Second ACL Workshop on Multiword Expressions: Integrating Processing

Proceedings of the Workshop

Workshop Chairs:
Takaaki Tanaka
Aline Villavicencio
Francis Bond
Anna Korhonen

26 July 2004
Barcelona, Spain

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
73 Landmark Center
East Stroudsburg, PA 18301
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ACL 2004 Workshop on Multiword Expressions: Integrating Processing, 26 July 2004

CO-CHAIRS:

Takaaki Tanaka (NTT Communication Science Laboratories, Japan)
Aline Villavicencio (University of Essex, UK; University of Cambridge, UK)
Francis Bond (NTT Communication Science Laboratories, Japan)
Anna Korhonen (University of Cambridge, UK)

PROGRAM COMMITTEE:

Timothy Baldwin (Stanford University, USA)
Colin Bannard (University of Edinburgh, UK)
Gael Dias (Beira Interior University, Portugal)
James Dowdall (University of Zurich, Switzerland)
Dan Flickinger (Stanford University, USA)
Matthew Hurst (Intelliseek, USA)
Stephan Oepen (Stanford University, USA; University of Oslo, Norway)
Kyonghee Paik (ATR Spoken Language Translation Research Laboratories, Japan)
Scott Piao (University of Lancaster, UK)
Beata Trawinski (University of Tübingen, Germany)
Kiyoko Uchiyama (Keio University, Japan)

CONFERENCE WEBSITE:

`http://www.cl.cam.ac.uk/users/alk23/mwe04/mwe.html`

Preface

This is the proceedings of the second ACL workshop on multiword expressions (MWEs). MWEs are increasingly being singled out as a problem for NLP, particularly for the many applications which require some degree of semantic interpretation and require tasks such as parsing and word sense disambiguation. In the call for papers we solicited papers that especially laid emphasis on integrating analysis, acquisition and treatment of various kinds of multiword expressions in natural language NLP. For example, research that combines a linguistic analysis with a method of automatically acquiring the classes described, work that combines the computational treatment of a class of MWEs with a solid linguistic analysis and research that extracts MWEs and either classifies them or uses them in some task.

We received 23 submissions (3 from Asia, 11 from Europe and 9 from the Americas), and accepted 11 of them for presentation, with two reserves. Each submission was reviewed by three members of the program committee, who not only judged each submission but also gave detailed comments to the authors. The overall quality of submissions was high, making the final selection very difficult. The papers in these proceedings are those which were finally selected for presentation. Many of the papers deal with MWEs in general, rather than aiming at specific subtypes, with examples from a wide range of languages (Basque, English, Japanese, Portuguese, Russian and Turkish). There were also a variety of formalisms considered (dependency grammar, finite state machines, lexical conceptual structure, HPSG, ...) as well as more decriptive papers. The main applications targeted were machine translation and information retrieval.

We would like to thank all the authors who submitted papers. We also thank all the members of the program committee for their time and effort in ensuring that the papers were fairly assessed.

The workshop was supported by

- Research Collaboration between NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation and CSLI, Stanford University
- UK EPSRC project GR/N36493 "Robust Accurate Statistical Parsing (RASP)"

Finally, we wish to thank the organizers of the main conference, in particular the conference workshop co-chairs, Srinivas Bangalore, Christopher Manning, Helen Meng and Marcello Federico.

Takaaki Tanaka, Aline Villavicencio, Francis Bond, Anna Korhonen

June 2004

Table of Contents

Technical Program Schedule

Monday, May 2

9:30-9:35 Welcome

9:35-10:00 *Statistical Measures of the Semi-Productivity of Light Verb Constructions*
Suzanne Stevenson, Afsaneh Fazly and Ryan North

10:00-10:25 *Paraphrasing of Japanese Light-verb Constructions Based on Lexical Conceptual Structure*
Atsushi Fujita, Kentaro Furihata, Kentaro Inui, Yuji Matsumoto and Koichi Takeuchi

10:25-10:50 *What is at Stake: a Case Study of Russian Expressions Starting with a Preposition*
Serge Sharoff

10:50-11:20 Break

11:20-11:45 *Translation by Machine of Complex Nominals: Getting it Right*
Timothy Baldwin and Takaaki Tanaka

11:45-12:10 *MWEs as Non-propositional Content Indicators*
Kosho Shudo, Toshifumi Tanabe, Masahito Takahashi and Kenji Yoshimura

12:10-12:35 *Multiword Expression Filtering for Building Knowledge Maps*
Shailaja Venkatsubramanyan and Jose Perez-Carballo

12:35-14:00 Lunch

14:00-14:25 *Representation and Treatment of Multiword Expressions in Basque*
Iñaki Alegria, Olatz Ansa, Xabier Artola, Nerea Ezeiza, Koldo Gojenola and Ruben Urizar

14:25-14:50 *Multiword Expressions as Dependency Subgraphs*
Ralph Debusmann

14:50-15:15 *Integrating Morphology with Multi-word Expression Processing in Turkish*
Kemal Oflazer, Özlem Çetinoğlu and Bilge Say

15:15-15:45 Break

15:45-16:10 *Frozen Sentences of Portuguese: Formal Descriptions for NLP*
Jorge Baptista, Anabela Correia and Graça Fernandes

16:10-16:35 *Lexical Encoding of MWEs*
Aline Villavicencio, Ann Copestake, Benjamin Waldron and Fabre Lambeau

THIS IS A BLANK PAGE PLEASE IGNORE

Author Index