ACL HLT 2011

LAW V

Fifth Linguitic Annotation Workshop

Proceedings of the Workshop

23-24 June 2011

Portland, Oregon, USA

Production and Manufacturing by
Omnipress, Inc.
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

ISBN 978-1-932432-93-0

Introduction

The Linguistic Annotation Workshop (The LAW) provides a forum to facilitate the exchange and propagation of research results concerned with the annotation, manipulation, and exploitation of corpora; work towards the harmonization and interoperability from the perspective of the increasingly large number of tools and frameworks for annotated language resources; and work towards a consensus on all issues crucial to the advancement of the field of corpus annotation. Although this year's LAW is officially the fifth edition, LAW itself is the convergence of several previous workshops-including NLPXML, FLAC, LINC, and Frontiers in Corpus Annotation-dating back to the first NLPXML in 2001. This series of workshops attests to the rapid developments in the creation and use of annotated data in both language technology and empirical approaches to linguistic studies over the past 10 years.

We received a sizeable number of papers this year. A total of 37 submissions were received. After careful review, the program committee accepted 10 papers and 11 posters. One of the papers selected for oral presentation was withdrawn later, leaving the total of full papers to 9. Selection of the papers was not an easy task, as the papers cover the full range of linguistic facts and their corresponding annotation frameworks, from predicate-argument to discourse structure, speech to social networks, and learner corpus to CVs. The papers also deal with a range of annotation levels, from the macro perspective on infrastructure for international collaboration and interoperability, to the micro perspective on tools to deal with inter-annotator inconsistencies. It is this richness of the topics that attest to the growing maturity of field. This year we tried a slightly different approach where we allowed the posters to be full length papers and have a ten minute talk associated with each.

We would like to thank SIGANN for its continuing endorsement of the LAW workshops. We would also like to thank the the ACL workshop co-chairs John Carroll and Hal Daume III and the publication chair Guodong Zhou for their support and help in producing the LAW V proceedings. Most of all, we would like to thank all our program committee members and reviewers for their dedication and helpful review comments. Without them, LAW V could not be implemented successfully.

Sameer Pradhan and Katrin Tomanek, Program Committee Co-chairs
Nancy Ide and Adam Meyers, Organizers

Workshop Organizers

Organizers:

Nancy Ide, Vassar College
Adam Meyers, New York University

Organizing Committee:

Sameer Pradhan (Program Co-chair), BBN Technologies
Katrin Tomanek (Program Co-chair), Friedrich-Schiller-Universität Jena
Chu-Ren Huang, The Hong Kong Polytechnic University
Antonio Pareja-Lora, Universidad Complutense de Madrid
Massimo Poesio, University of Trento
Manfred Stede, Universitat Potsdam
Nianwen Xue, Brandeis University

Program Committee:

Collin Baker	ICSI/University of California, Berkeley
Pushpak Bhattacharyya	IIT Bombay
Nicoletta Calzolari	ILC/CNR
Richard Eckart de Castilho	Technische Universitt Darmstadt
Mona Diab	Columbia University
Tomaz Erjavec	Josef Stefan Institute
Alex Chengyu Fang	City University of Hong Kong
Christiane Fellbaum	Princeton University
Charles Fillmore	ICSI/UC Berkeley
Eduard Hovy	USC/ISI
Chu-Ren Huang	Hong Kong Polytechnic
Nancy Ide	Vassar College
Richard Johansson	Lund University
Aravind Joshi	University of Pennsylvania
Edward Loper	BBN Technologies
Adam Meyers	New York University
Antonio Pareja-Lora	Universidad Complutense de Madrid
Martha Palmer	University of Colorado
Massimo Poesio	University of Trento
Rashmi Prasad	University of Pennsylvania
Vasin Punyakanok	BBN Technologies
James Pustejovsky	Brandeis University
Manfred Stede	Universitt Potsdam
Nianwen Xue	Brandeis University

Table of Contents

Conference Program

Thursday, June 23, 2011

8:45–9:00 Welcome

Session I:

9:00–9:30 *On the Development of the RST Spanish Treebank*
Iria da Cunha, Juan-Manuel Torres-Moreno and Gerardo Sierra

9:30–10:00 *OWL/DL formalization of the MULTEXT-East morphosyntactic specifications*
Christian Chiarcos and Tomaž Erjavec

10:00–10:30 *Analysis of the Hindi Proposition Bank using Dependency Structure*
Ashwini Vaidya, Jinho Choi, Martha Palmer and Bhuvana Narasimhan

10:30–11:00 Coffee Break

11:00–11:30 *How Good is the Crowd at "real" WSD?*
Jisup Hong and Collin F. Baker

11:30–12:00 *Consistency Maintenance in Prosodic Labeling for Reliable Prediction of Prosodic Breaks*
Youngim Jung and Hyuk-Chul Kwon

12:00–12:30 *An Annotation Scheme for Automated Bias Detection in Wikipedia*
Livnat Herzig, Alex Nunes and Batia Snir

12:30–14:00 Lunch Break

Thursday, June 23, 2011 (continued)

Session II:

14:00–14:10 *A Collaborative Annotation between Human Annotators and a Statistical Parser*
Shun'ya Iwasawa, Hiroki Hanaoka, Takuya Matsuzaki, Yusuke Miyao and Jun'ichi Tsujii

14:10–14:20 *Reducing the Need for Double Annotation*
Dmitriy Dligach and Martha Palmer

14:20–14:30 *Crowdsourcing Word Sense Definition*
Anna Rumshisky

14:30–14:40 *A scaleable automated quality assurance technique for semantic representations and proposition banks*
K. Bretonnel Cohen, Lawrence Hunter and Martha Palmer

14:40–14:50 *Proposal for an Extension of Traditional Named Entities: From Guidelines to Evaluation, an Overview*
Cyril Grouin, Sophie Rosset, Pierre Zweigenbaum, Karën Fort, Olivier Galibert and Ludovic Quintard

14:50–15:00 *Assessing the practical usability of an automatically annotated corpus*
Md. Faisal Mahbub Chowdhury and Alberto Lavelli

15:00–15:10 *Subjectivity and Sentiment Annotation of Modern Standard Arabic Newswire*
Muhammad Abdul-Mageed and Mona Diab

15:10–15:20 *Creating an Annotated Tamil Corpus as a Discourse Resource*
Ravi Teja Rachakonda and Dipti Misra Sharma

15:30–16:00 Coffee Break

Thursday, June 23, 2011 (continued)

Session III:

16:00–16:10 *A Gold Standard Corpus of Early Modern German*
Silke Scheible, Richard J. Whitt, Martin Durrell and Paul Bennett

16:10–16:20 *MAE and MAI: Lightweight Annotation and Adjudication Tools*
Amber Stubbs

16:20–16:30 *Empty Categories in Hindi Dependency Treebank: Analysis and Recovery*
Chaitanya GSK, Samar Husain and Prashanth Mannem

16:30–17:00 SIGANN task announcement

17:00–17:30 SILT challange annoncement

Friday, June 24, 2011

8:45–9:00 Opening

Session IV:

9:00–10:30 Poster Session

10:30–11:00 Coffee Break

11:00–11:30 *Annotating Events, Temporal Expressions and Relations in Italian: the It-Timeml Experience for the Ita-TimeBank*
Tommaso Caselli, Valentina Bartalesi Lenzi, Rachele Sprugnoli, Emanuele Pianta and Irina Prodanof

11:30–12:00 *Increasing Informativeness in Temporal Annotation*
James Pustejovsky and Amber Stubbs

12:00–12:30 *Discourse-constrained Temporal Annotation*
Yuping Zhou and Nianwen Xue