LAW VIII

The 8th Linguistic Annotation Workshop in conjunction with COLING 2014

Proceedings of the Workshop

August 23-24, 2014
Dublin, Ireland

©2014 The Authors

The papers in this volume are licensed by the authors under a Creative Commons Attribution 4.0 International License.

ISBN 978-1-941643-29-7
Proceedings of the 8th Linguistic Annotation Workshop (LAW-VIII)
Lori Levin, Manfred Stede (eds.)

Preface

The Linguistic Annotation Workshop (The LAW) is organized annually by the Association for Computational Linguistics Special Interest Group for Annotation (ACL SIGANN). It provides a forum to facilitate the exchange and propagation of research results concerned with the annotation, manipulation, and exploitation of corpora; work towards 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.

The series is now in its eighth year, with these proceedings including papers that were presented at LAW VIII, held in conjunction with the COLING conference in Dublin, Ireland, on August 23-24 2014. As in previous years, more than 40 submissions have originally been received in response to the call for papers. After careful review, the program committee accepted 11 long papers and three short papers for oral presentation, together with eight additional papers to be presented as posters. The topics of the long papers revolve quite nicely around major linguistic levels of description: part of speech, syntax, semantics, and discourse; and thus we arranged them in theses groups in the program. The short papers report on interesting experiments or new tools.

Our thanks go to SIGANN, our organizing committee, for its continuing organization of the LAW workshops, and to the COLING 2014 workshop chairs for their support: Jennifer Foster, Dan Gildea and Tim Baldwin. Also, we thank the COLING 2014 publication chairs for their help with these proceedings.

Most of all, we would like to thank all the authors for submitting their papers to the workshop, and our program committee members for their dedication and their thoughtful reviews.

Lori Levin and Manfred Stede, program co-chairs

Workshop Chairs

Lori Levin (Carnegie Mellon University)
Manfred Stede (University of Potsdam)

Organizing Committee

Stefanie Dipper (Ruhr University Bochum)
Chu-Ren Huang (The Hong Kong Polytechnic University)
Nancy Ide (Vassar College)
Adam Meyers (New York University)
Antonio Pareja-Lora (SIC & ILSA, UCM / ATLAS, UNED)
Massimo Poesio (University of Trento)
Sameer Pradhan (Harvard University)
Katrin Tomanek (University of Jena)
Fei Xia (University of Washington)
Nianwen Xue (Brandeis University)

Program Committee

Collin Baker (UC Berkeley)
Archna Bhatia (Carnegie Mellon University)
Nicoletta Calzolari (ILC/CNR)
Christian Chiarcos (University of Frankfurt)
Stefanie Dipper (Ruhr University Bochum)
Tomaž Erjavec (Josef Stefan Institute)
Dan Flickinger (Stanford University)
Udo Hahn (University of Jena)
Chu-Ren Huang (The Hong Kong Polytechnic University)
Nancy Ide (Vassar College)
Aravind Joshi (University of Pennsylvania)
Valia Kordoni (Humboldt University Berlin)
Adam Meyers (New York University)
Antonio Pareja-Lora (SIC & ILSA, UCM / ATLAS, UNED)
Massimo Poesio (University of Trento)
Sameer Pradhan (Harvard University)
James Pustejovsky (Brandeis University)
Katrin Tomanek (University of Jena)
Yulia Tsvetkov (Carnegie Mellon University)
Andreas Witt (IDS Mannheim)
Marie-Paule Péry-Woodley (Université de Toulouse 2)
Fei Xia (University of Washington)
Nianwen Xue (Brandeis University)
Heike Zinsmeister (University of Hamburg)

Table of Contents

Workshop Program

Saturday, August 23

8.50-9:00 Opening remarks

Parts of Speech

9:00–9:30 *STTS 2.0? Improving the Tagset for the Part-of-Speech-Tagging of German Spoken Data*
Swantje Westpfahl

9:30–10:00 *Use of Coreference in Automatic Searching for Multiword Discourse Markers in the Prague Dependency Treebank*
Magdalena Rysova and Jiří Mírovský

10:00–10:30 *POS Error Detection in Automatically Annotated Corpora*
Ines Rehbein

10:30-11:00 Break

Syntax

11:00–11:30 *Aligning Chinese-English Parallel Parse Trees: Is it Feasible?*
Dun Deng and Nianwen Xue

11:30–12:00 *Sentence Diagrams: their Evaluation and Combination*
Jirka Hana, Barbora Hladka and Ivana Luksova

12:00-12:30 Poster Boasters

12:30-14:00 Lunch

Short Papers

14:00–14:20 *Finding Your "Inner-Annotator": An Experiment in Annotator Independence for Rating Discourse Coherence Quality in Essays*
Jill Burstein, Swapna Somasundaran and Martin Chodorow

14:20–14:40 *Optimizing Annotation Efforts to Build Reliable Annotated Corpora for Training Statistical Models*
Cyril Grouin, Thomas Lavergne and Aurelie Neveol

14:40–15:00 *A Web-based Geo-resolution Annotation and Evaluation Tool*
Beatrice Alex, Kate Byrne, Claire Grover and Richard Tobin

15:00-15:30 Break

Saturday, August 23 (continued)

(15:30-17:00) Poster session

Annotating Uncertainty in Hungarian Webtext
Veronika Vincze, Katalin Ilona Simkó and Viktor Varga

A Corpus Study for Identifying Evidence on Microblogs
Paul Reisert, Junta Mizuno, Miwa Kanno, Naoaki Okazaki and Kentaro Inui

Semi-Semantic Part of Speech Annotation and Evaluation
Qaiser Abbas

Multiple Views as Aid to Linguistic Annotation Error Analysis
Marilena Di Bari, Serge Sharoff and Martin Thomas

Annotating a Large Representative Corpus of Clinical Notes for Parts of Speech
Narayan Choudhary, Parth Pathak, Pinal Patel and Vishal Panchal

Part-of-speech Tagset and Corpus Development for Igbo, an African Language
Ikechukwu Onyenwe, Chinedu Uchechukwu and Mark Hepple

Annotating Descriptively Incomplete Language phenomena
Fabian Barteld, Sarah Ihden, Ingrid Schröder and Heike Zinsmeister

Annotating Discourse Connectives in Spoken Turkish
Isin Demirsahin and Deniz Zeyrek

Sunday, August 24

Discourse

9:00–9:30 *Exploiting the Human Computational Effort Dedicated to Message Reply Formatting for Training Discursive Email Segmenters*
Nicolas Hernandez and Soufian Salim

9:30–10:00 *Annotating Multiparty Discourse: Challenges for Agreement Metrics*
Nina Wacholder, Smaranda Muresan, Debanjan Ghosh and Mark Aakhus

10:00–10:30 *Towards Automatic Annotation of Clinical Decision-Making Style*
Limor Hochberg, Cecilia Ovesdotter Alm, Esa M. Rantanen, Qi Yu, Caroline M. DeLong and Anne Haake

10:30-11:00 Break

Semantics

11:00–11:30 *Interactive Annotation for Event Modality in Modern Standard and Egyptian Arabic Tweets*
Rania Al-Sabbagh, Roxana Girju and Jana Diesner

11:30–12:00 *Situation Entity Annotation*
Annemarie Friedrich and Alexis Palmer

12:00–12:30 *Focus Annotation in Reading Comprehension Data*
Ramon Ziai and Detmar Meurers

12:30-14:00 Lunch

14:00-15:00 Presentation of NSF Community Infrastructure proposal

15:00-15:30 Break

15:30-16:30 Discussion of NSF Community Infrastructure proposal

16:30-17:30 LAW Business Meeting and Discussion of Shared Task