

EACL 2014

**14th Conference of the European Chapter of the
Association for Computational Linguistics**



**Proceedings of the 10th Workshop on Multiword Expressions
(MWE 2014)**

26-27 April 2014
Gothenburg, Sweden

©2014 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 978-1-937284-87-9

Introduction

The 10th Workshop on Multiword Expressions (MWE 2014) took place on April 26 and 27, 2014 in Gothenburg, Sweden in conjunction with the 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2014) and was endorsed by the Special Interest Group on the Lexicon of the Association for Computational Linguistics (SIGLEX), as well as SIGLEX's Section dedicated to the study and research of Multiword Expressions (SIGLEX-MWE). Moreover, this year's edition of the MWE workshop was also supported by the IC1207 COST1 action PARSEME2 dedicated to Parsing and Multiword Expressions. This European initiative, which started in 2013, gathers 29 European COST member countries, one COST cooperating state and 3 non-COST institutions from the USA and Brazil. Its objective is to increase and enhance the information and communication technology support of the European multilingual heritage by bringing about a substantial progress in the understanding and modelling of MWEs within advanced multilingual NLP techniques, notably deep parsing. The special track of the MWE 2014 workshop endorsed by PARSEME gathered 8 papers where links between lexical and grammatical aspects of MWEs, as well as their role in deep parsing and NLP applications, such as machine translation, were addressed.

The workshop has been held almost every year since 2003 in conjunction with ACL, EACL, NAACL, COLING and LREC. It provides an important venue for interaction, sharing of resources and tools and collaboration efforts for advancing the computational treatment of Multiword Expressions (MWEs), attracting the attention of an ever-growing community working on a variety of languages and MWE types.

MWEs include idioms (storm in a teacup, sweep under the rug), fixed phrases (in vitro, by and large, rock'n roll), noun compounds (olive oil, laser printer), compound verbs (take a nap, bring about), among others. These, while easily mastered by native speakers, are a key issue and a current weakness for natural language parsing and generation, as well as real-life applications depending on some degree of semantic interpretation, such as machine translation, just to name a prominent one among many. However, thanks to the joint efforts of researchers from several fields working on MWEs, significant progress has been made in recent years, especially concerning the construction of large-scale language resources. For instance, there is a large number of recent papers that focus on acquisition of MWEs from corpora, and others that describe a variety of techniques to find paraphrases for MWEs. Current methods use a plethora of tools such as association measures, machine learning, syntactic patterns, web queries, etc.

In the call for papers we solicited submissions about major challenges in the overall process of MWE treatment, both from the theoretical and the computational viewpoint, focusing on original research related (but not limited) to the following topics:

- Manually and automatically constructed resources
- Representation of MWEs in dictionaries and ontologies
- MWEs and user interaction

- Multilingual acquisition
- Multilingualism and MWE processing
- Models of first and second language acquisition of MWEs
- Crosslinguistic studies on MWEs
- The role of MWEs in the domain adaptation of parsers
- Integration of MWEs into NLP applications
- Evaluation of MWE treatment techniques
- Lexical, syntactic or semantic aspects of MWEs

Submission modalities included Long Papers and Short Papers. From a total of 36 submissions, 14 were long papers and 22 were short papers, and we accepted 6 long papers for oral presentation and 2 as posters. We further accepted 6 short papers for oral presentation and 8 as posters. The overall acceptance rate is 58%. The workshop also featured 3 invited talks.

Acknowledgements

We would like to thank the members of the Program Committee for the timely reviews and the authors for their valuable contributions. We also want to thank PARSEME which funded travel and stay for at least 20 participants from the Action's member countries.

Valia Kordoni, Agata Savary, Markus Egg, Eric Wehrli, Stefan Evert
Co-Organizers



Organizers:

Valia Kordoni, Humboldt Universität zu Berlin (Germany)
Markus Egg, Humboldt Universität zu Berlin (Germany)
Agata Savary, special track organizer, Université François Rabelais Tours (France)
Eric Wehrli, special track organizer, Université de Genève (Switzerland)
Stefan Evert, Friedrich-Alexander-Universität Erlangen-Nürnberg (Germany)

Program Committee:

Iñaki Alegria, University of the Basque Country (Spain)
Dimitra Anastasiou, University of Bremen (Germany)
Doug Arnold, University of Essex (UK)
Eleftherios Avramidis, DFKI GmbH (Germany)
Tim Baldwin, University of Melbourne (Australia)
Núria Bel, Universitat Pompeu Fabra (Spain)
Chris Biemann, Technische Universität Darmstadt (Germany)
Francis Bond, Nanyang Technological University (Singapore)
Lars Borin, University of Gothenburg (Sweden)
António Branco, University of Lisbon (Portugal)
Miriam Butt, Universität Konstanz (Germany)
Aoife Cahill, ETS (USA)
Ken Church, IBM Research (USA)
Matthieu Constant, Université Paris-Est Marne-la-Vallée (France)
Paul Cook, University of Melbourne (Australia)
Béatrice Daille, Nantes University (France)
Koenraad De Smedt, University of Bergen (Norway)
Gaël Dias, University of Caen Basse-Normandie (France)
Gülşen Eryiğit, Istanbul Technical University (Turkey)
Tomaž Erjavec, Jožef Stefan Institute (Slovenia)
Joaquim Ferreira da Silva, New University of Lisbon (Portugal)
Roxana Girju, University of Illinois at Urbana-Champaign (USA)
Chikara Hashimoto, National Institute of Information and Communications Technology (Japan)
Ulrich Heid, Universität Hildesheim (Germany)
Kyo Kageura, University of Tokyo (Japan)
Ioannis Korkontzelos, University of Manchester (UK)
Brigitte Krenn, Austrian Research Institute for Artificial Intelligence (Austria)
Cvetana Krstev, University of Belgrade (Serbia)
Marie-Catherine de Marneffe, The Ohio State University (USA)
Takuya Matsuzaki, National Institute of Informatics (Japan)
Preslav Nakov, Qatar Computing Research Institute (Qatar)
Malvina Nissim, University of Bologna (Italy)

Joakim Nivre, Uppsala University (Sweden)
Diarmuid Ó Séaghdha, University of Cambridge (UK)
Jan Odijk, University of Utrecht (The Netherlands)
Yannick Parmentier, Université d'Orléans (France)
Pavel Pecina, Charles University in Prague (Czech Republic)
Scott Piao, Lancaster University (UK)
Adam Przepiórkowski, Institute of Computer Science, Polish Academy of Sciences (Poland)
Victoria Rosén, University of Bergen (Norway)
Carlos Ramisch, Aix-Marseille University (France)
Manfred Sailer, Goethe University Frankfurt am Main (Germany)
Magali Sanches Duran, University of São Paulo (Brazil)
Violeta Seretan, University of Geneva (Switzerland)
Ekaterina Shutova, University of California, Berkeley (USA)
Jan Šnajder, University of Zagreb (Croatia)
Pavel Straňák, Charles University in Prague (Czech Republic)
Sara Stymne, Uppsala University (Sweden)
Stan Szpakowicz, University of Ottawa (Canada)
Beata Trawinski, Institut für Deutsche Sprache (IDS), Mannheim (Germany)
Yulia Tsvetkov, Carnegie Mellon University (USA)
Yuancheng Tu, Microsoft (USA)
Ruben Urizar, University of the Basque Country (Spain)
Gertjan van Noord, University of Groningen (The Netherlands)
Aline Villavicencio, Federal University of Rio Grande do Sul (Brazil)
Veronika Vincze, Hungarian Academy of Sciences (Hungary)
Martin Volk, University of Zurich (Switzerland)
Tom Wasow, Stanford University (USA)
Shuly Wintner, University of Haifa (Israel)
Dekai Wu, The Hong Kong University of Science & Technology (Hong Kong)

Invited Speakers:

Ekaterina Shutova, ICSI, UC Berkeley (USA)
Preslav Nakov, Qatar Computing Research Institute (Qatar)
One more invited speaker to be confirmed

Table of Contents

<i>Breaking Bad: Extraction of Verb-Particle Constructions from a Parallel Subtitles Corpus</i> Aaron Smith	1
<i>A Supervised Model for Extraction of Multiword Expressions, Based on Statistical Context Features</i> Meghdad Farahmand and Ronaldo Martins	10
<i>VPCTagger: Detecting Verb-Particle Constructions With Syntax-Based Methods</i> István Nagy T. and Veronika Vincze	17
<i>The Relevance of Collocations for Parsing</i> Eric Wehrli	26
<i>Parsing Modern Greek verb MWEs with LFG/XLE grammars</i> Niki Samaridi and Stella Markantonatou	33
<i>Evaluation of a Substitution Method for Idiom Transformation in Statistical Machine Translation</i> Giancarlo Salton, Robert Ross and John Kelleher	38
<i>Encoding MWEs in a conceptual lexicon</i> Aggeliki Fotopoulou, Stella Markantonatou and Voula Giouli	43
<i>German Compounds and Statistical Machine Translation. Can they get along?</i> Carla Parra Escartín, Stephan Peitz and Hermann Ney	48
<i>Extracting MWEs from Italian corpora: A case study for refining the POS-pattern methodology</i> Sara Castagnoli, Malvina Nassim and Francesca Masini	57
<i>Mickey Mouse is not a Phrase: Improving Relevance in E-Commerce with Multiword Expressions</i> Prathyusha Senthil Kumar, Vamsi Salaka, Tracy Holloway King and Brian Johnson	62
<i>Encoding of Compounds in Swedish FrameNet</i> Karin Friberg Heppin and Miriam R L Petruck	67
<i>Extraction of Nominal Multiword Expressions in French</i> Marie Dubremetz and Joakim Nivre	72
<i>Towards an Empirical Subcategorization of Multiword Expressions</i> Luigi Squillante	77
<i>Contexts, Patterns, Interrelations - New Ways of Presenting Multi-word Expressions</i> Kathrin Steyer and Annelen Brunner	82
<i>Detecting change and emergence for multiword expressions</i> Martin Emms and Arun Jayapal	89
<i>An Approach to Take Multi-Word Expressions</i> Claire Bonial, Meredith Green, Jenette Preciado and Martha Palmer	94

<i>Paraphrasing Swedish Compound Nouns in Machine Translation</i>	
Edvin Ullman and Joakim Nivre	99
<i>Feature Norms of German Noun Compounds</i>	
Stephen Roller and Sabine Schulte im Walde	104
<i>Identifying collocations using cross-lingual association measures</i>	
Lis Pereira, Elga Strafella, Kevin Duh and Yuji Matsumoto	109
<i>Unsupervised Construction of a Lexicon and a Repository of Variation Patterns for Arabic Modal Multiword Expressions</i>	
Rania Al-Sabbagh, Roxana Girju and Jana diesner	114
<i>Issues in Translating Verb-Particle Constructions from German to English</i>	
Nina Schottmüller and Joakim Nivre	124

Conference Program

Saturday, April 26, 2014

8:45–9:00 Opening Remarks

Oral Session 1: Detection and Extraction of MWEs

9:00–9:30 *Breaking Bad: Extraction of Verb-Particle Constructions from a Parallel Subtitles Corpus*
Aaron Smith

9:30–10:00 *A Supervised Model for Extraction of Multiword Expressions, Based on Statistical Context Features*
Meghdad Farahmand and Ronaldo Martins

Oral Session 2: PARSEME I – Parsing MWEs

10:00–10:30 *VPCTagger: Detecting Verb-Particle Constructions With Syntax-Based Methods*
István Nagy T. and Veronika Vincze

10:30–11:00 Coffee Break

11:00–12:00 Invited Talk 1: TBA

Oral Session 2: PARSEME I – Parsing MWEs (continued)

12:00–12:30 *The Relevance of Collocations for Parsing*
Eric Wehrli

12:30–14:00 Lunch

Oral Session 3: Short papers – PARSEME II

14:00–14:20 *Parsing Modern Greek verb MWEs with LFG/XLE grammars*
Niki Samaridi and Stella Markantonatou

14:20–14:40 *Evaluation of a Substitution Method for Idiom Transformation in Statistical Machine Translation*
Giancarlo Salton, Robert Ross and John Kelleher

14:40–15:00 *Encoding MWEs in a conceptual lexicon*
Aggeliki Fotopoulou, Stella Markantonatou and Voula Giouli

Saturday, April 26, 2014 (continued)

15:00–15:30 **Poster Booster Session (4 minutes per poster)**

German Compounds and Statistical Machine Translation. Can they get along?

Carla Parra Escartín, Stephan Peitz and Hermann Ney

Extracting MWEs from Italian corpora: A case study for refining the POS-pattern methodology

Sara Castagnoli, Malvina Nissim and Francesca Masini

Mickey Mouse is not a Phrase: Improving Relevance in E-Commerce with Multiword Expressions

Prathyusha Senthil Kumar, Vamsi Salaka, Tracy Holloway King and Brian Johnson

Encoding of Compounds in Swedish FrameNet

Karin Friberg Heppin and Miriam R L Petruck

Extraction of Nominal Multiword Expressions in French

Marie Dubremetz and Joakim Nivre

Towards an Empirical Subcategorization of Multiword Expressions

Luigi Squillante

Contexts, Patterns, Interrelations - New Ways of Presenting Multi-word Expressions

Kathrin Steyer and Annelen Brunner

Detecting change and emergence for multiword expressions

Martin Emms and Arun Jayapal

An Approach to Take Multi-Word Expressions

Claire Bonial, Meredith Green, Jenette Preciado and Martha Palmer

15:30–16:00 Coffee Break

16:00–17:30 Poster Session

Sunday, April 27, 2014

9:30–10:30 Invited Talk 2: TBA

10:30–11:00 Coffee Break

Oral Session 5: Short papers – MWEs in multilingual applications

11:00–11:20 *Paraphrasing Swedish Compound Nouns in Machine Translation*
Edvin Ullman and Joakim Nivre

11:20–11:40 *Feature Norms of German Noun Compounds*
Stephen Roller and Sabine Schulte im Walde

11:40–12:00 *Identifying collocations using cross-lingual association measures*
Lis Pereira, Elga Strafella, Kevin Duh and Yuji Matsumoto

Oral Session 6: Issues in lexicon construction and Machine Translation

12:00–12:30 *Unsupervised Construction of a Lexicon and a Repository of Variation Patterns for Arabic Modal Multiword Expressions*
Rania Al-Sabbagh, Roxana Girju and Jana diesner

12:30–14:00 Lunch

14:00–14:30 *Issues in Translating Verb-Particle Constructions from German to English*
Nina Schottmüller and Joakim Nivre

14:30–15:30 Invited Talk 3: TBA

15:30–15:45 Closing remarks

