NAACL HLT 2015

11th Workshop on Multiword Expressions
MWE 2014

Proceedings of the Workshop

June 4, 2015
Denver, Colorado, USA

Order print-on-demand copies from:

Curran Associates
57 Morehouse Lane
Red Hook, New York 12571
USA
Tel: +1-845-758-0400
Fax: +1-845-758-2633
curran@proceedings.com

ISBN 978-1-941643-38-9

Introduction

The 11th Workshop on Multiword Expressions (MWE 2015) took place on June 4, 2015 in Denver, Colorado, USA, in conjunction with the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015) 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).

The workshop has been held almost every year since 2003 in conjunction with ACL, EACL, NAACL, COLING and LREC. By now, it provides the main venue of the field 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 from all around the world 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), 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:

- Lexicon-grammar interface for MWEs
- Parsing techniques for MWEs
- Hybrid parsing of MWEs
- Annotating MWEs in treebanks
- MWEs in Machine Translation and Translation Technology
- 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 27 submissions, of which 14 were long papers and 13 were short papers, we accepted 5 long papers for oral presentation and 3 as posters. We further accepted 3 short papers for oral presentation and 3 as posters. The overall acceptance rate is 52%.

The workshop also featured an invited talk by Paul Kay (International Computer Science Institute, UC Berkeley) and Laura A. Michaelis (Department of Linguistics and Institute of Cognitive Science, University of Colorado Boulder) on "How Constructions Mean".

Acknowledgements

We would like to thank the members of the Program Committee for the timely reviews and the authors for their valuable contributions.

Valia Kordoni, Kostadin Cholakov, Markus Egg, Stella Markantonatou, Shuly Wintner
Co-Organizers

Organizers:

Valia Kordoni, Humboldt Universität zu Berlin (Germany)
Kostadin Cholakov, Humboldt Universität zu Berlin (Germany)
Markus Egg, Humboldt Universität zu Berlin (Germany)
Stella Markantonatou, Institute for Language and Speech Processing (ILSP) - Athena Research Center (Greece)
Shuly Wintner, University of Haifa (Israel)

Program Committee:

Dimitra Anastasiou, University of Bremen (Germany)
Eleftherios Avramidis, DFKI GmbH (Germany)
Tim Baldwin, University of Melbourne (Australia)
Núria Bel, Pompeu Fabra University (Spain)
Lars Borin, University of Gothenburg (Sweden)
Jill Burstein, ETS (USA)
Aoife Cahill, ETS (USA)
Helena Caseli, Federal University of Sao Carlos (Brazil)
Ken Church, IBM Research (USA)
Paul Cook, University of New Brunswick (Canada)
Béatrice Daille, Nantes University (France)
Gaël Dias, University of Caen Basse-Normandie (France)
Stefan Evert, Friedrich-Alexander-Universität Erlangen-Nürnberg (Germany)
Roxana Girju, University of Illinois at Urbana-Champaign (USA)
Ed Hovy, Carnegie Mellon University (USA)
Kyo Kageura, University of Tokyo (Japan)
Su Nam Kim, Monash University (Australia)
Dimitrios Kokkinakis, University of Gothenburg (Sweden)
Ioannis Korkontzelos, University of Manchester (UK)
Lori Levin, Carnegie Mellon University (USA)
Patricia Lichtenstein, University of California, Merced (USA)
Marie-Catherine de Marneffe, The Ohio State University (USA)
Takuya Matsuzaki, Nagoya University (Japan)
Yusuke Miyao, National Institute of Informatics (Japan)
Preslav Nakov, Qatar Computing Research Institute - Qatar Foundation (Qatar)
Malvina Nissim, University of Bologna (Italy)
Joakim Nivre, University of Uppsala (Sweden)
Diarmuid Ó Séaghdha, University of Cambridge and VocalIQ (UK)
Jan Odijk, University of Utrecht (The Netherlands)
Yannick Parmentier, Universite d'Orleans (France)
Pavel Pecina, Charles University Prague (Czech Republic)

Scott Piao, Lancaster University (UK)
Barbara Plank, University of Copenhagen (Denmark)
Carlos Ramisch, Aix-Marseille University (France)
Martin Riedl, University of Darmstadt (Germany)
Will Roberts, Humboldt University Berlin (Germany)
Agata Savary, Université François Rabelais Tours (France)
Violeta Seretan, University of Geneva (Switzerland)
Ekaterina Shutova, University of California, Berkeley (USA)
Beata Trawinski, IDS Mannheim (Germany)
Yulia Tsvetkov, Carnegie Mellon University (USA)
Yuancheng Tu, Microsoft (USA)
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)
Eric Wehrli, University of Geneva (Switzerland)

Invited Speaker:

Laura A. Michaelis

Table of Contents

Conference Program

Thursday, June 4, 2014

Oral Session 1

09:00–09:30 *A Method of Accounting Bigrams in Topic Models*
Michael Nokel and Natalia Loukachevitch

09:30–10:00 *Multiword Expression Identification with Recurring Tree Fragments and Association Measures*
Federico Sangati and Andreas van Cranenburgh

10:00–10:30 *How to Account for Idiomatic German Support Verb Constructions in Statistical Machine Translation*
Fabienne Cap, Manju Nirmal, Marion Weller and Sabine Schulte im Walde

10:30–11:00 ***Coffee Break***

Oral Session 2

11:00–11:20 *A Multiword Expression Data Set: Annotating Non-Compositionality and Conventionalization for English Noun Compounds*
Meghdad Farahmand, Aaron Smith and Joakim Nivre

11:20–11:40 *Modeling the Statistical Idiosyncrasy of Multiword Expressions*
Meghdad Farahmand and Joakim Nivre

11:40–12:00 *Clustering-based Approach to Multiword Expression Extraction and Ranking*
Elena Tutubalina

Thursday, June 4, 2014 (continued)

Invited Talk by Laura A. Michaelis

12:00–13:00 *How Constructions Mean*
Paul Kay and Laura A. Michaelis

13:00–14:00 ***Lunch***

14:00–14:30 ***Poster Booster Session (5 minutes per poster)***

Never-Ending Multiword Expressions Learning
Alexandre Rondon, Helena Caseli and Carlos Ramisch

The Impact of Multiword Expression Compositionality on Machine Translation Evaluation
Bahar Salehi, Nitika Mathur, Paul Cook and Timothy Baldwin

The Bare Necessities: Increasing Lexical Coverage for Multi-Word Domain Terms with Less Lexical Data
Branimir Boguraev, Esme Manandise and Benjamin Segal

Phrase translation using a bilingual dictionary and n-gram data: A case study from Vietnamese to English
Khang Nhut Lam, Feras Al Tarouti and Jugal Kalita

Annotation and Extraction of Multiword Expressions in Turkish Treebanks
Gülşen Eryiğit, Kübra ADALI, Dilara Torunoğlu-Selamet, Umut Sulubacak and Tuğba Pamay

Event Categorization beyond Verb Senses
Aron Marvel and Jean-Pierre Koenig

14:30–15:30 ***Poster Session***

15:30–16:00 ***Coffee Break***

Thursday, June 4, 2014 (continued)

Oral Session 3

16:00–16:30 *Muddying The Multiword Expression Waters: How Cognitive Demand Affects Multiword Expression Production*
Adam Goodkind and Andrew Rosenberg

16:30–17:00 *Building a Lexicon of Formulaic Language for Language Learners*
Julian Brooke, Adam Hammond, David Jacob, Vivian Tsang, Graeme Hirst and Fraser Shein