EACL-2006

11th Conference
of the European Chapter of the
Association for Computational Linguistics

Proceedings of the workshop on

Multi-word-expressions in a multilingual context

April, 3rd, 2006
Trento, Italy

The conference, the workshop and the tutorials are sponsored by:

Center for the Evaluation of Language and Communication Technologies

Celct
c/o BIC, Via dei Solteri, 38
38100 Trento, Italy
http://www.celct.it

Xerox Research Centre Europe
6 Chemin de Maupertuis
38240 Meylan, France
http://www.xrce.xerox.com

CELI s.r.l.
Corso Moncalieri, 21
10131 Torino, Italy
http://www.celi.it

Thales
45 rue de Villiers
92526 Neuilly-sur-Seine Cedex, France
http://www.thalesgroup.com

EACL-2006 is supported by

Trentino S.p.a.  and Metalsistem Group 

Order copies of ACL proceedings from:
Priscilla Rasmussen,
Association for Computational Linguistics (ACL),
3 Landmark Center,
East Stroudsburg, PA 18301 USA

Phone +1-570-476-8006
Fax +1-570-476-0860
E-mail: acl@aclweb.org
On-line order form: http://www.aclweb.org/

PREFACE

This volume contains the ten papers accepted for presentation at Multi-word-expressions in a multilingual context, an EACL 2006 workshop held on April 3rd, 2006, preceding the 11th Conference of the European Chapter of the Association for Computational Linguistics, taking place in Trento, Italy.

For many years, interest in the natural language processing community of the problems that multiword-expressions (MWE) posed was focussed mainly on English. Recently, for example at the ACL2004 workshop on multiword expressions, attention has begun to expand to other languages such as Japanese, Russian, Basque and Turkish. This necessitates a re-evaluation of earlier rule-based, statistical and hybrid techniques for MWE identification and classification. In English, MWE types such as phrasal verbs, noun phrases, proper names, and true non-compositional idioms, are considered. However, in other languages MWE types can be represented as single words, e.g. phrasal verbs in English are generally expressed as verbs with a prefix in Russian. At the same time, research on MWEs for languages other than English is confronted with new problems, such as the number of word forms per lemma or free word order. In the call for papers for this workshop, we invited submissions incorporating the requirements from different areas such as translation, language engineering and those studying computational techniques for the processing of MWE of language learners and how all these requirements differ across languages. This had a deliberately wide scope to enable cross-disciplinary contact between descriptive, contrastive, educational and computational approaches. Of the 23 papers submitted, we accepted 10 for presentation. Each submission was reviewed by at least two members of the programme committee, who gave detailed comments to the authors. The papers presented here deal with a number of themes, from translation, extraction of MWEs, language description and modelling of dictionaries and lexicons.

We gratefully acknowledge the assistance of members of the programme committee and the additional reviewer for performing their task within such a tight schedule. We also acknowledge the support of UK-EPSRC project EP/C004574/1 “Automated Semantic Assistance for Translators (ASSIST)”. Finally, we wish to thank the organisers of the main conference, in particular the conference workshop co-chairs, Maarten de Rijke and Caroline Sporleder.

Paul Rayson
Serge Sharoff
Svenja Adolphs

February 2006

WORKSHOP ORGANISERS

Paul Rayson	Lancaster University, UK
Serge Sharoff	University of Leeds, UK
Svenja Adolphs	University of Nottingham, UK

PROGRAMME COMMITTEE

Dawn Archer	University of Central Lancashire, UK
Timothy Baldwin	University of Melbourne, Australia
Francis Bond	NTT Communication Science Laboratories, Japan
Key-Sun Choi	KAIST, Korea
Béatrice Daille	University of Nantes, France
Sylviane Granger	Université catholique de Louvain, Belgium
Chikara Hashimoto	Kyoto University, Japan
Ulrich Heid	Universität Stuttgart, Germany
Laura Löfberg	University of Tampere, Finland
Anke Lüdeling	Humboldt-Universität zu Berlin, Germany
Olga Mudraya	Lancaster University, UK
Kyonghee Paik	ATR Spoken Language Translation Research Laboratories, Japan
Scott Piao	Lancaster University, UK
Norbert Schmitt	University of Nottingham, UK

ADDITIONAL REVIEWER

Andrew Hardie	Lancaster University, UK

WEBSITE

`http://ucrel.lancs.ac.uk/EACL06MWEmc/`

WORKSHOP PROGRAMME

MORNING:

9.00	Arrivals and welcome Workshop co-chairs
9.30	*Named Entities Translation Based on Comparable Corpora* Iñaki Alegria, Nerea Ezeiza, and Izaskun Fernandez
10.00	*Grouping Multi-word Expressions According to Part-Of-Speech in Statistical Machine Translation* Patrik Lambert and Rafael Banchs
10.30	COFFEE BREAK
11.00	*Automatic Extraction of Chinese Multiword Expressions with a Statistical Tool* Scott S.L. Piao, Guangfan Sun, Paul Rayson, and Qi Yuan
11.30	*Chunking Japanese Compound Functional Expressions by Machine Learning* Masatoshi Tsuchiya, Takao Shime, Toshihiro Takagi, Takehito Utsuro, Kiyotaka Uchimoto, Suguru Matsuyoshi, Satoshi Sato and Seiichi Nakagawa
12.00	*Identifying idiomatic expressions using automatic word-alignment* Begoña Villada Moirón and Jörg Tiedemann

LUNCH BREAK:

12.30 - 14.30

AFTERNOON:

14.30	*Collocation Extraction: Needs, Feeds and Results of an Extraction System for German* Julia Ritz
15.00	*Extending corpus-based identification of light verb constructions using a supervised learning framework* Yee Fan Tan, Min-Yen Kan and Hang Cui
15.30	*Multi-word verbs in a flective language: the case of Estonian* Heiki-Jaan Kaalep and Kadri Muischnek
16.00	COFFEE BREAK
16.30	*Modeling Monolingual and Bilingual Collocation Dictionaries in Description Logics* Dennis Spohr and Ulrich Heid
17.00	*Multiword Units in an MT Lexicon* Tamás Váradi
17.30	Closing discussion

TABLE OF CONTENTS