

NAACL HLT 2013

**9th Workshop on Multiword Expressions
MWE 2013**

Proceedings of the Workshop

13-14 June 2013
Atlanta, Georgia, USA

©2013 The Association for Computational Linguistics

209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-937284-47-3

Introduction

The *9th Workshop on Multiword Expressions (MWE 2013)*¹ took place on June 13 and 14, 2013 in Atlanta, Georgia, USA in conjunction with the 2013 Conference of the North American Chapter of the (NAACL HLT 2013), and was endorsed by the Special Interest Group on the Lexicon of the Association for Computational Linguistics (SIGLEX)². 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 to the following topics:

- Manually and automatically constructed resources
- Representation of MWEs in dictionaries and ontologies
- MWEs and user interaction
- Multilingual acquisition
- Crosslinguistic studies on MWEs
- Integration of MWEs into NLP applications
- Lexical, syntactic or semantic aspects of MWEs

Submission modalities included Long Papers and Short Papers. From a total of 27 submissions, 15 were long papers and 12 were short papers, and we accepted 7 long papers for oral presentation and 3 as posters: an acceptance rate of 66.6%. We further accepted 5 short papers for oral presentation and 3

¹<http://multiword.sourceforge.net/mwe2013>

²<http://www.siglex.org>

as posters (66.6% acceptance). The workshop also featured 3 invited talks, by Jill Burstein (Educational Testing Service, USA), Malvina Nissim (University of Bologna, Italy) and Martha Palmer (University of Colorado at Boulder, USA).

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 projects CAPES/COFECUB 707/11 Cameleon, CNPq 482520/2012-4, 478222/2011-4, 312184/2012-3 and 551964/2011-1.

Valia Kordoni, Carlos Ramisch, Aline Villavicencio
Co-Organizers

Organizers:

Valia Kordoni, Humboldt-Universität zu Berlin, Germany
Carlos Ramisch, Joseph Fourier University, France
Aline Villavicencio, Federal University of Rio Grande do Sul, Brazil

Program Committee:

Iñaki Alegria, University of the Basque Country, Spain
Dimitra Anastasiou, University of Bremen, Germany
Doug Arnold, University of Essex, UK
Giuseppe Attardi, Università di Pisa, Italy
Eleftherios Avramidis, DFKI GmbH, Germany
Timothy Baldwin, The University of Melbourne, Australia
Chris Biemann, Technische Universität Darmstadt, Germany
Francis Bond, Nanyang Technological University, Singapore
Antonio Branco, University of Lisbon, Portugal
Aoife Cahill, Educational Testing Service, USA
Helena Caseli, Federal University of São Carlos, Brazil
Ken Church, IBM Research, USA
Matthieu Constant, Université Paris-Est Marne-la-Vallée, France
Paul Cook, The University of Melbourne, Australia
Béatrice Daille, Nantes University, France
Koenraad de Smedt, University of Bergen, Norway
Markus Egg, Humboldt-Universität zu Berlin, Germany
Stefan Evert, Friedrich-Alexander Universität Erlangen-Nürnberg, Germany
Afsaneh Fazly, University of Toronto, Canada
Joaquim Ferreira da Silva, New University of Lisbon, Portugal
Chikara Hashimoto, National Institute of Information and Communications Technology, Japan
Kyo Kageura, University of Tokyo, Japan
Su Nam Kim, Monash University, Australia
Ioannis Korkontzelos, University of Manchester, UK
Brigitte Krenn, Austrian Research Institute for Artificial Intelligence, Austria
Evita Linardaki, Hellenic Open University, Greece
Takuya Matsuzaki, National Institute of Informatics, Japan
Yusuke Miyao, National Institute of Informatics, Japan
Preslav Nakov, Qatar Computing Research Institute - Qatar Foundation, Qatar
Joakim Nivre, Uppsala University, Sweden
Diarmuid Ó Séaghdha, University of Cambridge, UK
Jan Odijk, Utrecht University, The Netherlands
Yannick Parmentier, Université d'Orléans, France
Pavel Pecina, Charles University Prague, Czech Republic
Scott Piao, Lancaster University, UK

Adam Przepiórkowski, Polish Academy of Sciences, Poland
Magali Sanches Duran, University of São Paulo, Brazil
Agata Savary, Université François Rabelais Tours, France
Ekaterina Shutova, University of California at Berkeley, USA
Mark Steedman, University of Edinburgh, UK
Sara Stymne, Uppsalla University, Sweden
Stan Szpakowicz, University of Ottawa, Canada
Beata Trawinski, University of Vienna, Austria
Yulia Tsvetkov, Carnegie Mellon University, USA
Yuancheng Tu, Microsoft, USA
Kyrioko Uchiyama, National Institute of Informatics, Japan
Ruben Urizar, University of the Basque Country, Spain
Tony Veale, University College Dublin, Ireland
David Vilar, DFKI GmbH, Germany
Veronika Vincze, Hungarian Academy of Sciences, Hungary
Tom Wasow, Stanford University, USA
Eric Wehrli, University of Geneva, Switzerland

Additional Reviewers:

Silvana Hartmann, Technische Universität Darmstadt, Germany
Bahar Salehi, The University of Melbourne, Australia

Invited Speakers:

Jill Burstein, Educational Testing Service, USA
Malvina Nissim, University of Bologna, Italy
Martha Palmer, University of Colorado at Boulder, USA

Table of Contents

<i>Managing Multiword Expressions in a Lexicon-Based Sentiment Analysis System for Spanish</i> Antonio Moreno-Ortiz, Chantal Perez-Hernandez and Maria Del-Olmo	1
<i>Introducing PersPred, a Syntactic and Semantic Database for Persian Complex Predicates</i> Pollet Samvelian and Pegah Faghiri	11
<i>Improving Word Translation Disambiguation by Capturing Multiword Expressions with Dictionaries</i> Lars Bungum, Björn Gambäck, André Lynum and Erwin Marsi	21
<i>Complex Predicates are Multi-Word Expressions</i> Martha Palmer	31
<i>The (Un)expected Effects of Applying Standard Cleansing Models to Human Ratings on Compositionality</i> Stephen Roller, Sabine Schulte im Walde and Silke Scheible	32
<i>Determining Compositionality of Word Expressions Using Word Space Models</i> Lubomír Krčmář, Karel Ježek and Pavel Pecina	42
<i>Modelling the Internal Variability of MWEs</i> Malvina Nissim	51
<i>Automatically Assessing Whether a Text Is Cliched, with Applications to Literary Analysis</i> Paul Cook and Graeme Hirst	52
<i>An Analysis of Annotation of Verb-Noun Idiomatic Combinations in a Parallel Dependency Corpus</i> Zdenka Uresova, Jan Hajic, Eva Fucikova and Jana Sindlerova	58
<i>Automatic Identification of Bengali Noun-Noun Compounds Using Random Forest</i> Vivekananda Gayen and Kamal Sarkar	64
<i>Automatic Detection of Stable Grammatical Features in N-Grams</i> Mikhail Kopotev, Lidia Pivovarova, Natalia Kochetkova and Roman Yangarber	73
<i>Exploring MWEs for Knowledge Acquisition from Corporate Technical Documents</i> Bell Manrique-Losada, Carlos M. Zapata-Jaramillo and Diego A. Burgos	82
<i>MWE in Portuguese: Proposal for a Typology for Annotation in Running Text</i> Sandra Antunes and Amália Mendes	87
<i>Identifying Pronominal Verbs: Towards Automatic Disambiguation of the Clitic 'se' in Portuguese</i> Magali Sanches Duran, Carolina Evaristo Scarton, Sandra Maria Aluísio and Carlos Ramisch ..	93
<i>A Repository of Variation Patterns for Multiword Expressions</i> Malvina Nissim and Andrea Zaninello	101

<i>Syntactic Identification of Occurrences of Multiword Expressions in Text using a Lexicon with Dependency Structures</i>	
Eduard Bejček, Pavel Straňák and Pavel Pecina	106
<i>Combining Different Features of Idiomaticity for the Automatic Classification of Noun+Verb Expressions in Basque</i>	
Antton Gurrutxaga and Iñaki Alegria	116
<i>Semantic Roles for Nominal Predicates: Building a Lexical Resource</i>	
Ashwini Vaidya, Martha Palmer and Bhuvana Narasimhan	126
<i>Constructional Intensifying Adjectives in Italian</i>	
Sara Berlanda	132
<i>The Far Reach of Multiword Expressions in Educational Technology</i>	
Jill Burstein	138
<i>Construction of English MWE Dictionary and its Application to POS Tagging</i>	
Yutaro Shigeto, Ai Azuma, Sorami Hisamoto, Shuhei Kondo, Tomoya Kouse, Keisuke Sakaguchi, Akifumi Yoshimoto, Frances Yung and Yuji Matsumoto	139

Conference Program

Thursday, June 13 – Morning

09:00-09:15 Opening Remarks

Oral Session 1: Resources and Applications

09:15–09:40 *Managing Multiword Expressions in a Lexicon-Based Sentiment Analysis System for Spanish*

Antonio Moreno-Ortiz, Chantal Perez-Hernandez and Maria Del-Olmo

09:40–10:05 *Introducing PersPred, a Syntactic and Semantic Database for Persian Complex Predicates*

Pollet Samvelian and Pegah Faghiri

10:05–10:30 *Improving Word Translation Disambiguation by Capturing Multiword Expressions with Dictionaries*

Lars Bungum, Björn Gambäck, André Lynum and Erwin Marsi

10:30-11:00 **COFFEE BREAK**

11:00–12:00 **Invited Talk 1**

Complex Predicates are Multi-Word Expressions

Martha Palmer

Oral Session 2: Compositionality

12:00–12:25 *The (Un)expected Effects of Applying Standard Cleansing Models to Human Ratings on Compositionality*

Stephen Roller, Sabine Schulte im Walde and Silke Scheible

12:30-14:00 **LUNCH BREAK**

Thursday, June 13 – Afternoon

Oral Session 2: Compositionality (contd.)

14:05–14:30 *Determining Compositionality of Word Expressions Using Word Space Models*
Lubomír Krčmář, Karel Ježek and Pavel Pecina

14:30–15:30 Invited Talk 2

Modelling the Internal Variability of MWEs
Malvina Nissim

15:30-16:00 COFFEE BREAK**Oral Session 3: Short Papers**

16:00–16:15 *Automatically Assessing Whether a Text Is Cliched, with Applications to Literary Analysis*
Paul Cook and Graeme Hirst

16:15–16:30 *An Analysis of Annotation of Verb-Noun Idiomatic Combinations in a Parallel Dependency Corpus*
Zdenka Uresova, Jan Hajic, Eva Fucikova and Jana Sindlerova

16:30-17:40 Poster Session

16:30-16:40 Poster Boosters

Automatic Identification of Bengali Noun-Noun Compounds Using Random Forest
Vivekananda Gayen and Kamal Sarkar

Automatic Detection of Stable Grammatical Features in N-Grams

Mikhail Kopotev, Lidia Pivovarova, Natalia Kochetkova and Roman Yangarber

Exploring MWEs for Knowledge Acquisition from Corporate Technical Documents
Bell Manrique-Losada, Carlos M. Zapata-Jaramillo and Diego A. Burgos

MWE in Portuguese: Proposal for a Typology for Annotation in Running Text
Sandra Antunes and Amália Mendes

Identifying Pronominal Verbs: Towards Automatic Disambiguation of the Clitic 'se' in Portuguese

Magali Sanches Duran, Carolina Evaristo Scarton, Sandra Maria Aluísio and Carlos Ramisch

A Repository of Variation Patterns for Multiword Expressions
Malvina Nissim and Andrea Zaninello

Friday, June 14 – Morning

Oral Session 4: Identification and Classification

09:10–09:35 *Syntactic Identification of Occurrences of Multiword Expressions in Text using a Lexicon with Dependency Structures*

Eduard Bejček, Pavel Straňák and Pavel Pecina

09:35–10:00 *Combining Different Features of Idiomaticity for the Automatic Classification of Noun+Verb Expressions in Basque*

Antton Gurrutxaga and Iñaki Alegria

Oral Session 5: Short Papers

10:00–10:15 *Semantic Roles for Nominal Predicates: Building a Lexical Resource*

Ashwini Vaidya, Martha Palmer and Bhuvana Narasimhan

10:15–10:30 *Constructional Intensifying Adjectives in Italian*

Sara Berlanda

10:30-11:00 **COFFEE BREAK**

11:00–12:00 **Invited Talk 3**

The Far Reach of Multiword Expressions in Educational Technology

Jill Burstein

Oral Session 5: Short Papers (contd.)

12:00–12:15 *Construction of English MWE Dictionary and its Application to POS Tagging*

Yutaro Shigeto, Ai Azuma, Sorami Hisamoto, Shuhei Kondo, Tomoya Kouse, Keisuke Sakaguchi, Akifumi Yoshimoto, Frances Yung and Yuji Matsumoto

12:15-12:30 Closing Remarks

