

# **Workshop on Resources and Technologies for Indigenous, Endangered and Lesser-resourced Languages in Eurasia (EURALI 2022)**

Held at LREC 2022

Marseille, France  
20 - 25 June 2022

## **Editors:**

**Atul Kr. Ojha  
Sina Ahmadi**

**Chao-Hong Liu  
John P. McCrae**

ISBN: 978-1-7138-6119-5

**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© (2022) by the Association for Computational Linguistics  
All rights reserved.

Copyright for individual papers remains with the authors and are licensed under a Creative Commons 4.0 license, CC-BY-ND. (<https://creativecommons.org/licenses/by-nd/4.0/>)

Printed with permission by Curran Associates, Inc. (2022)

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>NLP Pipeline for Annotating (Endangered) Tibetan and Newar Varieties</i><br>Christian Faggionato, Nathan Hill and Marieke Meelen .....  | 1  |
| <i>Towards an Ontology for Toponyms in Nepalese Historical Documents</i><br>Sabine Tittel .....  | 7  |
| <i>Semiautomatic Speech Alignment for Under-Resourced Languages</i><br>Juho Leinonen, Niko Partanen, Sami Virpioja and Mikko Kurimo .....  | 17 |
| <i>How to Digitize Completely: Interactive Geovizualization of a Sketch Map from the Kuzmina Archive</i><br>Elena Lazarenko and Aleksandr Riaposov .....   | 22 |
| <i>Word Class Based Language Modeling: A Case of Upper Sorbian</i><br>Isidor Maier, Johannes Kuhn, Frank Duckhorn, Ivan Kraljevski, Daniel Sobe, Matthias Wolff and Constanze Tschöpe .....                  | 28 |
| <i>Bringing Together Version Control and Quality Assurance of Language Data with LAMA</i><br>Aleksandr Riaposov, Elena Lazarenko and Timm Lehmborg .....   | 36 |
| <i>Automatic Verb Classifier for Abui (AVC-abz)</i><br>Frantisek Kratochvil, George Saad, Jiří Vomlel and Václav Kratochvíl .....  | 42 |
| <i>Dialogue Act and Slot Recognition in Italian Complex Dialogues</i><br>Irene Sucameli, Michele De Quattro, Arash Eshghi, Alessandro Suglia and Maria Simi .....  | 51 |
| <i>Digital Resources for the Shughni Language</i><br>Yury Makarov, Maksim Melenchenko and Dmitry Novokshanov .....   | 61 |
| <i>German Dialect Identification and Mapping for Preservation and Recovery</i><br>Aynalem Tesfaye Misganaw and Sabine Roller .....   | 65 |
| <i>Exploring Transfer Learning for Urdu Speech Synthesis</i><br>Sahar Jamal, Sadaf Abdul Rauf and Quratulain Majid .....   | 70 |
| <i>Towards Bengali WordNet Enrichment using Knowledge Graph Completion Techniques</i><br>Sree Bhattacharyya and Abhik Jana .....   | 75 |
| <i>Enriching Hindi WordNet Using Knowledge Graph Completion Approach</i><br>Sushil Awale and Abhik Jana .....  | 81 |
| <i>A Digital Swedish-Yiddish/Yiddish-Swedish Dictionary: A Web-Based Dictionary that is also Available Offline</i><br>Magnus Ahltop, Jean Hessel, Gunnar Eriksson, Maria Skeppstedt and Rickard Domeij ..... | 86 |
| <i>An Online Dictionary for Dialects of North Frisian</i><br>Michael Wehar and Tanno Hüttenrauch .....   | 88 |
| <i>Towards a Unified Tool for the Management of Data and Technologies in Field Linguistics and Computational Linguistics - LiFE</i><br>Siddharth Singh, Ritesh Kumar, Shyam Ratan and Sonal Sinha .....      | 90 |
| <i>Universal Dependencies Treebank for Tatar: Incorporating Intra-Word Code-Switching Information</i><br>Chihiro Taguchi, Sei Iwata and Taro Watanabe .....  | 95 |

*Preparing an endangered language for the digital age: The Case of Judeo-Spanish*

Alp Öktem, Rodolfo Zevallos, Yasmin Moslem, Özgür Güneş Öztürk and Karen Gerson Şarhon 105