

MWE 2017

13th Workshop on Multiword Expressions

Proceedings of the Workshop

EACL 2017 Workshop

April 4, 2017

Valencia, Spain

Sponsors



P A R S  M E

©2017 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 978-1-945626-48-7

Introduction

The 13th edition of the annual *Workshop on Multiword Expressions* (MWE 2017)¹ took place on April 4, 2017 in Valencia, Spain, in conjunction with the 15th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017). It was endorsed by the Special Interest Group on the Lexicon of the Association for Computational Linguistics (SIGLEX)² and by the European IC1207 COST Action PARSEME.³ The workshop featured a dedicated track for the Shared Task on the Identification of Verbal Multiword Expressions.

The workshop has been held since 2001 in conjunction with major computational linguistics conferences (ACL, COLING, LREC, EACL). It attracts the attention of an ever-growing community working on a variety of languages and linguistic phenomena and represents an important venue for the community to interact, share resources and tools, and collaborate on efforts for advancing the computational treatment of multiword expressions.

In this 13th edition of the workshop, we have called for papers on major challenges in MWE processing, both from theoretical and computational viewpoints, focusing on research related (but not limited) to the following topics:

- Manually and automatically constructed lexical resources
- MWE representation in lexical resources
- MWE annotation in corpora and treebanks
- MWEs in non-standard language (e.g. tweets, forums, spontaneous speech)
- Original MWE discovery methods (e.g. using word embeddings, parallel corpora)
- Original MWE in-context identification methods (e.g. using deep learning, topic models)
- MWE processing in syntactic frameworks (e.g. HPSG, LFG, TAG, universal dependencies)
- MWE processing in semantic frameworks (e.g. WSD, semantic parsing)
- MWE processing in end-user applications (e.g. summarization, machine translation)
- Orchestration of MWE processing with respect to applications
- Evaluation of MWE processing techniques
- Models of first and second language acquisition of MWEs
- Theoretical and psycholinguistic studies on MWEs
- Crosslinguistic studies on MWEs

Submissions included both long and short papers. In total, 34 papers (22 short and 12 long) were submitted to the *main track* of the workshop. From those, 7 papers (4 short and 3 long) were accepted as oral presentations and 14 papers (11 short and 3 long) were accepted as posters. The overall acceptance rate is 62% including oral presentations and posters, short and long papers.

Additionally, 6 system description papers and 1 shared task description paper were submitted to the *shared task track*. The former were all selected as posters and the latter as an oral presentation. The reviewing modalities were different in this track, therefore we do not count these papers in the workshop acceptance rate.

In addition to the oral and poster sessions, the workshop featured an invited talk by Paul Cook, from the University of New Brunswick, Canada. The presentation was entitled *Exploiting multilingual lexical*

¹<http://multiword.sf.net/mwe2017>

²<http://www.siglex.org/>

³<http://parseme.eu/>

resources to predict the compositionality of MWEs. The program also included a panel discussion on the future directions of the MWE community and the SIGLEX Section.

We would like to thank the members of the program committee for the timely reviews, authors for their valuable contributions, shared task organizers, annotators, and system developers for their hard work, and all involved participants for their interest in the workshop. We also want to thank the IC1207 COST Action PARSEME and SIGLEX for their endorsement and support, as well as the EACL 2017 organizers.

Stella Markantonatou, Carlos Ramisch, Agata Savary, and Veronika Vincze

Organizers:

Stella Markantonatou, Institute for Language and Speech Processing/RC “Athena” (Greece)
Carlos Ramisch, Aix Marseille University (France)
Agata Savary, Université François Rabelais Tours (France)
Veronika Vincze, Hungarian Academy of Sciences (Hungary)

Program Committee - Main Track:

Iñaki Alegria, University of the Basque Country (Spain)
Anna Anastassiadis-Symeonidis, Aristotle University of Thessaloniki (Greece)
Dimitra Anastasiou, Luxemburg Institute of Science and Technology (Luxembourg)
Doug Arnold, University of Essex (UK)
Timothy Baldwin, The University of Melbourne (Australia)
Eduard Bejček, Charles University (Czech Republic)
Francis Bond, Nanyang Technological University (Singapore)
Antonio Branco, University of Lisbon (Portugal)
Miriam Butt, Universität Konstanz (Germany)
Marie Candito, Paris Diderot University (France)
Fabienne Cap, Uppsala University (Sweden)
Marine Carpuat, University of Maryland (USA)
Helena Caseli, Federal University of São Carlos (Brazil)
Anastasia Christofidou, Academy of Athens (Greece)
Ken Church, IBM Watson (USA)
Matthieu Constant, Université de Lorraine (France)
Silvio Cordeiro, Aix Marseille University (France) and Federal University of Rio Grande do Sul (Brazil)
Béatrice Daille, Nantes University (France)
Koenraad de Smedt, University of Bergen (Norway)
Mona Diab, The George Washington University (USA)
Gaël Dias, University of Caen Basse-Normandie (France)
Ismail El Maarouf, Adarga Ltd (UK)
Gülşen Eryiğit, Istanbul Technical University (Turkey)
Stefan Evert, FAU Erlangen-Nürnberg (Germany)
Meghdad Farahmand, University of Geneva (Switzerland)
Joaquim Ferreira da Silva, New University of Lisbon (Portugal)
Dan Flickinger, Stanford University (USA)
Angeliki Fotopoulou, Institute for Language and Speech Processing/RC “Athena” (Greece)
Voula Giouli, Institute for Language and Speech Processing/RC “Athena” (Greece)
Antton Gurrutxaga, Elhuyar Foundation (Basque Country, Spain)
Chikara Hashimoto, Yahoo!Japan (Japan)
Kyo Kageura, University of Tokyo (Japan)
Philipp Koehn, John Hopkins University (USA)
Dimitris Kokkinakis, University of Gothenburg (Sweden)
Yannis Korkontzelos, Edge Hill University (UK)
Brigitte Krenn, Austrian Research Institute for Artificial Intelligence (Austria)
Cvetana Krstev, University of Belgrade (Serbia)
Eric Laporte, University Paris-Est Marne-la-Vallee (France)
Evita Linardaki, Hellenic Open University (Greece)

Hector Martínez Alonso, INRIA (France)
Diana McCarthy, University of Cambridge (UK)
Johanna Monti, “L’Orientale” University of Naples (Italy)
Preslav Nakov, Qatar Computing Research Institute, HBKU (Qatar)
Joakim Nivre, Uppsala University (Sweden)
Diarmuid Ó Séaghdha, Apple (UK)
Michael Oakes, University of Wolverhampton (UK)
Jan Odijk, University of Utrecht (The Netherlands)
Petya Osenova, Bulgarian Academy of Sciences (Bulgaria)
Haris Papageorgiou, Institute for Language and Speech Processing/RC “Athena” (Greece)
Yannick Parmentier, Université d’Orléans (France)
Carla Parra Escartín, ADAPT Centre, Dublin City University (Ireland)
Agnieszka Patejuk, Institute of Computer Science, Polish Academy of Sciences (Poland)
Pavel Pecina, Charles University (Czech Republic)
Scott Piao, Lancaster University (UK)
Thierry Poibeau, CNRS and École Normale Supérieure (France)
Martin Riedl, University of Hamburg (Germany)
Mike Rosner, University of Malta (Malta)
Manfred Sailer, Goethe-Universität Frankfurt am Main (Germany)
Magali Sanches Duran, University of São Paulo (Brazil)
Federico Sangati, Independent researcher (Italy)
Nathan Schneider, Georgetown University (USA)
Sabine Schulte im Walde, University of Stuttgart (Germany)
Serge Sharoff, University of Leeds (UK)
Kiril Simov, Bulgarian Academy of Sciences (Bulgaria)
Sara Stymne, Uppsala University (Sweden)
Stan Szpakowicz, University of Ottawa (Canada)
Beata Trawinski, Institut für Deutsche Sprache Mannheim (Germany)
Yulia Tsvetkov, Stanford University (USA)
Yuancheng Tu, Microsoft (USA)
Ruben Urizar, University of the Basque Country (Spain)
Lonneke van der Plas, University of Malta (Malta)
Gertjan van Noord, University of Groningen (The Netherlands)
Aline Villavicencio, Federal University of Rio Grande do Sul (Brazil)
Tom Wasow, Stanford University (USA)
Marion Weller-Di Marco, University of Stuttgart (Germany)
Shuly Wintner, University of Haifa (Israel)

Program Committee - Shared Task Track:

Verginica Barbu Mititelu, ICIA, Romanian Academy (Romania)
Eduard Bejček, Charles University (Czech Republic)
Marie Candito, Paris Diderot University (France)
Fabienne Cap, Uppsala University (Sweden)
Matthieu Constant, Université de Lorraine (France)
Silvio Cordeiro, Aix Marseille University (France) and Federal University of Rio Grande do Sul (Brazil)
Koenraad De Smedt, University of Bergen (Norway)
Gaël Dias, University of Caen Basse-Normandie (France)
Ismail El Maarouf, Adarga Ltd (UK)
Gülşen Eryiğit, Istanbul Technical University (Turkey)

Voula Giouli, Institute for Language and Speech Processing/RC “Athena” (Greece)
Natalia Klyueva, Charles University (Czech Republic)
Yannis Korkontzelos, Edge Hill University (UK)
Johanna Monti, “L’Orientale” University of Naples (Italy)
Luka Nerima, University of Geneva (Switzerland)
Michael Oakes, University of Wolverhampton (UK)
Carla Parra Escartín, ADAPT Centre, Dublin City University (Ireland)
Federico Sangati, Independent researcher (Italy)
Nathan Schneider, Georgetown University (USA)
Katalin Ilona Simkó, University of Szeged (Hungary)
Kiril Simov, Bulgarian Academy of Sciences (Bulgaria)
Lonneke van der Plas, University of Malta (Malta)

Invited Speaker:

Paul Cook, University of New Brunswick (Canada)

Table of Contents

<i>ParaDi: Dictionary of Paraphrases of Czech Complex Predicates with Light Verbs</i> Petra Barancikova and Václava Kettnerová	1
<i>Multi-word Entity Classification in a Highly Multilingual Environment</i> Sophie Chesney, Guillaume Jacquet, Ralf Steinberger and Jakub Piskorski	11
<i>Using bilingual word-embeddings for multilingual collocation extraction</i> Marcos Garcia, Marcos García-Salido and Margarita Alonso-Ramos	21
<i>The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions</i> Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang Qasem-izadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova and Antoine Doucet	31
<i>USzeged: Identifying Verbal Multiword Expressions with POS Tagging and Parsing Techniques</i> Katalin Ilona Simkó, Viktória Kovács and Veronika Vincze	48
<i>Parsing and MWE Detection: Fips at the PARSEME Shared Task</i> Luka Nerima, Vasiliki Foufi and Eric Wehrli	54
<i>Neural Networks for Multi-Word Expression Detection</i> Natalia Klyueva, Antoine Doucet and Milan Straka	60
<i>Factoring Ambiguity out of the Prediction of Compositionality for German Multi-Word Expressions</i> Stefan Bott and Sabine Schulte im Walde	66
<i>Multiword expressions and lexicalism: the view from LFG</i> Jamie Y. Findlay	73
<i>Understanding Idiomatic Variation</i> Kristina Geeraert, R. Harald Baayen and John Newman	80
<i>Discovering Light Verb Constructions and their Translations from Parallel Corpora without Word Alignment</i> Natalie Vargas, Carlos Ramisch and Helena Caseli	91
<i>Identification of Multiword Expressions for Latvian and Lithuanian: Hybrid Approach</i> Justina Mandravickaite and Tomas Krilavičius	97
<i>Show Me Your Variance and I Tell You Who You Are - Deriving Compound Compositionality from Word Alignments</i> Fabienne Cap	102
<i>Semantic annotation to characterize contextual variation in terminological noun compounds: a pilot study</i> Melania Cabezas-García and Antonio San Martín	108
<i>Detection of Verbal Multi-Word Expressions via Conditional Random Fields with Syntactic Dependency Features and Semantic Re-Ranking</i> Alfredo Maldonado, Lifeng Han, Erwan Moreau, Ashjan Alsulaimani, Koel Dutta Chowdhury, Carl Vogel and Qun Liu	114

<i>A data-driven approach to verbal multiword expression detection. PARSEME Shared Task system description paper</i>	
Tiberiu Boroş, Sonia Pipa, Verginica Barbu Mititelu and Dan Tufiş	121
<i>The ATILF-LLF System for Parseme Shared Task: a Transition-based Verbal Multiword Expression Tagger</i>	
Hazem Al Saied, Matthieu Constant and Marie Candito	127
<i>Investigating the Opacity of Verb-Noun Multiword Expression Usages in Context</i>	
Shiva Taslimipour, Omid Rohanian, Ruslan Mitkov and Afsaneh Fazly	133
<i>Compositionality in Verb-Particle Constructions</i>	
Archana Bhatia, Choh Man Teng and James Allen	139
<i>Rule-Based Translation of Spanish Verb-Noun Combinations into Basque</i>	
Uxoa Iñurrieta, Itziar Aduriz, Arantza Diaz de Ilarraza, Gorka Labaka and Kepa Sarasola	149
<i>Verb-Particle Constructions in Questions</i>	
Veronika Vincze	155
<i>Simple Compound Splitting for German</i>	
Marion Weller-Di Marco	161
<i>Identification of Ambiguous Multiword Expressions Using Sequence Models and Lexical Resources</i>	
Manon Scholivet and Carlos Ramisch	167
<i>Comparing Recurring Lexico-Syntactic Trees (RLTs) and Ngram Techniques for Extended Phraseology Extraction</i>	
Agnès Tutin and Olivier Kraif	176
<i>Benchmarking Joint Lexical and Syntactic Analysis on Multiword-Rich Data</i>	
Matthieu Constant and Héctor Martínez Alonso	181
<i>Semi-Automated Resolution of Inconsistency for a Harmonized Multiword Expression and Dependency Parse Annotation</i>	
King Chan, Julian Brooke and Timothy Baldwin	187
<i>Combining Linguistic Features for the Detection of Croatian Multiword Expressions</i>	
Maja Buljan and Jan Šnajder	194
<i>Complex Verbs are Different: Exploring the Visual Modality in Multi-Modal Models to Predict Compositionality</i>	
Maximilian Köper and Sabine Schulte im Walde	200

Workshop Program

Tuesday, April 4 2017

09:15–09:20 *Opening Remarks*

09:20–09:30 *COST – Creating Spaces for People and Ideas to Grow*
Ralph Stuebner

Oral Session - Long Papers

09:30–10:00 *ParaDi: Dictionary of Paraphrases of Czech Complex Predicates with Light Verbs*
Petra Barancikova and Václava Kettnerová

10:00–10:30 *Multi-word Entity Classification in a Highly Multilingual Environment*
Sophie Chesney, Guillaume Jacquet, Ralf Steinberger and Jakub Piskorski

10:30–11:00 *Using bilingual word-embeddings for multilingual collocation extraction*
Marcos Garcia, Marcos García-Salido and Margarita Alonso-Ramos

11:00–11:30 COFFEE BREAK

Invited Talk

11:30–12:30 *Exploiting multilingual lexical resources to predict the compositionality of MWEs*
Paul Cook

Shared Task Session

12:30–12:50 *The PARSEME Shared Task on Automatic Identification of Verbal Multiword Expressions*

Agata Savary, Carlos Ramisch, Silvio Cordeiro, Federico Sangati, Veronika Vincze, Behrang QasemiZadeh, Marie Candito, Fabienne Cap, Voula Giouli, Ivelina Stoyanova and Antoine Doucet

12:50–13:00 *Shared Task Poster Boosters (1min per poster - ST1 to ST6)*

13:00–14:30 LUNCH

14:30–14:50 *Poster Session Boosters (1min per poster - A01 to A07, B08 to B14)*

14:50–15:25 Poster Session A

ST1 *USzeged: Identifying Verbal Multiword Expressions with POS Tagging and Parsing Techniques*

Katalin Ilona Simkó, Viktória Kovács and Veronika Vincze

ST2 *Parsing and MWE Detection: Fips at the PARSEME Shared Task*

Luka Nerima, Vasiliki Foufi and Eric Wehrli

ST3 *Neural Networks for Multi-Word Expression Detection*

Natalia Klyueva, Antoine Doucet and Milan Straka

A01 *Factoring Ambiguity out of the Prediction of Compositionality for German Multi-Word Expressions*

Stefan Bott and Sabine Schulte im Walde

A02 *Multiword expressions and lexicalism: the view from LFG*

Jamie Y. Findlay

A03 *Understanding Idiomatic Variation*

Kristina Geeraert, R. Harald Baayen and John Newman

A04 *Discovering Light Verb Constructions and their Translations from Parallel Corpora without Word Alignment*

Natalie Vargas, Carlos Ramisch and Helena Caseli

A05 *Identification of Multiword Expressions for Latvian and Lithuanian: Hybrid Approach*

Justina Mandravickaite and Tomas Krilavičius

- 14:50–15:25** *Poster Session A (continued)*
- A06 *Show Me Your Variance and I Tell You Who You Are - Deriving Compound Compositionality from Word Alignments*
Fabienne Cap
- A07 *Semantic annotation to characterize contextual variation in terminological noun compounds: a pilot study*
Melania Cabezas-García and Antonio San Martín
- 15:25–16:00** *Poster Session B*
- ST4 *Detection of Verbal Multi-Word Expressions via Conditional Random Fields with Syntactic Dependency Features and Semantic Re-Ranking*
Alfredo Maldonado, Lifeng Han, Erwan Moreau, Ashjan Alsulaimani, Koel Dutta Chowdhury, Carl Vogel and Qun Liu
- ST5 *A data-driven approach to verbal multiword expression detection. PARSEME Shared Task system description paper*
Tiberiu Boroş, Sonia Pipa, Verginica Barbu Mititelu and Dan Tufiş
- ST6 *The ATILF-LLF System for Parseme Shared Task: a Transition-based Verbal Multi-word Expression Tagger*
Hazem Al Saied, Matthieu Constant and Marie Candito
- B08 *Investigating the Opacity of Verb-Noun Multiword Expression Usages in Context*
Shiva Taslimipour, Omid Rohanian, Ruslan Mitkov and Afsaneh Fazly
- B09 *Compositionality in Verb-Particle Constructions*
Archana Bhatia, Choh Man Teng and James Allen
- B10 *Rule-Based Translation of Spanish Verb-Noun Combinations into Basque*
Uxoia Iñurrieta, Itziar Aduriz, Arantza Diaz de Ilarraza, Gorka Labaka and Kepa Sarasola
- B11 *Verb-Particle Constructions in Questions*
Veronika Vincze
- B12 *Simple Compound Splitting for German*
Marion Weller-Di Marco
- B13 *Identification of Ambiguous Multiword Expressions Using Sequence Models and Lexical Resources*
Manon Scholivet and Carlos Ramisch
- B14 *Comparing Recurring Lexico-Syntactic Trees (RLTs) and Ngram Techniques for Extended Phraseology Extraction*
Agnès Tutin and Olivier Kraif
- 16:00–16:30** **COFFEE BREAK**
- Oral Session - Short Papers**
- 16:30–16:50 *Benchmarking Joint Lexical and Syntactic Analysis on Multiword-Rich Data*
Matthieu Constant and Héctor Martínez Alonso
- 16:50–17:10 *Semi-Automated Resolution of Inconsistency for a Harmonized Multiword Expression and Dependency Parse Annotation*
King Chan, Julian Brooke and Timothy Baldwin
- 17:10–17:30 *Combining Linguistic Features for the Detection of Croatian Multiword Expressions*
Maja Buljan and Jan Šnajder
- 17:30–17:50 *Complex Verbs are Different: Exploring the Visual Modality in Multi-Modal Models to Predict Compositionality*
Maximilian Köper and Sabine Schulte im Walde
- 17:50–18:30 *Panel Discussion - Multiword Expressions: Getting the Taste of Things to Come*