2006

COLING • ACL

COLING • ACL 2006

Multiword Expressions: Identifying and Exploiting Underlying Properties

Proceedings of the Workshop

Chairs:
Begoña Villada Moirón, Aline Villavicencio,
Diana McCarthy, Stefan Evert and Suzanne Stevenson

23 July 2006
Sydney, Australia

Production and Manufacturing by
BPA Digital
11 Evans St
Burwood VIC 3125
AUSTRALIA

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 1-932432-84-1

Table of Contents

Preface

This volume contains the papers accepted for presentation at the *Workshop on Multiword Expressions: Identifying and Exploiting Underlying Properties*. The workshop is endorsed by the Association for Computational Linguistics Special Interest Group on the Lexicon (SIGLEX) and is hosted in conjunction with the COLING/ACL 2006 on July 23rd, 2006 in Sydney, Australia.

There has been a growing awareness in the NLP community of the problems that multiword expressions (MWEs) pose. Developments in areas such as machine translation, text summarization, paraphrasing, grammar development and parsing, information retrieval, and question answering (to mention a few) have acknowledged difficulties due to the idiosyncratic nature of multiword expressions. This workshop continues a tradition of ACL workshops on Collocations (2001) and Multiword Expressions (2003 and 2004). Its specific objective is to focus on the underlying properties of MWEs. The call for papers expressed our interest in several topics such as the definition of MWEs, properties of MWEs and their impact on NLP applications, representation and treatment of the different classes of MWEs, linguistic and psycholinguistic analyses of MWEs, evaluation of extraction techniques and the importance of (non-)compositionality.

We received 23 submissions in total. Each submission was reviewed by (at least) three members of the program committee who not only judged each submission but also gave detailed comments to the authors. Among the received papers, 10 were selected for presentation at the workshop. After 3 papers have been withdrawn by their authors, seven papers are included in these proceedings.

The intention of this workshop is to focus on some fundamental questions on the nature of MWEs. To do this we will allow plenty of time for discussion to pursue some of the interesting, open and difficult questions that MWEs raise. As well as a discussion period after each session of papers, we will be organising group discussions at the end of the workshop. These will focus on problems of defining, characterising and evaluating MWEs, given what we know about the range of phenomena that they encompass as well as any important questions that have arisen during the workshop.

We would like to thank all the authors for submitting their research and the members of the program committee for their careful reviews and useful suggestions to the authors. We are indebted to Timothy Baldwin who will give an invited talk at the workshop. We would also like to thank the COLING/ACL 2006 organising committee that made this workshop possible and SIGLEX for agreeing to endorse this workshop. Finally, we hope that this workshop will provide food for thought for all participants.

Begoña Villada Moirón
Aline Villavicencio
Diana McCarthy
Stefan Evert
Suzanne Stevenson
June 2006

Organizers

Chairs:

Begoña Villada Moirón, University of Groningen (The Netherlands)
Aline Villavicencio, Federal University of Rio Grande do Sul (Brazil)
Diana McCarthy, University of Sussex (UK)
Stefan Evert, University of Osnabrueck (Germany)
Suzanne Stevenson, University of Toronto (Canada)

Program Committee:

Timothy Baldwin, Stanford University (USA); Melbourne University (Australia)
Colin Bannard, University of Edinburgh (UK)
Francis Bond, NTT Communication Science Laboratories (Japan)
Gosse Bouma, University of Groningen (The Netherlands)
Beatrice Daille, Nantes University (France)
Gael Dias, Beira Interior University (Portugal)
James Dowdall, University of Sussex (UK)
Afsaneh Fazly, University of Toronto (Canada)
Christiane Fellbaum, Princeton University (USA)
Nicole Gregoire, Utrecht University (The Netherlands)
Matthew Hurst, Inteliseek (USA)
Nancy Ide, Vassar College (USA)
Aravind Joshi, University of Pennsylvania (USA)
Kyo Kageura, National Institute of Informatics (Japan)
Anna Korhonen, University of Cambridge (UK)
Brigitte Krenn, OFAI, Vienna (Austria)
Mirella Lapata, University of Edinburgh (UK)
Roger Levy, University of Edinburgh (UK)
Rosamund Moon, University of Birmingham (UK)
Stephan Oepen, Stanford University (USA); University of Oslo (Norway)
Kentaro Ogura, NTT Cyber Space Laboratories (Japan)
Darren Pearce, London Knowledge Lab, (UK)
Scott Piao, University of Lancaster (UK)
Ivan Sag, University of Stanford (USA)
Violeta Seretan, University of Geneva (Switzerland)
Beata Trawinski, University of Tuebingen (Germany)
Kiyoko Uchiyama, Keio University (Japan)
Tom Wasow, Stanford University (USA)
Annie Zaenen, PARC (USA)

Invited Speaker:

Timothy Baldwin, Melbourne University (Australia)

Workshop Program

Sunday, 23 July 2006

9:15–9:30 Opening Remarks

9:30–10:30 *Compositionality and Multiword Expressions: Six of One, Half a Dozen of the Other?*
Timothy Baldwin (*Invited Speaker*)

10:30–11:00 Coffee break

Session 1: Compositionality and its Applications

11:00–11:30 *Measuring MWE Compositionality Using Semantic Annotation*
Scott S.L. Piao, Paul Rayson, Olga Mudraya, Andrew Wilson and Roger Garside

11:30–12:00 *Automatic Identification of Non-Compositional Multi-Word Expressions using Latent Semantic Analysis*
Graham Katz and Eugenie Giesbrecht

12:00–12:30 *Using Information about Multi-word Expressions for the Word-Alignment Task*
Sriram Venkatapathy and Aravind K. Joshi

12:30–12:45 Discussion

12:45–14:15 Lunch break

Session 2: Identification

14:15–14:45 *Detecting Complex Predicates in Hindi using POS Projection across Parallel Corpora*
Amitabha Mukerjee, Ankit Soni and Achla M Raina

14:45–15:15 *Automated Multiword Expression Prediction for Grammar Engineering*
Yi Zhang, Valia Kordoni, Aline Villavicencio and Marco Idiart

15:15–15:30 Discussion

15:30–16:00 Coffee break

Sunday, 23 July 2006 (continued)

Session 3: Classes and Underlying Semantics

16:00–16:30 *Classifying Particle Semantics in English Verb-Particle Constructions*
Paul Cook and Suzanne Stevenson

16:30–17:00 *Interpretation of Compound Nominalisations using Corpus and Web Statistics*
Jeremy Nicholson and Timothy Baldwin

17:00–17:15 Discussion

17:15–17:45 Group discussions

17:45–18:15 Summaries from group discussions

18:15 Closing Remarks