

# **Second Workshop on NLP Applications to Field Linguistics (Field Matters 2023)**

Dubrovnik, Croatia  
6 May 2023

ISBN: 978-1-7138-9050-8

**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>Automated speech recognition of Indonesian-English language lessons on YouTube using transfer learning</i> Zara Maxwell-smith and Ben Foley .....	1
<i>Application of Speech Processes for the Documentation of Kréyòl Gwadeloupéyen</i> Éric Le Ferrand, Fabiola Henri, Benjamin Lecouteux and Emmanuel Schang .....	17
<i>Unsupervised part-of-speech induction for language description: Modeling documentation materials in Kolyma Yukaghir</i> Albert Ventayol-boada, Nathan Roll and Simon Todd .....	23
<i>Speech Database (Speech-DB) – An on-line platform for storing, validating, searching, and recording spoken language data</i> Jolene Poulin, Daniel Dacanay and Antti Arppe .....	30
<i>ASR pipeline for low-resourced languages: A case study on Pomak</i> Chara Tsoukala, Kosmas Kritsis, Ioannis Douros, Athanasios Katsamanis, Nikolaos Kokkas, Vasileios Arampatzakis, Vasileios Sevetlidis, Stella Markantonatou and George Pavlidis .....	40
<i>Improving Low-resource RRG Parsing with Structured Gloss Embeddings</i> Roland Eibers, Kilian Evang and Laura Kallmeyer .....	46
<i>Approaches to Corpus Creation for Low-Resource Language Technology: the Case of Southern Kurdish and Laki</i> Sina Ahmadi, Zahra Azin, Sara Belelli and Antonios Anastasopoulos .....	52
<i>AraDiaWER: An Explainable Metric For Dialectical Arabic ASR</i> Abdulwahab Sahyoun and Shady Shehata .....	64
<i>A Quest for Paradigm Coverage: The Story of Nen</i> Saliha Muradoglu, Hanna Suominen and Nicholas Evans .....	74
<i>Multilingual Automatic Extraction of Linguistic Data from Grammars</i> Albert Kornilov .....	86