

ACL 2019

**Joint Workshop on Multiword Expressions and WordNet
(MWE-WN 2019)**

Proceedings of the Workshop

August 2, 2019
Florence, Italy

©2019 The Association for Computational Linguistics

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 978-1-950737-26-0

Introduction

The Joint Workshop on Multiword Expressions and WordNet (MWE-WN 2019)¹ took place on August 2, 2019 in Florence (Italy), in conjunction with the 57th Annual Meeting of the Association for Computational Linguistics (ACL 2019). This was the 15th edition of the Workshop on Multiword Expressions (MWE 2019). The event was organized and sponsored by the Special Interest Group on the Lexicon (SIGLEX)² of the Association for Computational Linguistics (ACL). It was also endorsed by the Global WordNet Association (GWA)³.

The workshop brought together two research communities studying multiword expressions and wordnets. *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. Computational research on MWEs encompasses NLP modeling and processing, as well as annotation. *Wordnets* include MWEs and link their meanings into a shared network. For instance, the following simple words and multiword expressions *talk*, *blab*, *sing*, *spill the beans*, *let the cat out of the bag*, *tattle*, *peach*, *babble*, *babble out*, *blab out* are all part of the same synset, which has the gloss “divulge confidential information or secrets”. Over 50% of entries in the Princeton WordNet are MWEs and most other projects have a similarly high percentage.

In order to allow better convergence and scientific innovation within these two largely complementary scientific communities, we called for papers on joint topics on MWEs and wordnets, on the one hand, and on MWE-specific topics, on the other hand. With the intention to also perpetuate previous converging effects with the Construction Grammar community (see the LAW-MWE-CxG 2018 workshop), we extended the traditional MWE scope to grammatical constructions. The topics included, but were not limited to:

- Joint topics on MWEs and wordnets:
 - Encoding MWEs in wordnets – how we can take advantage of the existing rich structure of wordnets
 - Encoding MWEs in wordnets – consequences for a lexical-semantic organization of MWEs
 - Linking wordnets with existing MWE lexicons
 - Word sense disambiguation for single-word and multiword expressions
 - Cross-wordnet and cross-language comparisons of MWEs
 - MWEs in sense-annotated corpora
 - Semantic relations in wordnets related to MWEs
- MWE-specific topics:
 - Computationally-applicable theoretical studies on MWEs and constructions in psycholinguistics, corpus linguistics and formal grammars
 - MWE and construction annotation in corpora and treebanks
 - MWE and construction representation in manually/automatically constructed lexical resources
 - Processing of MWEs and constructions in syntactic and semantic frameworks (e.g. CCG, CxG, HPSG, LFG, TAG, UD, etc.), and in end-user applications (e.g. information extraction, machine translation and summarization)
 - Original discovery and identification methods for MWEs and constructions

¹<http://multiword.sourceforge.net/mwewn2019/>

²<http://alt.qcri.org/siglex/>

³<http://globalwordnet.org/>

- MWEs and constructions in language acquisition and in non-standard language (e.g. tweets, forums, spontaneous speech)
- Evaluation of annotation and processing techniques for MWEs and constructions
- Retrospective comparative analyses from the PARSEME shared tasks on automatic identification of MWEs

We received 37 submissions (21 long and 16 short papers). We selected 12 long papers and 8 short ones. From those, 6 papers were presented orally and the remaining 14 as posters. The overall acceptance rate was 54%. Of the 20 presented papers, 6 concerned both wordnets and MWEs, which makes us believe that the intended synergy effect has been achieved.

In addition to the oral and poster sessions, the workshop featured an invited talk, given by Aline Villavicencio.

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 to all the workshop participants for their interest in this event. Our thanks also go to the ACL 2019 organizers for their support, as well as to SIGLEX and GWA for their endorsement.

Agata Savary, Carla Parra Escartín, Francis Bond, Jelena Mitrović, Verginica Barbu Mititelu

Organizers

Organizers:

Agata Savary, University of Tours (France)
Carla Parra Escartín, Unbabel, Lisbon (Portugal)
Francis Bond, Nanyang Technological University (Singapore)
Jelena Mitrović, University of Passau (Germany)
Verginica Barbu Mititelu, Romanian Academy Research Institute for Artificial Intelligence (Romania)

Program Committee:

Eneko Agirre, University of the Basque Country (Spain)
Tim Baldwin, University of Melbourne (Australia)
Archana Bhatia, Florida Institute for Human and Machine Cognition (USA)
Sonja Bosch, Department of African Languages, University of South Africa (South Africa)
Miriam Butt, Universität Konstanz (Germany)
Aoife Cahill, ETS (USA)
Marie Candito, Paris Diderot University (France)
Annalina Caputo, ADAPT Centre / Trinity College Dublin (Ireland)
Helena Caseli, Federal University of Sao Carlos (Brazil)
Anastasia Christofidou, Academy of Athens (Greece)
Matthieu Constant, Université de Lorraine (France)
Silvio Cordeiro, Federal University of Rio Grande do Sul (Brazil)
Janos Csirik, University of Szeged (Hungary)
Gaël Dias, University of Caen Basse-Normandie (France)
Gülşen Eryiğit, Istanbul Technical University (Turkey)
Stefan Evert, FAU Erlangen-Nürnberg (Germany)
Christiane Fellbaum, Princeton University (USA)
Joaquim Ferreira da Silva, New University of Lisbon (Portugal)
Darja Fišer, University of Ljubljana (Slovenia)
Dan Flickinger, Stanford University (USA)
Aggeliki Fotopoulou, ILSP/RC "Athena" (Greece)
Voula Giouli, Institute for Language and Speech Processing (Greece)
Chikara Hashimoto, Yahoo! Japan (Japan)
Ales Horak, Masaryk University (Czech Republic)
Shu-Kai Hsieh, National Taiwan Normal University (Taiwan)
Hitoshi Isahara, Toyohashi University of Technology (Japan)
Kyo Kageura, University of Tokyo (Japan)
Diptesh Kanojia, IIT Bombay (India)
Kyoko Kanzaki, Toyohashi University of Technology (Japan)
Philipp Koehn, University of Edinburgh (UK)
Dimitris Kokkinakis, University of Gothenburg (Sweden)
Olga Kolesnikova, Instituto Politécnico Nacional (Mexico)
Ioannis Korkontzelos, Edge Hill University (UK)
Cvetana Krstev, University of Belgrade (Serbia)
Tim Lichte, University of Duesseldorf (Germany)
Irina Lobzhanidze, Iliia State University (Georgia)
Ismail el Maarouf, Adarga Ltd (UK)

Stella Markantonatou, Institute for Language and Speech Processing (Greece)
 Héctor Martínez Alonso, Apple (UK)
 John P. McCrae, National University of Ireland, Galway (Ireland)
 Nurit Melnik, The Open University of Israel (Israel)
 Gerard de Melo, Rutgers University (USA)
 Johanna Monti, "L'Orientale" University of Naples (Italy)
 Preslav Nakov, Qatar Computing Research Institute, HBKU (Qatar)
 Joakim Nivre, Uppsala University (Sweden)
 Jan Odijk, University of Utrecht (Netherlands)
 Antoni Oliver Gonzales, Universitat Oberta de Catalunya (Spain)
 Heili Orav, University of Tartu (Estonia)
 Petya Osenova, Bulgarian Academy of Sciences (Bulgaria)
 Haris Papageorgiou, Institute for Language and Speech Processing (Greece)
 Yannick Parmentier, Université d'Orléans (France)
 Agnieszka Patejuk, University of Oxford (UK); Institute of Computer Science, Polish Academy of Sciences (Poland)
 Marie-Sophie Pausé, University of Paris 3 (France)
 Adam Pease, Articulate Software (USA)
 Pavel Pecina, Charles University (Czech Republic)
 Bolette Pedersen, University of Copenhagen (Denmark)
 Ted Pedersen, University of Minnesota (USA)
 Scott Piao, Lancaster University (UK)
 Maciej Piasecki, Wroclaw University of Technology (Poland)
 Alain Polguère, Université de Lorraine (France)
 Marten Postma, Vrije Universiteit Amsterdam (Netherlands)
 Behrang QuasemiZadeh, University of Duesseldorf (Germany)
 Alexandre Rademaker, IBM Research Brazil and EMap/FGV (Brazil)
 Carlos Ramisch, Aix Marseille University (France)
 German Rigau, University of the Basque Country (Spain)
 Mike Rosner, University of Malta (Malta)
 Ewa Rudnicka, Wrocław University of Technology (Poland)
 Manfred Sailer, Goethe-Universität Frankfurt am Main (Germany)
 Federico Sangati, Independent researcher (Italy)
 Kevin Scannell, Saint Louis University (USA)
 Nathan Schneider, Georgetown University (USA)
 Sabine Schulte im Walde, University of Stuttgart (Germany)
 Kiril Simov, Bulgarian Academy of Sciences (Bulgaria)
 Jan Šnajder, University of Zagreb (Croatia)
 Ranka Stanković, University of Belgrade (Serbia)
 Ivelina Stoyanova, Bulgarian Academy of Sciences (Bulgaria)
 Stan Szpakowicz, University of Ottawa (Canada)
 Beata Trawinski, Institut für Deutsche Sprache Mannheim (Germany)
 Dan Tufiş, Romanina Academey (Romania)
 Ruben Urizar, University of the Basque Country (Spain)
 E Umamaheswari Vasanthakumar, Nanyang Technological University (Singapore)
 Veronika Vincze, Hungarian Academy of Sciences (Hungary)
 Piek Vossen, VU University Amsterdam. (Netherlands)
 Shan Wang, University of Macau (China)
 Jakub Waszczuk, University of Duesseldorf (Germany)
 Marion Weller-Di Marco, University of Amsterdam (Netherlands)

Invited Speaker:

Aline Villavicencio, Federal University of Rio Grande do Sul (Brazil); University of Essex (UK)

Table of Contents

Summary of the Invited Talk

<i>When the Whole Is Greater Than the Sum of its Parts: Multiword Expressions and Idiomaticity</i> Aline Villavicencio	1
---	---

Long Papers

<i>Hear about Verbal Multiword Expressions in the Bulgarian and the Romanian Wordnets Straight from the Horse's Mouth</i> Verginica Barbu Mititelu, Ivelina Stoyanova, Svetlozara Leseva, Maria Mitrofan, Tsvetana Dimitrova and Maria Todorova	2
<i>The Romanian Corpus Annotated with Verbal Multiword Expressions</i> Verginica Barbu Mititelu, Mihaela Cristescu and Mihaela Onofrei	13
<i>Using OntoLex-Lemon for Representing and Interlinking German Multiword Expressions in OdeNet and MMORPH</i> Thierry Declerck, Melanie Siegel and Stefania Racioppa	22
<i>Learning to Predict Novel Noun-Noun Compounds</i> Prajit Dhar and Lonneke van der Plas	30
<i>Unsupervised Compositional Translation of Multiword Expressions</i> Pablo Gamallo and Marcos Garcia	40
<i>A Comparison of Statistical Association Measures for Identifying Dependency-based Collocations in Various Languages.</i> Marcos Garcia, Marcos García Salido and Margarita Alonso-Ramos	49
<i>L2 Processing Advantages of Multiword Sequences: Evidence from Eye-Tracking</i> Elma Kerz, Arndt Heilmann and Stella Neumann	60
<i>Modeling MWEs in BTB-WN</i> Laska Laskova, Petya Osenova, Kiril Simov, Ivajlo Radev and Zara Kancheva	70
<i>Without Lexicons, Multiword Expression Identification Will Never Fly: A Position Statement</i> Agata Savary, Silvio Cordeiro and Carlos Ramisch	79
<i>A Systematic Comparison of English Noun Compound Representations</i> Vered Shwartz	92
<i>Semantic Modelling of Adjective-Noun Collocations Using FrameNet</i> Yana Strakatova and Erhard Hinrichs	104
<i>A Neural Graph-based Approach to Verbal MWE Identification</i> Jakub Waszczuk, Rafael Ehren, Regina Stodden and Laura Kallmeyer	114

Short Papers

<i>Confirming the Non-compositionality of Idioms for Sentiment Analysis</i> Alyssa Hwang and Christopher Hidey	125
---	-----

<i>IDION: A Database for Modern Greek Multiword Expressions</i> Stella Markantonatou, Panagiotis Minos, George Zakis, Vassiliki Moutzouri and Maria Chantou	130
<i>Identification of Adjective-Noun Neologisms Using Pretrained Language Models</i> John Philip McCrae	135
<i>Neural Lemmatization of Multiword Expressions</i> Marine Schmitt and Mathieu Constant	142
<i>Evaluating Automatic Term Extraction Methods on Individual Documents</i> Antonio Šajatović, Maja Buljan, Jan Šnajder and Bojana Dalbelo Bašić	149
<i>Cross-lingual Transfer Learning and Multitask Learning for Capturing Multiword Expressions</i> Shiva Taslimipour, Omid Rohanian and Le An Ha	155
<i>Ilfhocail: A Lexicon of Irish MWEs</i> Abigail Walsh, Teresa Lynn and Jennifer Foster	162
<i>The Impact of Word Representations on Sequential Neural MWE Identification</i> Nicolas Zampieri, Carlos Ramisch and Geraldine Damnati	169

Workshop Program

- 8:55–9:00** *Opening*
- 9:00–10:00** **Session 1: Invited Talk**
When the Whole Is Greater Than the Sum of its Parts: Multiword Expressions and Idiomaticity
Aline Villavicencio
- 10:00–10:30** **Session 2: Poster boosters**
- 10:30–11:00 *Coffee break*
- 11:00–12:30** **Session 3: Posters**
The Romanian Corpus Annotated with Verbal Multiword Expressions
Verginica Barbu Mititelu, Mihaela Cristescu and Mihaela Onofrei
Learning to Predict Novel Noun-Noun Compounds
Prajit Dhar and Lonneke van der Plas
A Comparison of Statistical Association Measures for Identifying Dependency-based Collocations in Various Languages.
Marcos Garcia, Marcos García Salido and Margarita Alonso-Ramos
Confirming the Non-compositionality of Idioms for Sentiment Analysis
Alyssa Hwang and Christopher Hidey
L2 Processing Advantages of Multiword Sequences: Evidence from Eye-Tracking
Elma Kerz, Arndt Heilmann and Stella Neumann
Identification of Adjective-Noun Neologisms Using Pretrained Language Models
John Philip McCrae
IDION: A Database for Modern Greek Multiword Expressions
Stella Markantonatou, Panagiotis Minos, George Zakis, Vassiliki Moutzouri and Maria Chantou
Evaluating Automatic Term Extraction Methods on Individual Documents
Antonio Šajatović, Maja Buljan, Jan Šnajder and Bojana Dalbelo Bašić
A Systematic Comparison of English Noun Compound Representations
Vered Shwartz
Semantic Modelling of Adjective-Noun Collocations Using FrameNet
Yana Strakatova and Erhard Hinrichs
Cross-lingual Transfer Learning and Multitask Learning for Capturing Multiword Expressions
Shiva Taslimipoor, Omid Rohanian and Le An Ha
The Impact of Word Representations on Sequential Neural MWE Identification
Nicolas Zampieri, Carlos Ramisch and Geraldine Damnati
Ilfhocail: A Lexicon of Irish MWEs
Abigail Walsh, Teresa Lynn and Jennifer Foster
A Neural Graph-based Approach to Verbal MWE Identification
Jakub Waszczuk, Rafael Ehren, Regina Stodden and Laura Kallmeyer

- 12:30–14:00 *Lunch break*
- Session 4: Multiword Expressions and WordNet**
- 14:00–14:30 *Hear about Verbal Multiword Expressions in the Bulgarian and the Romanian Wordnets Straight from the Horse’s Mouth*
Verginica Barbu Mititelu, Ivelina Stoyanova, Svetlozara Leseva, Maria Mitrofan, Tsvetana Dimitrova and Maria Todorova
- 14:30–15:00 *Modeling MWEs in BTB-WN*
Laska Laskova, Petya Osenova, Kiril Simov, Ivajlo Radev and Zara Kancheva
- 15:00–15:30 *Using OntoLex-Lemon for Representing and Interlinking German Multiword Expressions in OdeNet and MMORPH*
Thierry Declerck, Melanie Siegel and Stefania Racioppa
- 15:30–16:00 *Coffee break*
- Session 5: Multiword Expressions – translation, lemmatization and identification**
- 16:00–16:30 *Unsupervised Compositional Translation of Multiword Expressions*
Pablo Gamallo and Marcos Garcia
- 16:30–16:50 *Neural Lemmatization of Multiword Expressions*
Marine Schmitt and Mathieu Constant
- 16:50–17:20 *Without Lexicons, Multiword Expression Identification Will Never Fly: A Position Statement*
Agata Savary, Silvio Cordeiro and Carlos Ramisch
- 17:20–18:00 Session 6: Community discussion**
News, future work, SIGLEX MWE section