sources: LCT-EHU at AmericasNLP Shared Task</i> Nouman Ahmed, Natalia Flechas Manrique and Antonije Petrovi .....	156

<i>ChatGPT is not a good indigenous translator</i> David Stap and Ali Araabi .....	163
<i>Few-shot Spanish-Aymara Machine Translation Using English-Aymara Lexicon</i> Liling Tan .....	168
<i>PlayGround Low Resource Machine Translation System for the 2023 AmericasNLP Shared Task</i> Tianrui Gu, Kaie Chen, Siqi Ouyang and Lei Li .....	173
<i>Four Approaches to Low-Resource Multilingual NMT: The Helsinki Submission to the AmericasNLP 2023 Shared Task</i> Ona De Gibert, Ral Vzquez, Mikko Aulamo, Yves Scherrer, Sami Virpioja and Jrg Tiedemann	177
<i>Sheffield’s Submission to the AmericasNLP Shared Task on Machine Translation into Indigenous Languages</i> Edward Gow-smith and Danae Snchez Villegas .....	192
<i>Enhancing Translation for Indigenous Languages: Experiments with Multilingual Models</i> Atnafu Lambebo Tonja, Hellina Hailu Nigatu, Olga Kolesnikova, Grigori Sidorov, Alexander Gelbukh and Jugal Kalita .....	200
<i>Findings of the AmericasNLP 2023 Shared Task on Machine Translation into Indigenous Languages</i> Abteen Ebrahimi, Manuel Mager, Shruti Rijhwani, Enora Rice, Arturo Oncevay, Claudia Baltazar, María Cortés, Cynthia Montaña, John E. Ortega, Rolando Coto-solano, Hilaria Cruz, Alexis Palmer and Katharina Kann .....	206