ACL 2007

The LAW
Proceedings of The Linguistic Annotation Workshop

June 28-29, 2007
Prague, Czech Republic

The Association for Computational Linguistics

Production and Manufacturing by
Omnipress
2600 Anderson Street
Madison, WI 53704
USA

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`

Preface

Welcome to The Linguistic Annotation Workshop (The LAW).

Linguistically annotated corpora play a major role in parsing, information extraction, question answering, machine translation and many other areas of computational linguistics, and provide an empirical testbed for theoretical linguistics research. This has led to a proliferation of annotation systems, frameworks, formats, and schemes. Recognition of the need to harmonize annotation practices and frameworks has become increasingly critical, as witnessed by numerous workshops dealing with different aspects of linguistic annotation over the past few years.

The LAW addresses all aspects of linguistic annotation in a single forum by merging two existing workshop series: NLPXML (Natural Language Processing and XML) and FLAC (Frontiers in Linguistically Annotated Corpora). The goals of the workshop include:

1. The exchange and propagation of research results with respect to the annotation, manipulation and exploitation of corpora, taking into account different applications and theoretical investigations in the field of language technology and research;
2. Working towards harmonization and interoperability from the perspective of the increasingly large number of tools and frameworks that support the creation, instantiation, manipulation, querying, and exploitation of annotated resources;
3. Working towards a consensus on all issues crucial to the advancement of the field of corpus annotation.

These proceedings include 11 long papers, 5 short papers, 4 demo descriptions and 8 posters selected by the program committee from 51 submissions for presentation at the workshop. In addition to these presentations, the workshop includes demonstrations of annotation tools, reports by working groups, and an open discussion session.

We would like to thank the members of the program committee for their timely reviews. We also thank the Workshops Chair and other organizers of ACL-2007 for their support. Finally, we congratulate Adriane Boyd, the winner of the Innovative Student Annotation Award for the paper *Discontinuity Revisited: An Improved Conversion to Context-Free Representations*.

Branimir Boguraev
Nancy Ide
Adam Meyers
Shigeko Nariyama
Manfred Stede
Janyce Wiebe
Graham Wilcock

Organizers

Workshop Chairs

Branimir Boguraev, IBM T. J. Watson Research Center
Nancy Ide, Vassar College
Adam Meyers, New York University
Shigeko Nariyama, University of Melbourne
Manfred Stede, University of Potsdam
Janyce Wiebe, University of Pittsburgh
Graham Wilcock, University of Helsinki

Program Committee

David Ahn, University of Amsterdam
Lars Ahrenberg, Linköping University
Timothy Baldwin, University of Melbourne
Francis Bond, NICT
Kalina Bontcheva, University of Sheffield
Paul Buitelaar, DFKI
Jean Carletta, University of Edinburgh
Key-Sun Choi, KAIST
Christopher Cieri, Linguistic Data Consortium/University of Pennsylvania
Hamish Cunningham, University of Sheffield
David Day, MITRE Corporation
Thierry Declerck, DFKI
Ludovic Denoyer, LIP6 - University of Paris 6
Tomaz Erjavec, Jozef Stefan Institute
David Farwell, New Mexico State University
Alex Chengyu Fang, City University of Hong Kong
Chuck Fillmore, International Computer Science Institute, Berkeley
Anette Frank, DFKI
John Fry, San Jose State University
Claire Grover, University of Edinburgh
Jan Hajic, Charles University
Ed Hovy, University of Southern California
Baden Hughes, University of Melbourne
Emi Izumi, NICT
Tsai Jia-Lin, Tung Nan Institute of Technology
Aravind Joshi, University of Pennsylvania
Ewan Klein, University of Edinburgh
Mounia Lalmas, University of London
Mike Maxwell, University of Maryland
Chieko Nakabasami, Toyo University
Stephan Oepen, University of Oslo
Kyonghee Paik, KLI

Martha Palmer, University of Colorado
Antonio Pareja-Lora, Universidad Complutense de Madrid / OEG - UPM
Manfred Pinkal, Saarland University
James Pustejovsky, Brandeis University
Owen Rambow, Columbia University
Laurent Romary, MPG-INRIA
Henry Thompson, University of Edinburgh
Erik Tjong Kim Sang, University of Amsterdam
Theresa Wilson, University of Edinburgh
Nainwen Xue, University of Colorado

Table of Contents

Workshop Program

Thursday, June 28, 2007

Session 1: Introduction and Long Papers

14:30–14:55 Introduction to the Workshop

14:55–15:20 *GrAF: A Graph-based Format for Linguistic Annotations*
Nancy Ide and Keith Suderman

15:20–15:45 *Efficient Annotation with the Jena ANnotation Environment (JANE)*
Katrin Tomanek, Joachim Wermter and Udo Hahn

15:45–16:15 Break

Session 2: Long Papers

16:15–16:40 *Mining Syntactically Annotated Corpora with XQuery*
Gosse Bouma and Geert Kloosterman

16:40–17:05 *Assocating Facial Displays with Syntactic Constituents for Generation*
Mary Ellen Foster

17:05–17:30 *An Annotation Type System for a Data-Driven NLP Pipeline*
Udo Hahn, Ekaterina Buyko, Katrin Tomanek, Scott Piao, John McNaught, Yoshimasa Tsuruoka and Sophia Ananiadou

Demonstration and Poster Session

17:30–18:30 Demonstrations and Posters
(Listed at end of program)

Friday, June 29, 2007

Session 3: Working Groups

Session 4: Short Papers

Friday, June 29, 2007 (continued)

Session 5: Long Papers

Session 6: Long Papers

Discussion Session

Demonstrations and Posters

Demonstrations

Posters

Demonstrations and Posters (continued)