The 14th Linguistic Annotation Workshop

Proceedings of the Workshop

December 12, 2020

Barcelona, Spain (Online)

Copyright of each paper stays with the respective authors (or their employers).

ISBN 978-1-952148-33-0

Introduction to the Workshop

The Linguistic Annotation Workshop (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. These proceedings include papers that were presented at LAW XIV, held online in conjunction with the annual meeting of the Association for Computational Linguistics (ACL) in Barcelona, Spain, on December 12, 2020.

The series is now in its fourteenth year. The first workshop took place in 2007 at the ACL in Prague. Since then, the LAW has been held every year, consistently drawing substantial participation (both in terms of paper/poster submissions and participation in the actual workshop) providing evidence that the LAW's overall focus continues to be an important area of interest in the field, a substantial part of which relies on supervised learning from gold standard data sets. This year's LAW has received 33 submissions, out of which 17 papers have been accepted to be presented at the workshop, 7 as talks and 10 as posters. In addition to oral and poster paper presentations, LAW XIV also features an invited talk by Nils Reiter.

Our thanks go to SIGANN, our organizing committee, for its continuing organization of the LAW workshops, and to the ACL 2020 workshop chairs for their support. Also, we thank Luke Gessler and Adam Roussel, the LAW XIV publicity and publications chairs, for their invaluable 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.

Special Theme: "Breaking the Mold"

This year's special theme is "Breaking the Mold", promoting novel or less studied types of annotations and data. This includes papers on less common annotations (for example metaphor, sense anaphora, gesture, humor); annotations in understudied settings or data types (new text types or novel collection methods); and annotation for languages other than English, especially beyond tagging/treebanking.

Stefanie Dipper and Amir Zeldes
Workshop chairs

Program Committee Co-chairs

Stefanie Dipper, Ruhr-Universität Bochum
Amir Zeldes, Georgetown University

Publicity and Publications Chairs

Luke Gessler, Georgetown University
Adam Roussel, Ruhr-Universität Bochum

Organizing Committee

Stefanie Dipper, Ruhr-Universität Bochum
Annemarie Friedrich, Bosch Research
Chu-Ren Huang, The Hong Kong Polytechnic University
Jena D. Hwang, Allen Institute for AI
Nancy Ide, Vassar College
Lori Levin, Carnegie Mellon University
Adam Meyers, New York University
Antonio Pareja-Lora, Universidad de Alcalá de Henares (UAH)/DMEG (UdG)/ATLAS (UNED)
Massimo Poesio, Queen Mary University of London
Sameer Pradhan, Boulder Learning, Inc.
Ines Rehbein, University of Mannheim
Nathan Schneider, Georgetown University
Manfred Stede, University of Potsdam
Katrin Tomanek, Google
Fei Xia, University of Washington
Deniz Zeyrek, Middle East Technical University
Heike Zinsmeister, Universität Hamburg

Program Committee

Ron Artstein, USC Institute for Creative Technologies
Nicoletta Calzolari, ILC-CNR
Emmanuele Chersoni, Hong Kong Polytechnic University
Kathryn Conger, University of Colorado, Boulder
Marie-Catherine de Marneffe, The Ohio State University
Jonathan Dunn, University of Canterbury
Richard Eckart de Castilho, UKP Lab, Technische Universität Darmstadt
Kilian Evang, Heinrich Heine University Düsseldorf
Federico Fancellu, Samsung AI Research Canada
Annemarie Friedrich, Bosch Research
Kim Gerdes, Sorbonne Nouvelle / Inria
Jena D. Hwang, Allen Institute for AI
Nancy Ide, Vassar College
Mikel Iruskieta, University of the Basque Country (UPV/EHU)
Sandra Kübler, Indiana University
John Lee, City University of Hong Kong
Els Lefever, LT3, Ghent University

Lori Levin, Carnegie Mellon University
Adam Meyers, New York University
Kemal Oflazer, Carnegie Mellon University in Qatar
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences
Antonio Pareja-Lora, Universidad de Alcalá de Henares (UAH)/DMEG (UdG)/ATLAS (UNED)
Miriam R. L. Petruck, International Computer Science Institute
Massimo Poesio, Queen Mary University of London
Ines Rehbein, University of Mannheim
Michael Roth, University of Stuttgart
Nathan Schneider, Georgetown University
Manfred Stede, University of Potsdam
Bonnie Webber, University of Edinburgh
Michael Wiegand, Alpen-Adria-Universität Klagenfurt
Nianwen Xue, Brandeis University
Deniz Zeyrek, Middle East Technical University
Heike Zinsmeister, Universität Hamburg

Table of Contents

Bayesian Methods for Semi-supervised Text Annotation
Kristian Miok, Gregor Pirs and Marko Robnik-Sikonja 1

Provenance for Linguistic Corpora through Nanopublications
Timo Lek, Anna de Groot, Tobias Kuhn and Roser Morante 13

A Sentiment-annotated Dataset of English Causal Connectives
Marta Andersson, Murathan Kurfalı and Robert Östling 24

A Novel Annotation Schema for Conversational Humor: Capturing the Cultural Nuances in Kanyasulkam
Vaishnavi Pamulapati, Gayatri Purigilla and Radhika Mamidi 34

Modeling Ambiguity with Many Annotators and Self-Assessments of Annotator Certainty
Melanie Andresen, Michael Vauth and Heike Zinsmeister 48

Representation Problems in Linguistic Annotations: Ambiguity, Variation, Uncertainty, Error and Bias
Christin Beck, Hannah Booth, Mennatallah El-Assady and Miriam Butt 60

Understanding the Tradeoff between Cost and Quality of Expert Annotations for Keyphrase Extraction
Hung Chau, Saeid Balaneshin, Kai Liu and Ondrej Linda 74

Cookpad Parsed Corpus: Linguistic Annotations of Japanese Recipes
Jun Harashima and Makoto Hiramatsu 87

Modelling and Annotating Interlinear Glossed Text from 280 Different Endangered Languages as Linked Data with LIGT
Sebastian Nordhoff 93

PASTRIE: A Corpus of Prepositions Annotated with Supersense Tags in Reddit International English
Michael Kranzlein, Emma Manning, Siyao Peng, Shira Wein, Aryaman Arora and Nathan Schneider 105

Supersense and Sensibility: Proxy Tasks for Semantic Annotation of Prepositions
Luke Gessler, Shira Wein and Nathan Schneider 117

Sprucing up Supersenses: Untangling the Semantic Clusters of Accompaniment and Purpose
Jena D. Hwang, Nathan Schneider and Vivek Srikumar 127

Querent Intent in Multi-Sentence Questions
Laurie Burchell, Jie Chi, Tom Hosking, Nina Markl and Bonnie Webber 138

Annotating Errors and Emotions in Human-Chatbot Interactions in Italian
Manuela Sanguinetti, Alessandro Mazzei, Viviana Patti, Marco Scalerandi, Dario Mana and Rossana Simeoni 148

Towards a Standardized, Fine-grained Manual Annotation Protocol for Verbal Fluency Data
Gabriel McKee, Joël Macoir, Lydia Gagnon and Pascale Tremblay 160

pyMMAX2: Deep Access to MMAX2 Projects from Python
Mark-Christoph Müller 167

Annotating Coherence Relations for Studying Topic Transitions in Social Talk
Alex Luu and Sophia A. Malamud 174

Conference Program

All times are in Central European Time (CET)

Saturday, December 12, 2020

14:00–14:10 *Introduction to the Workshop*
LAW XIV Chairs

14:10–15:00 *Invited Talk*
Nils Reiter

15:00–15:20 ***Coffee Break***

15:20–16:00 Session 1

15:20–15:30 *Bayesian Methods for Semi-supervised Text Annotation*
Kristian Miok, Gregor Pirs and Marko Robnik-Sikonja

15:30–15:40 *Provenance for Linguistic Corpora through Nanopublications*
Timo Lek, Anna de Groot, Tobias Kuhn and Roser Morante

15:40–15:50 *A Sentiment-annotated Dataset of English Causal Connectives*
Marta Andersson, Murathan Kurfalı and Robert Östling

15:50–16:00 *A Novel Annotation Schema for Conversational Humor: Capturing the Cultural Nuances in Kanyasulkam*
Vaishnavi Pamulapati, Gayatri Purigilla and Radhika Mamidi

Saturday, December 12, 2020 (continued)

16:00–17:00 Poster Session 1

Modeling Ambiguity with Many Annotators and Self-Assessments of Annotator Certainty
Melanie Andresen, Michael Vauth and Heike Zinsmeister

Representation Problems in Linguistic Annotations: Ambiguity, Variation, Uncertainty, Error and Bias
Christin Beck, Hannah Booth, Mennatallah El-Assady and Miriam Butt

Understanding the Tradeoff between Cost and Quality of Expert Annotations for Keyphrase Extraction
Hung Chau, Saeid Balaneshin, Kai Liu and Ondrej Linda

Cookpad Parsed Corpus: Linguistic Annotations of Japanese Recipes
Jun Harashima and Makoto Hiramatsu

Modelling and Annotating Interlinear Glossed Text from 280 Different Endangered Languages as Linked Data with LIGT
Sebastian Nordhoff

17:00–17:30 ***Coffee Break***

17:30–18:00 Session 2

17:30–17:40 *PASTRIE: A Corpus of Prepositions Annotated with Supersense Tags in Reddit International English*
Michael Kranzlein, Emma Manning, Siyao Peng, Shira Wein, Aryaman Arora and Nathan Schneider

17:40–17:50 *Supersense and Sensibility: Proxy Tasks for Semantic Annotation of Prepositions*
Luke Gessler, Shira Wein and Nathan Schneider

17:50–18:00 *Sprucing up Supersenses: Untangling the Semantic Clusters of Accompaniment and Purpose*
Jena D. Hwang, Nathan Schneider and Vivek Srikumar

Saturday, December 12, 2020 (continued)

18:00–19:00 Poster Session 2

Querent Intent in Multi-Sentence Questions
Laurie Burchell, Jie Chi, Tom Hosking, Nina Markl and Bonnie Webber

Annotating Errors and Emotions in Human-Chatbot Interactions in Italian
Manuela Sanguinetti, Alessandro Mazzei, Viviana Patti, Marco Scalerandi, Dario Mana and Rossana Simeoni

Towards a Standardized, Fine-grained Manual Annotation Protocol for Verbal Fluency Data
Gabriel McKee, Joël Macoir, Lydia Gagnon and Pascale Tremblay

pyMMAX2: Deep Access to MMAX2 Projects from Python
Mark-Christoph Müller

Annotating Coherence Relations for Studying Topic Transitions in Social Talk
Alex Luu and Sophia A. Malamud

19:00–19:10 ***Closing Remarks***