

ACL-IJCNLP 2009

LAW III

Third Linguistic Annotation Workshop

Proceedings of the Workshop

6-7 August 2009
Suntec, Singapore

Production and Manufacturing by
World Scientific Publishing Co Pte Ltd
5 Toh Tuck Link
Singapore 596224

©2009 The Association for Computational Linguistics
and The Asian Federation of Natural Language Processing

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 978-1-932432-52-7 / 1-932432-52-3

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 third 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.

The response to this year's Call for Papers was enthusiastic: 43 submissions were received. After careful review, the program committee accepted 10 long papers, 11 short papers, and 15 posters. 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 wordnets to treebanks, emotion to belief, and speech to discourse. 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, which will be represented in a special issue of the journal *Language Resources and Evaluation* devoted to selected papers from the workshop.

We would like to thank SIGANN for its continuing endorsement of the LAW workshops, as well as the support and comments from the ACL-IJCNLP 2009 workshop committee chairs: Jimmy Lin and Yuji Matsumoto. 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 III could not be implemented successfully.

Chu-Ren Huang and Manfred Stede, Program Committee Co-chairs
Nancy Ide and Adam Meyers, Organizers

Workshop Committees

Organizers:

Nancy Ide, Vassar College
Adam Meyers, New York University

Organizing Committee:

Manfred Stede (Program Co-chair), Universität Potsdam
Chu-Ren Huang (Program Co-chair), Hong Kong Polytechnic University / Academia Sinica
Antonio Pareja-Lora, SIC, UCM / OEG, UPM
Sameer Pradhan, BBN Technologies
Nianwen Xue, Brandeis University

Program Committee:

Collin Baker	<i>ICSI/UC Berkeley</i>
Timothy Baldwin	<i>University of Melbourne</i>
Francis Bond	<i>NICT</i>
Nicoletta Calzolari	<i>ILC/CNR</i>
Steve Cassidy	<i>Macquarie University</i>
Christopher Cieri	<i>Linguistic Data Consortium/University of Pennsylvania</i>
Tomaz Erjavec	<i>Josef Stefan Institute</i>
Katrin Erk	<i>University of Texas at Austin</i>
Alex Chengyu Fang	<i>City University of Hong Kong</i>
Christiane Fellbaum	<i>Princeton University</i>
Charles Fillmore	<i>ICSI/UC Berkeley</i>
Nancy Ide	<i>Vassar College</i>
Richard Johansson	<i>Lund University</i>
Aravind Joshi	<i>University of Pennsylvania</i>
Adam Meyers	<i>New York University</i>
Eleni Miltsakaki	<i>University of Pennsylvania</i>
Joakim Nivre	<i>Växjö University and Uppsala University</i>
Eric Nyberg	<i>Carnegie-Mellon University</i>
Antonio Pareja-Lora	<i>SIC, UCM / OEG, UPM</i>
Martha Palmer	<i>University of Colorado</i>
Sameer Pradhan	<i>BBN Technologies</i>
James Pustejovsky	<i>Brandeis University</i>
Mihai Surdeanu	<i>Yahoo! Research, Barcelona</i>
Theresa Wilson	<i>University of Edinburgh</i>
Andreas Witt	<i>IDS Mannheim</i>
Nianwen Xue	<i>Brandeis University</i>

Table of Contents

<i>A Cognitive-based Annotation System for Emotion Computing</i> Ying Chen, Sophia Y. M. Lee and Chu-Ren Huang	1
<i>Complex Linguistic Annotation No Easy Way Out! A Case from Bangla and Hindi POS Labeling Tasks</i> Sandipan Dandapat, Priyanka Biswas, Monojit Choudhury and Kalika Bali	10
<i>Assessing the benefits of partial automatic pre-labeling for frame-semantic annotation</i> Ines Rehbein, Josef Ruppenhofer and Caroline Sporleder	19
<i>Bridging the Gaps: Interoperability for GrAF, GATE, and UIMA</i> Nancy Ide and Keith Suderman	27
<i>By all these lovely tokens... Merging Conflicting Tokenizations</i> Christian Chiarcos, Julia Ritz and Manfred Stede	35
<i>Annotating Subordinators in the Turkish Discourse Bank</i> Deniz Zeyrek, Ümit Deniz Turan, Cem Bozsahin, Ruket Çakıcı, Ayıışıđı B. Sevdik-Çallı, Işın Demirşahin, Berfin Aktaş, İhsan Yalçınkaya and Hale Ögel	44
<i>Annotation of Events and Temporal Expressions in French Texts</i> André Bittar	48
<i>Designing a Language Game for Collecting Coreference Annotation</i> Barbora Hladká, Jiří Mírovský and Pavel Schlesinger	52
<i>Explorations in Automatic Image Annotation using Textual Features</i> Chee Wee Leong and Rada Mihalcea	56
<i>Human Evaluation of Article and Noun Number Usage: Influences of Context and Construction Variability</i> John Lee, Joel Tetreault and Martin Chodorow	60
<i>Stand-off TEI Annotation: the Case of the National Corpus of Polish</i> Piotr Banski and Adam Przepiórkowski	64
<i>Committed Belief Annotation and Tagging</i> Mona Diab, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaram and Weiwei Guo	68
<i>Annotation of Sentence Structure; Capturing the Relationship among Clauses in Czech Sentences</i> Markéta Lopatková, Natalia Klyueva and Petr Homola	74
<i>Schema and Variation: Digitizing Printed Dictionaries</i> Christian Schneider, Dietmar Seipel and Werner Wegstein	82
<i>Syntactic annotation of spoken utterances: A case study on the Czech Academic Corpus</i> Barbora Hladká and Zdenka Uresova	90
<i>High-Performance High-Volume Layered Corpora Annotation</i> Tiago Luís and David Martins de Matos	99

<i>The Coding Scheme for Annotating Extended Nominal Coreference and Bridging Anaphora in the Prague Dependency Treebank</i>	
Anna Nedoluzhko, Jiří Mírovský and Petr Pajas	108
<i>Timed Annotations — Enhancing MUC7 Metadata by the Time It Takes to Annotate Named Entities</i>	
Katrin Tomanek and Udo Hahn	112
<i>Transducing Logical Relations from Automatic and Manual GLARF</i>	
Adam Meyers, Michiko Kosaka, Heng Ji, Nianwen Xue, Mary Harper, Ang Sun, Wei Xu and Shasha Liao	116
<i>Using Parallel Propbanks to enhance Word-alignments</i>	
Jinho Choi, Martha Palmer and Nianwen Xue	121
<i>WordNet and FrameNet as Complementary Resources for Annotation</i>	
Collin F. Baker and Christiane Fellbaum	125
<i>Annotating language errors in texts: investigating argumentation and decision schemas</i>	
Camille Albert, Laurie Buscail, Marie Garnier, Arnaud Rykner and Patrick Saint-Dizier	130
<i>Developing Novel Multimodal and Linguistic Annotation Software</i>	
Alexey Podlasov, Kay O’Halloran, Sabine Tan, Bradley Smith and Arun Nagarajan	134
<i>Unsupervised Detection of Annotation Inconsistencies Using Apriori Algorithm</i>	
Václav Novák and Magda Razímová	138
<i>Towards a Methodology for Named Entities Annotation</i>	
Karěn Fort, Maud Ehrmann and Adeline Nazarenko	142
<i>Online Search Interface for the Sejong Korean-Japanese Bilingual Corpus and Auto-interpolation of Phrase Alignment</i>	
Sanghoun Song and Francis Bond	146
<i>Annotating Semantic Relations Combining Facts and Opinions</i>	
Koji Murakami, Shouko Masuda, Suguru Matsuyoshi, Eric Nichols, Kentaro Inui and Yuji Matsumoto	150
<i>Ingesting the Auslan Corpus into the DADA Annotation Store</i>	
Steve Cassidy and Trevor Johnston	154
<i>The Hindi Discourse Relation Bank</i>	
Umangi Oza, Rashmi Prasad, Sudheer Kolachina, Dipti Misra Sharma and Aravind Joshi	158
<i>Simple Parser for Indian Languages in a Dependency Framework</i>	
Akshar Bharati, Mridul Gupta, Vineet Yadav, Karthik Gali and Dipti Misra Sharma	162
<i>Annotating Discourse Anaphora</i>	
Stefanie Dipper and Heike Zinsmeister	166
<i>Annotating Wall Street Journal Texts Using a Hand-Crafted Deep Linguistic Grammar</i>	
Valia Kordoni and Yi Zhang	170
<i>A general scheme for broad-coverage multimodal annotation</i>	
Philippe Blache	174

<i>The SILT and FlaReNet International Collaboration for Interoperability</i>	
Nancy Ide, James Pustejovsky, Nicoletta Calzolari and Claudia Soria	178
<i>Building a Large Syntactically-Annotated Corpus of Vietnamese</i>	
Phuong Thai Nguyen, Xuan Luong Vu, Thi Minh Huyen Nguyen, Van Hiep Nguyen and Hong Phuong Le	182
<i>A Multi-Representational and Multi-Layered Treebank for Hindi/Urdu</i>	
Rajesh Bhatt, Bhuvana Narasimhan, Martha Palmer, Owen Rambow, Dipti Misra Sharma and Fei Xia	186

Conference Program

Thursday, August 6, 2009

- 8:50–9:00 Opening Remarks
- 9:00–9:30 *A Cognitive-based Annotation System for Emotion Computing*
Ying Chen, Sophia Y. M. Lee and Chu-Ren Huang
- 9:30–10:00 *Complex Linguistic Annotation—No Easy Way Out! A Case from Bangla and Hindi POS Labeling Tasks*
Sandipan Dandapat, Priyanka Biswas, Monojit Choudhury and Kalika Bali
- 10:00–10:30 Break
- 10:30–11:00 *Assessing the benefits of partial automatic pre-labeling for frame-semantic annotation*
Ines Rehbein, Josef Ruppenhofer and Caroline Sporleder
- 11:00–11:30 *Bridging the Gaps: Interoperability for GrAF, GATE, and UIMA*
Nancy Ide and Keith Suderman
- 11:30–12:00 *By all these lovely tokens... Merging Conflicting Tokenizations*
Christian Chiarcos, Julia Ritz and Manfred Stede
- 12:00–13:30 Lunch Break
- 13:30–13:50 *Annotating Subordinators in the Turkish Discourse Bank*
Deniz Zeyrek, Ümit Deniz Turan, Cem Bozsahin, Ruket Çakıcı, Ayışığı B. Sevdik-Çallı, Işın Demirşahin, Berfin Aktaş, İhsan Yalçinkaya and Hale Ögel
- 13:50–14:10 *Annotation of Events and Temporal Expressions in French Texts*
André Bittar
- 14:10–14:30 *Designing a Language Game for Collecting Coreference Annotation*
Barbora Hladká, Jiří Mírovský and Pavel Schlesinger
- 14:30–14:50 *Explorations in Automatic Image Annotation using Textual Features*
Chee Wee Leong and Rada Mihalcea
- 14:50–15:10 *Human Evaluation of Article and Noun Number Usage: Influences of Context and Construction Variability*
John Lee, Joel Tetreault and Martin Chodorow

Thursday, August 6, 2009 (continued)

15:10–15:30 *Stand-off TEI Annotation: the Case of the National Corpus of Polish*
Piotr Banski and Adam Przepiórkowski

15:30–16:00 Break

16:00–17:30 POSTER SESSION

17:30–18:00 SIGANN Annual Meeting

Friday, August 7, 2009

9:00–9:30 *Committed Belief Annotation and Tagging*
Mona Diab, Lori Levin, Teruko Mitamura, Owen Rambow, Vinodkumar Prabhakaram and Weiwei Guo

9:30–10:00 *Annotation of Sentence Structure: Capturing the Relationship among Clauses in Czech Sentences*
Markéta Lopatková, Natalia Klyueva and Petr Homola

10:00–10:30 Break

10:30–11:00 *Schema and Variation: Digitizing Printed Dictionaries*
Christian Schneider, Dietmar Seipel and Werner Wegstein

11:00–11:30 *Syntactic Annotation of Spoken Utterances: A Case Study on the Czech Academic Corpus*
Barbora Hladká and Zdenka Uresova

11:30–12:00 *High-Performance High-Volume Layered Corpora Annotation*
Tiago Luís and David Martins de Matos

12:00–13:30 Lunch Break

13:30–13:50 *The Coding Scheme for Annotating Extended Nominal Coreference and Bridging Anaphora in the Prague Dependency Treebank*
Anna Nedoluzhko, Jiří Mírovský and Petr Pajas

13:50–14:10 *Timed Annotations — Enhancing MUC7 Metadata by the Time It Takes to Annotate Named Entities*
Katrin Tomanek and Udo Hahn

Friday, August 7, 2009 (continued)

- 14:10–14:30 *Transducing Logical Relations from Automatic and Manual GLARF*
Adam Meyers, Michiko Kosaka, Heng Ji, Nianwen Xue, Mary Harper, Ang Sun, Wei Xu
and Shasha Liao
- 14:30–14:50 *Using Parallel Propbanks to enhance Word-alignments*
Jinho Choi, Martha Palmer and Nianwen Xue
- 14:50–15:10 *WordNet and FrameNet as Complementary Resources for Annotation*
Collin F. Baker and Christiane Fellbaum
- 15:10–15:30 SIGANN Working Group Reports
- 15:30–16:00 Break
- 16:00–17:50 PANEL: The Standards Debate: Pro and Con
- 17:50–18:00 Closing

