

LAW 2023

**The 17th Linguistic Annotation Workshop (LAW-XVII)
@ ACL 2023**

Proceedings of the Workshop

July 13, 2023

The LAW organizers gratefully acknowledge the support from the following sponsors.



©2023 Association for Computational Linguistics

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 978-1-959429-83-8

Introduction

Linguistic annotation of natural language corpora is the backbone of supervised methods of statistical natural language processing. The Linguistic Annotation Workshop (LAW) is the annual workshop of the ACL Special Interest Group on Annotation (SIGANN), and it provides a forum for the presentation and discussion of innovative research on all aspects of linguistic annotation, including the creation and evaluation of annotation schemes, methods for automatic and manual annotation, use and evaluation of annotation software and frameworks, representation of linguistic data and annotations, semi-supervised “human in the loop” methods of annotation, crowd-sourcing approaches, and more. As in the past, this year’s LAW provides a forum for annotation researchers to work towards standardization, best practices, and interoperability of annotation information and software. These proceedings include papers that were presented at the 17th Linguistic Annotation Workshop (LAW-XVII), co-located with ACL 2023 in Toronto, Canada, on July 13, 2023.

This edition of the workshop is the seventeenth meeting of the ACL Special Interest Group for Annotation. 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, we received 51 submissions, out of which 26 papers have been accepted to be presented at the workshop. In addition, 9 papers accepted to the Findings of ACL 2023 have been invited to be presented at the LAW.

The papers presented at LAW-XVII cover phenomena in a diverse range of 23 languages: Ancient Greek, English, Turkish, German, Czech, Spanish, Bengali, Italian, Hungarian, French, Chinese, Arapaho, Arabic, Gujarati, Hebrew, Hausa, Hindi, Indonesian, Javanese, Kannada, Sundanese, Swahili, and Yoruba.

The special theme of LAW-XVII is *Ethics and Annotation*. The workshop includes a discussion session about various aspects of ethics related to annotation work, e.g., the treatment of annotators, psychological health of annotators, bias, ethics in crowd-sourcing annotation scenarios, or annotation of information regarding ethics in text. LAW-XVII also features invited talks by Emily Bender (University of Washington, USA), Anne Lauscher (University of Hamburg, Germany), and Lilian Wanzare (Maseno University, Kenya).

Our thanks go to SIGANN, our organizing committee, for their continuing organization of the LAW workshops, and to the ACL 2023 workshop chairs, Eduardo Blanco, Yang Feng, and Annie Louis, for their support. 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 extremely thoughtful reviews.

The LAW-XVII Program Co-Chairs:
Jakob Prange and Annemarie Friedrich

Organizing Committee

Program Chairs

Jakob Prange, The Hong Kong Polytechnic University
Annemarie Friedrich, University of Augsburg

SIGANN President

Amir Zeldes, Georgetown University

SIGANN Secretary

Ines Rehbein, Mannheim University

Remote Session Chairs

Heike Zinsmeister, Universität Hamburg
Djamé Seddah, Inria

SIGANN Officers

Claire Bonial, US Army Research Laboratory
Stefanie Dipper, Ruhr-Universität Bochum
Chu-Ren Huang, The Hong Kong Polytechnic University
Jena D. Hwang, Allen Institute for AI
Sandra Kübler, Indiana University
Lori Levin, Carnegie-Mellon University
Adam Meyers, New York University
Antonio Pareja-Lora, Universidad de Alcalá (UAH) / FITISPos (UAH) / ATLAS (UNED) / DMEG (UdG)
Massimo Poesio, Queen Mary University of London
Sameer Pradhan, LDC, Cemantix
Nancy Ide, Vassar College
Nathan Schneider, Georgetown University
Manfred Stede, Universität Potsdam
Karin Tomanek, Google
Fei Xia, University of Washington
Nianwen Xue, Brandeis University
Deniz Zeyrek, Middle East Technical University

Invited Speakers

Emily M. Bender, University of Washington
Lilian D. A. Wanzare, Maseno University
Anne Lauscher, University of Hamburg

Program Committee

Program Committee

Omri Abend, The Hebrew University of Jerusalem
Melanie Andresen, Universität Stuttgart
Aditya Bhargava, University of Toronto
Claire Bonial, US Army Research Lab
Miriam Butt, University of Konstanz
Emmanuele Chersoni, The Hong Kong Polytechnic University
Christian Chiarcos, Goethe-Universität Frankfurt am Main
Kathryn Conger, University of Colorado, Boulder
Daniel Dakota, Indiana University
Marie-Catherine De Marneffe, The Ohio State University
Stefanie Dipper, Ruhr-Universität Bochum
Lucia Donatelli, Vrije Universiteit Amsterdam
Jonathan Dunn, University of Canterbury
Markus Egg, Humboldt-Universität zu Berlin
Kim Gerdes, Paris-Saclay University
Luke Gessler, Georgetown University
Jinghang Gu, The Hong Kong Polytechnic University
Udo Hahn, Friedrich-Schiller-Universität Jena
Andrea Horbach, Universität Hildesheim
Jena D. Hwang, Allen Institute for AI
Nancy Ide, Vassar College/Brandeis University
Michael Kranzlein, Georgetown University
Sandra Kübler, Indiana University
Ekaterina Lapshinova-Koltunski, Stiftung Universität Hildesheim
Katja Markert, Heidelberg University
Adam Meyers, New York University
Philippe Muller, IRIT, University of Toulouse
Anna Nedoluzhko, Charles University in Prague
Simon Ostermann, German Research Center for Artificial Intelligence (DFKI)
Alexis Palmer, University of Colorado Boulder
Antonio Pareja-Lora, Universidad de Alcalá (UAH) / FITISPos (UAH) / ATLAS (UNED) / DMEG (UdG)
Siyao Peng, Georgetown University
Miriam R. L. Petruck, FrameNet
Barbara Plank, LMU Munich
Massimo Poesio, Queen Mary University of London
James Pustejovsky, Brandeis University
Ines Rehbein, University of Mannheim
Michael Roth, University of Stuttgart
Josef Ruppenhofer, FernUniversität in Hagen
Nathan Schneider, Georgetown University
Djamé Seddah, Inria
Ludovic Tanguy, CLLE: University of Toulouse & CNRS
Joel Tetreault, Dataminr
Lilian Diana Awuor Wanzare, Maseno University
Bonnie Webber, University of Edinburgh

Michael Wiegand, Alpen-Adria-Universitaet Klagenfurt
Fei Xia, University of Washington
Nianwen Xue, Brandeis University
Amir Zeldes, Georgetown University
Winnie Huiheng Zeng, The Hong Kong Polytechnic University
Deniz Zeyrek, Middle East Technical University
Heike Zinsmeister, Universität Hamburg

Table of Contents

<i>Medieval Social Media: Manual and Automatic Annotation of Byzantine Greek Marginal Writing</i> Colin Swaelens, Ilse De Vos and Els Lefever	1
<i>Orpheus Came to His End by Being Struck by a Thunderbolt": Annotating Events in Mythological Sequences</i> Franziska Pannach	10
<i>Difficulties in Handling Mathematical Expressions in Universal Dependencies</i> Lauren Levine	19
<i>A Dataset for Physical and Abstract Plausibility and Sources of Human Disagreement</i> Annerose Eichel and Sabine Schulte Im Walde	31
<i>Annotating and Disambiguating the Discourse Usage of the Enclitic <i>da</i> in Turkish</i> Ebru Ersöyleyen, Deniz Zeyrek and Firat Öter	46
<i>An Active Learning Pipeline for NLU Error Detection in Conversational Agents</i> Damian Pascual, Aritz Bercher, Akansha Bhardwaj, Mingbo Cui, Dominic Kohler, Liam Van Der Poel and Paolo Rosso	55
<i>Multi-layered Annotation of Conversation-like Narratives in German</i> Magdalena Repp, Petra B. Schumacher and Fahime Same	61
<i>Crowdsourcing on Sensitive Data with Privacy-Preserving Text Rewriting</i> Nina Mouhammad, Johannes Daxenberger, Benjamin Schiller and Ivan Habernal	73
<i>Extending an Event-type Ontology: Adding Verbs and Classes Using Fine-tuned LLMs Suggestions</i> Jana Straková, Eva Fučíková, Jan Hajič and Zdeňka Urešová	85
<i>Temporal and Second Language Influence on Intra-Annotator Agreement and Stability in Hate Speech Labelling</i> Gavin Abercrombie, Dirk Hovy and Vinodkumar Prabhakaran	96
<i>BenCoref: A Multi-Domain Dataset of Nominal Phrases and Pronominal Reference Annotations</i> Shadman Rohan, Mojammel Hossain, Mohammad Mamun Or Rashid and Nabeel Mohammed	104
<i>Annotators-in-the-loop: Testing a Novel Annotation Procedure on Italian Case Law</i> Emma Zanoli, Matilde Barbini, Davide Riva, Sergio Picascia, Emanuela Furiosi, Stefano D'Ancona and Cristiano Chesi	118
<i>Annotating Decomposition in Time: Three Approaches for Again</i> Martin Kopf and Remus Gergel	129
<i>How Good Is the Model in Model-in-the-loop Event Coreference Resolution Annotation?</i> Shafiuddin Rehan Ahmed, Abhijnan Nath, Michael Regan, Adam Pollins, Nikhil Krishnaswamy and James H. Martin	136
<i>Pragmatic Annotation of Articles Related to Police Brutality</i> Tess Feyen, Alda Mari and Paul Portner	146
<i>The RST Continuity Corpus</i> Debopam Das and Markus Egg	154

<i>GENTLE: A Genre-Diverse Multilayer Challenge Set for English NLP and Linguistic Evaluation</i> Tatsuya Aoyama, Shabnam Behzad, Luke Gessler, Lauren Levine, Jessica Lin, Yang Janet Liu, Siyao Peng, Yilun Zhu and Amir Zeldes	166
<i>A Study on Annotation Interfaces for Summary Comparison</i> Sian Gooding, Lucas Werner and Victor Cărbune	179
<i>A Question Answering Benchmark Database for Hungarian</i> Attila Novák, Borbála Novák, Tamás Zombori, Gergő Szabó, Zsolt Szántó and Richárd Farkas	188
<i>No Strong Feelings One Way or Another: Re-operationalizing Neutrality in Natural Language Inference</i> Animesh Nighojkar, Antonio Laverghetta Jr. and John Licato	199
<i>UMR-Writer 2.0: Incorporating a New Keyboard Interface and Workflow into UMR-Writer</i> Sijia Ge, Jin Zhao, Kristin Wright-bettner, Skatje Myers, Nianwen Xue and Martha Palmer ..	211
<i>Unified Syntactic Annotation of English in the CGEL Framework</i> Brett Reynolds, Aryaman Arora and Nathan Schneider	220
<i>Annotating Discursive Roles of Sentences in Patent Descriptions</i> Lufei Liu, Xu Sun, François Veltz and Kim Gerdes	235
<i>The Effect of Alignment Correction on Cross-Lingual Annotation Projection</i> Shabnam Behzad, Seth Ebner, Marc Marone, Benjamin Van Durme and Mahsa Yarmohammadi	244
<i>When Do Annotator Demographics Matter? Measuring the Influence of Annotator Demographics with the POPQUORN Dataset</i> Jiaxin Pei and David Jurgens	252
<i>Enriching the NArabizi Treebank: A Multifaceted Approach to Supporting an Under-Resourced Lan- guage</i> Arij Riabi, Menel Mahamdi and Djamé Seddah	266

Program

Thursday, July 13, 2023

08:45 - 09:00 *Opening Remarks*

09:00 - 09:45 *Invited Talk 1*

09:45 - 11:15 *Onsite/remote Poster Sessions 1 (Coffee served from 10:30 to 11:00)*

11:15 - 12:00 *Invited Talk 2*

12:00 - 14:00 *Lunch Break*

14:00 - 15:00 *Invited Talk 3*

15:00 - 15:30 *Theme Paper Highlights Session*

15:30 - 17:00 *Onsite/remote Poster Sessions 2 (Coffee served from 15:30 to 16:00)*

17:00 - 17:45 *Discussion Session*

17:45 - 18:00 *Closing Remarks*