

9th Workshop on Slavic Natural Language Processing (SlavicNLP 2023)

Dubrovnik, Croatia
6 May 2023

ISBN: 978-1-7138-7379-2

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. (2023)

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>Named Entity Recognition for Low-Resource Languages - Profiting from Language Families</i> Sunna Torge, Andrei Politov, Christoph Lehmann, Bochra Saffar and Ziyang Tao	1
<i>MAUPQA: Massive Automatically-created Polish Question Answering Dataset</i> Piotr Rybak	11
<i>TreBERT: A pre-trained encoder for Polish Twitter</i> Wojciech Szmyd, Alicja Kotyla, Michał Zobniów, Piotr Falkiewicz, Jakub Bartczuk and Artur Zygałdo	17
<i>Croatian Film Review Dataset (Cro-FiReDa): A Sentiment Annotated Dataset of Film Reviews</i> Gaurish Thakkar, Nives Mikelic Preradovic and Marko Tadić	25
<i>Too Many Cooks Spoil the Model: Are Bilingual Models for Slovene Better than a Large Multilingual Model?</i> Pranaydeep Singh, Aaron Maladry and Els Lefever	32
<i>Machine-translated texts from English to Polish show a potential for typological explanations in Source Language Identification</i> Damiaan Reijnaers and Elize Herrewijnen	40
<i>Comparing domain-specific and domain-general BERT variants for inferred real-world knowledge through rare grammatical features in Serbian</i> Sofia Lee and Jelke Bloem	47
<i>Dispersing the clouds of doubt: can cosine similarity of word embeddings help identify relation-level metaphors in Slovene?</i> Mojca Brglez	61
<i>Automatic text simplification of Russian texts using control tokens</i> Anna Dmitrieva	70
<i>Target Two Birds With One SToNe: Entity-Level Sentiment and Tone Analysis in Croatian News Headlines</i> Ana Barić, Laura Majer, David Dukić, Marijana Grbeša-zenzerović and Jan Snajder	78
<i>Is German secretly a Slavic language? What BERT probing can tell us about language groups</i> Aleksandra Mysiak and Jacek Cyranka	86
<i>Resources and Few-shot Learners for In-context Learning in Slavic Languages</i> Michal Štefánik, Marek Kadlčík, Piotr Gramacki and Petr Sojka	94
<i>Analysis of Transfer Learning for Named Entity Recognition in South-Slavic Languages</i> Nikola Ivačić, Thi Hong Hanh Tran, Boshko Koloski, Senja Pollak and Matthew Purver	106
<i>Information Extraction from Polish Radiology Reports Using Language Models</i> Aleksander Obuchowski, Barbara Klauedel and Patryk Jasik	113
<i>Can BERT eat RuCoLA? Topological Data Analysis to Explain</i> Irina Proskurina, Ekaterina Artemova and Irina Piontkovskaya	123
<i>WikiGoldSK: Annotated Dataset, Baselines and Few-Shot Learning Experiments for Slovak Named Entity Recognition</i> David Suba, Marek Suppa, Jozef Kubik, Endre Hamerlik and Martin Takac	138

<i>Measuring Gender Bias in West Slavic Language Models</i> Sandra Martinková, Karolina Stanczak and Isabelle Augenstein	146
<i>On Experiments of Detecting Persuasion Techniques in Polish and Russian Online News: Preliminary Study</i> Nikolaos Nikolaidis, Nicolas Stefanovitch and Jakub Piskorski	155
<i>Exploring the Use of Foundation Models for Named Entity Recognition and Lemmatization Tasks in Slavic Languages</i> Gabriela Pařka and Artur Nowakowski	165
<i>Large Language Models for Multilingual Slavic Named Entity Linking</i> Rinalds Vĩksna, Inguna Skadiņa, Daiga Deksnē and Roberts Rozis	172
<i>Slav-NER: the 4th Cross-lingual Challenge on Recognition, Normalization, Classification, and Linking of Named Entities across Slavic languages</i> Roman Yangarber, Jakub Piskorski, Anna Dmitrieva, Michał Marcińczuk, Pavel Přebáň, Piotr Rybak and Josef Steinberger	179