

# **LREC WORKSHOP**

## **MERGING AND LAYERING LINGUISTIC INFORMATION**

To be held in conjunction with  
The 5th International Language Resources and Evaluation Conference

Magazzini del Cotone Conference Centre  
Genoa, Italy  
May 23, 2006

### **ORGANIZERS**

Erhard Hinrichs, University of Tuebingen, Germany  
Nancy Ide, Vassar College, USA  
Martha Palmer, University of Colorado, Boulder, USA  
James Pustejovsky, Brandeis University, USA

Treebanks and other theme-specific annotation schemes, together with stand-alone resources such as syntactic and semantic lexicons, wordnets, and framenets, enable annotation of natural language at different structural levels. These resources have become crucially important for the development of data-driven approaches to NLP, human language technologies, grammar extraction, and linguistic research in general. However, most of these resources and schemes have been developed by different groups working at different sites around the world, and their design is often driven by different linguistic theories and/or application requirements. Efforts to merge resources and annotations in order to exploit the information in all of them have shown how difficult the problem of mapping categories and features reflecting a particular conceptual design can be.

This workshop is designed to bring together researchers involved in the development and/or use of theme-specific annotation schemes and supporting language resources to share experiences and methodologies, in order to provide a basis for addressing the obstacles to future resource and annotation development efforts. Another goal of the workshop is to move towards agreement on linguistic annotation standards for different levels of representation; that is, frameworks that will allow (a) individual annotations to cohabit with one another (providing consistency), (b) specification components from different annotation schemas to communicate with one another, in order to refer to merged information (creating integration), (c) underspecification of annotation information at all levels (enabling incremental addition of information over the processing history), (d) maintenance of individual annotations as separate schemas for development, acquisition, and processing purposes; and (e) annotation of multi-lingual and multi-modal data. Finally, the workshop is intended to promote collaboration within the international research community on the harmonization of representations for linguistic information for use in both language resources and annotations.

We invite submission of papers on topics relevant to resource and annotation formalisms, including but not limited to:

- design principles and annotation schemes for theme-specific annotations and resources such as treebanks, lexicons, etc.
- experiences with and methods for merging information in existing resources, including both resources of the same type (e.g. lexical/ semantic resources) and those containing linguistic information of different types (e.g., syntax, co-reference, discourse, etc.)
- experiences with and methods for merging annotations for different linguistic phenomena;
- the role of linguistic theories in annotation development;
- representation frameworks for multi-layered linguistic annotations;
- methods for and results of evaluation of annotation standards;
- tools for creation and management of integrated annotation schemas;
- applications of resources and theme-specific annotations in acquiring linguistic knowledge for NLP.

## **DATES**

Paper submission : March 8, 2006

Author notification : April 7, 2006

Workshop date : May 23, 2006

## **SUBMISSION INFORMATION**

Papers should be no more than 8 pages in length and follow the format for submissions to the main LREC conference. Submissions in pdf format should be sent to [merging@cs.vassar.edu](mailto:merging@cs.vassar.edu).

## **CONTACT**

Please send inquiries to [merging@cs.vassar.edu](mailto:merging@cs.vassar.edu).

## **PROGRAM COMMITTEE**

Eneko Agirre, Basque Country University (Spain)

Collin Baker, International Computer Science Institute (USA)

Gosse Bouma (University of Groningen, The Netherlands)

Monserrat Civit (Centre de Llenguatges i Computació, University of Barcelona)

Hamish Cunningham, University of Sheffield (UK)

Bonnie Dorr, University of Maryland (USA)

Eva Ejerhed (U. of Umea, Umea, Sweden)

Tomaz Erjavec, Institut Josef Stefan (Slovenia)

David Farwell (CRL New Mexico State University, Las Cruces, NM)

Christiane Fellbaum, Princeton University (USA)

Charles J. Fillmore (International Computer Science Institute, Berkeley)

Jan Hajic (Center for Computational Linguistics, Charles University, Prague)

Eva Hajicova (Center for Computational Linguistics, Charles University, Prague)

Eduard Hovy, International Sciences Institute (USA)

Sandra Kübler (U. of Tübingen, Germany)

Alessandro Lenci (University of Pisa, Italy)  
Lori Levin (LTI, CMU, Pittsburgh, PA)  
Inderjeet Mani (MITRE, Bedford, MA)  
Adam Meyer (NYU, New York, NY)  
Rada Mihalcea, University of North Texas (USA)  
Sergei Nirenburg (University of Maryland, Baltimore County)  
Joakim Nivre (Växjö University, Sweden)  
Boyan A. Onyshkevych (U.S. Dept. of Defense)  
Karel Pala, (Masaryk University, Brno)  
Gerald Penn (University of Toronto, Toronto)  
Wim Peters, University of Sheffield (UK)  
Manfred Pinkal (DFKI, Saarbruecken, Germany)  
Massimo Poesio, University of Essex (UK)  
Adam Przepiorkowski (Polish Academy of Sciences, Warsaw, Poland)  
Owen Rambow (Columbia University, NYC)  
Kiril Simov (CLPP, Sofia, Bulgaria)  
Beth Sundheim (SPAWAR Systems Center, San Diego)  
Piek Vossen (Iriion technologies, The Netherlands)  
Fei Xia (IBM Watson, Hawthorne, NY)  
Bert Xue (UPENN, Philadelphia, PA)  
Dietmar Zaefferer (Ludwig-Maximilians-Universitaet, Muenchen, Germany)  
Annie Zaenen, (PARC, Palo Alto, CA)