

19th Workshop on Multiword Expressions (MWE 2023)

Dubrovnik, Croatia
6 May 2023

ISBN: 978-1-7138-7386-0

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>Token-level Identification of Multiword Expressions using Pre-trained Multilingual Language Models</i> Raghuraman Swaminathan and Paul Cook	1
<i>Romanian Multiword Expression Detection Using Multilingual Adversarial Training and Lateral Inhibition</i> Andrei Avram, Verginica Barbu Mititelu and Dumitru-Clementin Cercel	7
<i>Predicting Compositionality of Verbal Multiword Expressions in Persian</i> Mahtab Sarlak, Yalda Yarandi and Mehrnoush Shamsfard	14
<i>PARSEME corpus release 1.3</i> Agata Savary, Cherifa Ben Khelil, Carlos Ramisch, Voula Giouli, Verginica Barbu Mititelu, Najet Hadj Mohamed, Cvetana Krstev, Chaya Liebeskind, Hongzhi Xu, Sara Stymne, Tunga Güngör, Thomas Pickard, Bruno Guillaume, Eduard Bejček, Archana Bhatia, Marie Candito, Polona Gantar, Uxoia Iñurrieta, Albert Gatt, Jolanta Kovalevskaite, Timm Lichte, Nikola Ljubešić, Johanna Monti, Carla Parra Escartín, Mehrnoush Shamsfard, Ivelina Stoyanova, Veronika Vincze and Abigail Walsh	24
<i>Investigating the Effects of MWE Identification in Structural Topic Modelling</i> Dimitrios Kokkinakis, Ricardo Sánchez, Sebastianus Bruinsma and Mia-Marie Hammarlin	36
<i>Idioms, Probing and Dangerous Things: Towards Structural Probing for Idiomaticity in Vector Space</i> Filip Klubička, Vasudevan Nedumpozhimana and John Kelleher	45
<i>Graph-based multi-layer querying in Parseme Corpora</i> Bruno Guillaume	58
<i>Enriching Multiword Terms in Wiktionary with Pronunciation Information</i> Lenka Bajcetic, Thierry Declerck and Gilles Sérasset	65
<i>Detecting Idiomatic Multiword Expressions in Clinical Terminology using Definition-Based Representation Learning</i> François Remy, Alfiya Khabibullina and Thomas Demeester	73
<i>Automatic Generation of Vocabulary Lists with Multiword Expressions</i> John Lee and Adilet Uvaliyev	81
<i>Are Frequent Phrases Directly Retrieved like Idioms? An Investigation with Self-Paced Reading and Language Models</i> Giulia Rambelli, Emmanuele Chersoni, Marco S. G. Senaldi, Philippe Blache and Alessandro Lenci	87
<i>Annotation of lexical bundles with discourse functions in a Spanish academic corpus</i> Eleonora Guzzi, Margarita Alonso-Ramos, Marcos Garcia and Marcos García Salido	99
<i>A Survey of MWE Identification Experiments: The Devil is in the Details</i> Carlos Ramisch, Abigail Walsh, Thomas Blanchard and Shiva Taslimipoor	106
<i>A MWE lexicon formalism optimised for observational adequacy</i> Adam Lion-Bouton, Agata Savary and Jean-Yves Antoine	121