

First Workshop on Natural Language Processing for Indigenous Languages of the Americas (AmericasNLP 2021)

Online
11 June 2021

ISBN: 978-1-7138-3015-3

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© (2021) by the Association for Computational Linguistics
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

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

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
Web: www.proceedings.com

Table of Contents

<i>qxoRef 1.0: A coreference corpus and mention-pair baseline for coreference resolution in Conchucos Quechua</i>	
Elizabeth Pankratz	1
<i>A corpus of K'iche' annotated for morphosyntactic structure</i>	
Francis Tyers and Robert Henderson	10
<i>Investigating variation in written forms of Nahuatl using character-based language models</i>	
Robert Pugh and Francis Tyers	21
<i>Apurinã Universal Dependencies Treebank</i>	
Jack Rueter, Marília Fernanda Pereira de Freitas, Sidney Da Silva Facundes, Mika Hämäläinen and Niko Partanen	28
<i>Automatic Interlinear Glossing for Otomi language</i>	
Diego Barriga Martínez, Victor Mijangos and Ximena Gutierrez-Vasques	34
<i>A survey of part-of-speech tagging approaches applied to K'iche'</i>	
Francis Tyers and Nick Howell	44
<i>Highland Puebla Nahuatl Speech Translation Corpus for Endangered Language Documentation</i>	
Jiatong Shi, Jonathan D. Amith, Xuankai Chang, Siddharth Dalmia, Brian Yan and Shinji Watanabe	53
<i>End-to-End Automatic Speech Recognition: Its Impact on the Workflow in Documenting Yoloxóchitl Mixtec</i>	
Jonathan D. Amith, Jiatong Shi and Rey Castillo García	64
<i>A finite-state morphological analyser for Paraguayan Guaraní</i>	
Anastasia Kuznetsova and Francis Tyers	81
<i>Morphological Segmentation for Seneca</i>	
Zoey Liu, Robert Jimerson and Emily Prud'hommeaux	90
<i>Representation of Yine [Arawak] Morphology by Finite State Transducer Formalism</i>	
Adriano Ingunza Torres, John Miller, Arturo Oncevay and Roberto Zariquiey Biondi	102
<i>Leveraging English Word Embeddings for Semi-Automatic Semantic Classification in Nêhiyawêwin (Plains Cree)</i>	
Atticus Harrigan and Antti Arppe	113
<i>Restoring the Sister: Reconstructing a Lexicon from Sister Languages using Neural Machine Translation</i>	
Remo Nitschke	122
<i>Expanding Universal Dependencies for Polysynthetic Languages: A Case of St. Lawrence Island Yupik</i>	
Hyunji Park, Lane Schwartz and Francis Tyers	131
<i>The More Detail, the Better? – Investigating the Effects of Semantic Ontology Specificity on Vector Semantic Classification with a Plains Cree / nêhiyawêwin Dictionary</i>	
Daniel Dacanay, Atticus Harrigan, Arok Wolvengrey and Antti Arppe	143

<i>Experiments on a Guarani Corpus of News and Social Media</i> Santiago Góngora, Nicolás Giossa and Luis Chiruzzo	153
<i>Towards a First Automatic Unsupervised Morphological Segmentation for Inuinnaqtun</i> Ngoc Tan Le and Fatiha Sadat	159
<i>Toward Creation of Ancash Lexical Resources from OCR</i> Johanna CORDOVA and Damien Nouvel	163
<i>Ayuuk-Spanish Neural Machine Translator</i> Delfino Zacarías Márquez and Ivan Vladimir Meza Ruiz	168
<i>Explicit Tone Transcription Improves ASR Performance in Extremely Low-Resource Languages: A Case Study in Bribri</i> Rolando Coto-Solano	173
<i>Towards a morphological transducer and orthography converter for Western Tlacolula Valley Zapotec</i> Jonathan Washington, Felipe Lopez and Brook Lillehaugen	185
<i>Peru is Multilingual, Its Machine Translation Should Be Too?</i> Arturo Oncevay	194
<i>Findings of the AmericasNLP 2021 Shared Task on Open Machine Translation for Indigenous Languages of the Americas</i> Manuel Mager, Arturo Oncevay, Abteen Ebrahimi, John Ortega, Annette Rios, Angela Fan, Ximena Gutierrez-Vasques, Luis Chiruzzo, Gustavo Giménez-Lugo, Ricardo Ramos, Ivan Vladimir Meza Ruiz, Rolando Coto-Solano, Alexis Palmer, Elisabeth Mager-Hois, Vishrav Chaudhary, Graham Neubig, Ngoc Thang Vu and Katharina Kann	202
<i>Open Machine Translation for Low Resource South American Languages (AmericasNLP 2021 Shared Task Contribution)</i> Shantipriya Parida, Subhadarshi Panda, Amulya Dash, Esau Villatoro-Tello, A. Seza Doğruöz, Rosa M. Ortega-Mendoza, Amadeo Hernández, Yashvardhan Sharma and Petr Motlicek	218
<i>NRC-CNRC Machine Translation Systems for the 2021 AmericasNLP Shared Task</i> Rebecca Knowles, Darlene Stewart, Samuel Larkin and Patrick Littell	224
<i>Low-Resource Machine Translation Using Cross-Lingual Language Model Pretraining</i> Francis Zheng, Machel Reid, Edison Marrese-Taylor and Yutaka Matsuo	234
<i>The REPU CS' Spanish-Quechua Submission to the AmericasNLP 2021 Shared Task on Open Machine Translation</i> Oscar Moreno	241
<i>Moses and the Character-Based Random Babbling Baseline: CoAStal at AmericasNLP 2021 Shared Task</i> Marcel Bollmann, Rahul Aralikkatte, Héctor Murrieta Bello, Daniel Hershovich, Miryam de Lhoneux and Anders Søgaard	248
<i>The Helsinki submission to the AmericasNLP shared task</i> Raúl Vázquez, Yves Scherrer, Sami Virpioja and Jörg Tiedemann	255
<i>IndT5: A Text-to-Text Transformer for 10 Indigenous Languages</i> El Moatez Billah Nagoudi, Wei-Rui Chen, Muhammad Abdul-Mageed and Hasan Cavusoglu ..	265