Call for papers

Multiword expressions: hard going or plain sailing?

Special issue of the International Journal of Language Resources and Evaluation http://www.springer.com/journal/10579/

Guest editors:

Paul Rayson (Lancaster University, UK)
Begoña Villada Moirón (University of Groningen, The Netherlands)
Serge Sharoff (University of Leeds, UK)
Scott Piao (University of Manchester, UK)
Stefan Evert (University of Osnabrueck, Germany)

This special issue is concerned with language resources and evaluation in the area of multiword expressions. For a number of years, the focus in the natural language processing (NLP) community on the problems posed by multiword expressions (MWE) was on English. Recently, for example at the ACL and EACL workshops on multiword expressions, attention has expanded to other languages including Dutch, Chinese, Japanese, German, Estonian, Russian, Basque, Turkish and Hindi. This necessitates a re-evaluation of earlier rule-based, statistical and hybrid techniques for MWE extraction and classification. In English, MWE types such as phrasal verbs, light verb constructions, noun compounds, proper names, and non-compositional idioms, have been considered. However, in other languages some MWE types can be represented as compound words, e.g. phrasal verbs in English are generally expressed as verb-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, case marking, word order or word segmentation.

The focus of this special issue is on the acquisition and analysis of language resources related to MWE and the methods for evaluation of the extraction procedures and resulting MWE resources. Language resources include written or spoken corpora marked up for MWE, terminology or domain specific databases and dictionaries of MWE, as well as software tools for their acquisition and analysis.

Topics to be addressed include, but are not limited to:

1. *Linguistic analysis*: of MWE based on language resources (such as corpora) and the impact that these studies have on NLP applications. We also welcome articles which undertake cross-linguistic analyses of MWE or which identify variation across languages. Here we include studies which investigate how application of techniques developed for one language can be transferred to another language and how successful bilingual (or multilingual) approaches are.
2. *Typologies of MWE*: descriptions of different classes of MWE and their representation in language resources, and in addition to evaluate how well computational techniques transfer across different types of MWE. Investigations of the variability of MWEs.
3. *Extraction methods*: Do methods generalise across languages? What is the interaction between linguistic descriptions/analyses of MWE and extraction methods? (i.e. to what extent are linguistically informed extraction methods useful?) Is fully automatic extraction of MWE feasible, or will manual validation/intervention always be necessary?
4. *Evaluation strategies*: creation of gold standard MWE language resources. Comparative studies using human subjects including experts and non-experts and computational evaluations using language resources derived from the web and elsewhere. Task-based evaluation of MWE lexical resources in NLP applications
5. *Compositionality*: comparative evaluation of how well humans and machines make reliable judgements on compositionality for the various classes of MWE. Also, the extent to which compositionality is a key indicator for extraction and classification of MWE.

6. *Applications*: of the theories, techniques and tools developed in MWE research to practical tasks such as IR, Text Mining etc.

Important Dates

- Deadline for paper submission: 15th November 2007
- Notification of acceptance: 15th May 2008
- Camera-ready version of accepted paper: 15th July 2008
- Target publication date: 4th quarter 2008

Instructions for Authors

Submissions should be not more than 20 pages long, must be in English, and follow the submission guidelines at

http://www.springer.com/cda/content/document/cda_downloaddocument/instr_print_10579.06 0421.pdf?SGWID=0-0-45-126854-p35554703

Extended and revised versions of papers accepted at previous ACL and EACL workshops on multiword expressions, e.g. Workshops held in Trento (April 2006), Sydney (July 2006) and in Prague (June 2007) are encouraged.

Authors are advised to use the online manuscript submission for the journal. Authors are also encouraged to send a brief email to Paul Rayson (paul@comp.lancs.ac.uk) indicating their intention to participate as soon as possible, including their contact information and the topic they intend to address in their submissions. Enquiries regarding the special issue should be sent to the same address.