MWE 2021

The 17th Workshop on Multiword Expressions

Proceedings of the Workshop

August 6, 2021
Bangkok, Thailand (online)

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-954085-71-8

Introduction

The MWE 2021 workshop (MWE 2021)[1] took place in an online format on August 6, 2021 in conjunction with ACL-IJCNLP 2021[2]. This was the 17th edition of the Workshop on Multiword Expressions (MWE 2021). The event was organized and sponsored by the Special Interest Group on the Lexicon (SIGLEX)[3] of the Association for Computational Linguistics (ACL).

Multiword expressions (MWEs) are word combinations, such as *in the middle of nowhere*, *hot dog*, *to make a decision* or *to kick the bucket*, displaying lexical, syntactic, semantic, pragmatic and/or statistical idiosyncrasies. Because of their unpredictable behavior, notably their non-compositional semantics, MWEs pose problems in linguistic modelling (e.g. treebank annotation, grammar engineering), Natural Language Processing (NLP) pipelines (in particular when orchestrated with parsing), and end-user NLP applications such as natural language understanding, machine translation, information extraction, and social media mining.

The special topic in this edition was the role of MWEs in end-user applications. On the one hand, the PARSEME shared tasks (Ramisch et al. 2020[4], Ramisch et al. 2018[5], Savary et al. 2017[6]), among others, fostered significant progress in MWE identification, providing datasets, evaluation measures and tools that now allow fully integrating MWE identification into end-user applications. On the other hand, NLP seems to be shifting towards end-to-end neural models capable of solving complex end-user tasks with little or no intermediary linguistic symbols, questioning the extent to which MWEs should be implicitly or explicitly modelled. Therefore, one goal of this workshop was to bring together and encourage researchers in various NLP subfields to submit MWE-related research, so that approaches that deal with MWEs in various applications could benefit from each other.

Traditional MWE topics

- Computationally-applicable theoretical work on MWEs and constructions in psycholinguistics and corpus linguistics
- MWE and construction annotation and representation in resources such as corpora, treebanks, e-lexicons and WordNets
- Processing of MWEs and constructions in syntactic and semantic frameworks (e.g. CCG, CxG, HPSG, LFG, TAG, UD, etc.)
- Discovery and identification methods for MWEs and constructions
- MWEs and constructions in language acquisition, language learning, and non-standard language (e.g. tweets, speech)
- Evaluation of annotation and processing techniques for MWEs and constructions
- Retrospective comparative analyses from the PARSEME shared tasks on automatic identification of MWEs

Topics on MWEs and end-user applications

- Processing of MWEs and constructions in end-user applications (e.g. MT, NLU, summarisation, social media mining, computer assisted language learning)

[1] https://multiword.org/mwe2021/

[2] https://2021.aclweb.org/

[3] https://siglex.org/

[4] https://www.aclweb.org/anthology/2020.mwe-1.14/

[5] https://www.aclweb.org/anthology/W18-4925/

[6] https://www.aclweb.org/anthology/W17-1704/

- Implicit and explicit representation of MWEs and constructions in end-user applications
- Evaluation of end-user applications concerning MWEs and constructions
- Resources and tools for MWEs and constructions (e.g. lexicons, identifiers) in end-user applications

Pursuing the MWE Section's tradition of synergies with other communities and in accordance with ACL-IJCNLP 2021's theme track on NLP for social good, a joint discussion panel was organized with the Workshop on Online Abuse and Harm (WOAH)[7].

This year, we received 19 submissions, among which 7 were accepted for presentation. The overall acceptance rate was 36%. In addition to the presentations, the workshop featured an invited talk that was given by Vered Shwartz, University of Washington.

We are grateful to the paper authors for their valuable contributions, the members of the Program Committee for their thorough and timely reviews, all members of the organizing committee for the fruitful collaboration, and all the workshop participants for their interest in this event. Our thanks also go to the ACL-IJCNLP 2021 organizers for their support, as well as to SIGLEX for their endorsement.

Paul Cook, Jelena Mitrović, Carla Parra Escartín, Ashwini Vaidya, Petya Osenova, Shiva Taslimipoor, Carlos Ramisch

[7] https://www.workshoponlineabuse.com/

Organizers

Program Chairs

Paul Cook, University of New Brunswick (Canada)
Jelena Mitrović, University of Passau (Germany)
Carla Parra Escartín, Iconic Translation Machines, Ltd. (Ireland)
Ashwini Vaidya, Indian Institute of Technology in Delhi (India)

Publication chairs

Petya Osenova, Institute of Information and Communication Technologies (Bulgaria)
Shiva Taslimipoor, University of Cambridge (UK)

Communication chair

Carlos Ramisch, Aix Marseille University (France)

Program Committee

Margarita Alonso-Ramos, Universidade da Coruña (Spain)
Tim Baldwin, University of Melbourne (Australia)
Verginica Barbu Mititelu, Romanian Academy (Romania)
Fabienne Cap, Uppsala University (Sweden)
Anastasia Christofidou, Academy of Athens (Greece)
Ken Church, IBM Research (USA)
Matthieu Constant, Université de Lorraine (France)
Monika Czerepowicka, University of Warmia and Mazury (Poland)
Myriam de Lhonneux, University of Copenhagen (Denmark)
Gaël Dias, University of Caen Basse-Normandie (France)
Meghdad Farahmand, University of Geneva (Switzerland)
Christiane Fellbaum, Princeton University (USA)
Joaquim Ferreira da Silva, New University of Lisbon (Portugal)
Karën Fort, Sorbonne Université (France)
Aggeliki Fotopoulou, ILSP/RC "Athena" (Greece)
Marcos Garcia, University of Santiago de Compostela (Spain)
Voula Giouli, Institute for Language and Speech Processing (Greece)
Stefan Th. Gries, University of California (USA)
Bruno Guillaume, Université de Lorraine (France)
Chikara Hashimoto, Yahoo!Japan (Japan)
Uxoa Iñurrieta, University of the Basque Country (Spain)
Diptesh Kanojia, IIT Bombay (India)
Elma Kerz, RWTH Aachen (Germany)
Ekaterina Kochmar, University of Cambridge (UK)
Dimitrios Kokkinakis, University of Gothenburg (Sweden)
Ioannis Korkontzelos, Edge Hill University (UK)
Cvetana Krstev, University of Belgrade (Serbia)
Eric Laporte, University Paris-Est Marne-la-Vallee (France)
Timm Lichte, University of Duesseldorf (Germany)
Teresa Lynn, ADAPT Centre (Ireland)

Stella Markantonatou, Institute for Language and Speech Processing (Greece)
Yuji Matsumoto, Nara Institute of Science and Technology (Japan)
Nurit Melnik, The Open University of Israel (Israel)
Laura A. Michaelis, University of Colorado Boulder (USA)
Johanna Monti, “L’Orientale” University of Naples (Italy)
Preslav Nakov, Qatar Computing Research Institute, HBKU (Qatar)
Malvina Nissim, University of Groningen (Netherlands)
Diarmuid Ó Séaghdha, University of Cambridge (UK)
Jan Odijk, University of Utrecht (Netherlands)
Haris Papageorgiou, Institute for Language and Speech Processing (Greece)
Marie-Sophie Pausé, independent researcher (France)
Pavel Pecina, Charles University (Czech Republic)
Ted Pedersen, University of Minnesota (USA)
Scott Piao, Lancaster University (UK)
Maciej Piasecki, Wroclaw University of Technology (Poland)
Alain Polguère, Université de Lorraine (France)
Matīss Rikters, University of Tokyo (Japan)
Fatiha Sadat, Université du Québec à Montréal (Canada)
Manfred Sailer, Goethe-Universität Frankfurt am Main (Germany)
Magali Sanches Duran, University of São Paulo (Brazil)
Branislava Šandrih, University of Belgrade (Serbia)
Agata Savary, Université François Rabelais Tours (France)
Sabine Schulte im Walde, University of Stuttgart (Germany)
Matthew Shardlow, Manchester Metropolitan University (UK)
Vered Shwartz, Allen AI (USA)
Gyri Smørdal Losnegaard, University of Bergen (Norway)
Ranka Stanković, University of Belgrade (Serbia)
Ivelina Stoyanova, Bulgarian Academy of Sciences (Bulgaria)
Stan Szpakowicz, University of Ottawa (Canada)
Carole Tiberius, Dutch Language Institute (Netherlands)
Beata Trawinski, Institut für Deutsche Sprache Mannheim (Germany)
Ruben Urizar, University of the Basque Country (Spain)
Aline Villavicencio, Federal University of Rio Grande do Sul (Brazil)
Veronika Vincze, Hungarian Academy of Sciences (Hungary)
Martin Volk, University of Zürich (Switzerland)
Zeerak Waseem, University of Sheffield (UK)
Eric Wehrli, University of Geneva (Switzerland)
Seid Muhie Yimam, Universität Hamburg (Germany)

Invited Speaker

Vered Shwartz, University of Washington

Table of Contents

Abstract of the Invited Talk

Papers

Workshop Program

August 6, 2021
[All times are in CEST (UTC+2)]

14:00–14:10 Welcome and Preparation

14:10–15:50 Session 1: Long Papers

14:10–14:30 *Where Do Aspectual Variants of Light Verb Constructions Belong?*
Aggeliki Fotopoulou, Eric Laporte and Takuya Nakamura

14:30–14:50 *Data-driven Identification of Idioms in Song Lyrics*
Miriam Amin, Peter Fankhauser, Marc Kupietz and Roman Schneider

14:50–15:10 *(From Findings of ACL 2021) Transforming Term Extraction: Transformer-Based Approaches to Multilingual Term Extraction Across Domains*
Christian Lang, Lennart Wachowiak, Barbara Heinisch, Dagmar Gromann

15:10–15:30 *Contextualized Embeddings Encode Monolingual and Cross-lingual Knowledge of Idiomaticity*
Samin Fakharian and Paul Cook

15:30–15:50 *PIE: A Parallel Idiomatic Expression Corpus for Idiomatic Sentence Generation and Paraphrasing*
Jianing Zhou, Hongyu Gong and Suma Bhat

15:50–16:05 Break

16:05–17:05 Invited Talk

16:05–17:05 *A Long Hard Look at MWEs in the Age of Language Models*
Vered Shwartz

17:05–17:20 Break

17:20–18:05 Session 2: Short Papers

17:20–17:35 *Lexical Semantic Recognition*
Nelson F. Liu, Daniel Hershcovich, Michael Kranzlein and Nathan Schneider

17:35–17:50 *Finding BERT's Idiomatic Key*
Vasudevan Nedumpozhimana and John Kelleher

17:50–18:05 *Light Verb Constructions and Their Families - A Corpus Study on German 'stehen unter'-LVCs*
Jens Fleischhauer

18:05–18:20 Break

August 6, 2021 (continued)

18:20–19:00 Joint session with WOAH

19:00–19:20 Community discussion