LAW 2025

The 19th Linguistic Annotation Workshop (LAW-XIX)
Co-located with ACL 2025

Proceedings of the Workshop

July 31, 2025

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

©2025 The 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 979-8-89176-262-6

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 and ELRA 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 19th Linguistic Annotation Workshop (LAW-XIX), co-located with ACL 2025 in Vienna, Austria, on July 31, 2025.

This edition of the workshop is the nineteenth meeting of the ACL and ELRA 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 and trustworthy evaluation in the era of Large Language Models. This year, we received 66 submissions, out of which 30 papers have been accepted to be presented at the workshop, as long or short papers, or as posters.

In addition, LAW-XIX features two invited talks by Junyi Jessy Li (University of Texas at Austin) and Rotem Dror (Haifa University).

The special theme of LAW-XIX is "subjectivity and variation in linguistic annotations." As linguistic annotation increasingly supports diverse NLP applications, questions of annotator subjectivity, inter-annotator variation, and annotation uncertainty have become central to the field. Our special oral sessions aim to stimulate discussions on the challenges, methodological advances, and theoretical implications related to disagreement and subjectivity in annotation practice.

Our thanks go to SIGANN, our organizing committee, for their continuing organization of the LAW workshops. Most of all, we would like to thank all the authors for submitting their papers to the workshop, our program committee members for their dedication and their thoughtful reviews and our keynote speakers for sharing their insights on the topics of subjectivity, variation and social biases in linguistic annotation.

The LAW-XIX Program Co-Chairs:
Siyao Peng and Ines Rehbein

Organizing Committee

Program Chairs

Siyao Peng, LMU Munich
Ines Rehbein, Mannheim University

SIGANN President

Amir Zeldes, Georgetown University

SIGANN Secretary

Ines Rehbein, Mannheim University

SIGANN Officers

Claire Bonial, US Army Research Laboratory
Stefanie Dipper, Ruhr-Universität Bochum
Annemarie Friedrich, University of Augsburg
Sophie Henning, Bosch Center for Artificial Intelligence
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 (UAH) / FITISPos (UAH) / ATLAS (UNED) / DMEG (UdG)
Siyao Peng, LMU Munich
Massimo Poesio, Queen Mary University of London and University of Utrecht
Sameer Pradhan, LDC, Cemantix
Jakob Prange, University of Augsburg
Ines Rehbein, Mannheim University
Nathan Schneider, Georgetown University
Manfred Stede, University of Potsdam
Katrin Tomanek, Google
Fei Xia, University of Washington
Nianwen Xue, Brandeis University
Amir Zeldes, Georgetown University
Heike Zinsmeister, Hamburg University

Program Committee

Chairs

Siyao Peng, LMU Munich
Ines Rehbein, Mannheim University

Program Committee

Maria Becker, Heidelberg University
Verena Blaschke, LMU Munich
Claire Bonial, Army Research Lab
Miriam Butt, Konstanz University
Daniel Dakota, Indiana University
Marie-Catherine de Marneffe, FNRS - UCLouvain
Lucia Donatelli, Vrije Universiteit Amsterdam
Luise Dürlich, Uppsala University
Jonathan Dunn, University of Illinois, Urbana-Champaign
Pablo Picasso Feliciano de Faria, University of Campinas
Annemarie Friedrich, Augsburg University
Kim Gerdes, Université Paris-Saclay
Luke Gessler, Indiana University Bloomington
Ziwei Gong, Columbia University
Pingjun Hong, LMU Munich
Chengzhi Martin Hu, LMU Munich
Nancy Ide, Vassar College
Maxim Ionov, University of Cologne
Sandra Kübler, Indiana University
Ekaterina Lapshinova-Koltunski, University of Hildesheim
Ji-Ung Lee, Saarland University
Els Lefever, Ghent University
Lori Levin, Carnegie Mellon University
Lauren Levine, Georgetown University
Sebastian Loftus, LMU Munich
Abdou Mohamed Naira, SI2M Lab, INSEA
Anna Nedoluzhko, Charles University
Ercong Nie, LMU Munich
Juhyun Oh, KAIST
Teresa Paccosi, University of Trento
Alexis Palmer, University of Boulder
Antonio Pareja-Lora, University of Alcalá
Massimo Poesio, Utrecht University
Jakob Prange, Augsburg University
James Pustejovsky, Brandeis University
Sebastian Reimann, University of Jena
Josef Ruppenhofer, University of Hagen
Egil Rønningstad, University of Oslo
Tatjana Scheffler, University of Bochum
Nathan Schneider, Georgetown University
Merel Scholman, Utrecht University
Wesley Scivetti, Georgetown University
Indira Sen, Mannheim University

Soh-Eun Shim, LMU Munich
Manfred Stede, Potsdam University
Hakyung Sung, University of Oregon
Andreas Säuberli, LMU Munich
Ludovic Tanguy, CNRS & University of Toulouse
Joel Tetreault, Dataminr
Lilian Diana Awuor Wanzare, Maseno University
Bonnie Webber, University of Edinburgh
Leonie Weissweiler, UT Austin
Michael Wiegand, University of Vienna
Mohammad Yeghaneh Abkenar, Potsdam University
Karolina Zaczynska, Potsdam University
Amir Zeldes, Georgetown University
Deniz Zeyrek, Middle East Technical University
Wei Zhou, RWTH Aachen
Heike Zinsmeister, Hamburg University
Longfei Zuo, LMU Munich
Şaziye Betül Özateş, Boğaziçi University

Table of Contents

Understanding Disagreement: An Annotation Study of Sentiment and Emotional Language in Environmental Communication
Christina Barz, Melanie Siegel, Daniel Hanss and Michael Wiegand 1

Measuring Label Ambiguity in Subjective Tasks using Predictive Uncertainty Estimation
Richard Alies, Elena Merdjanovska and Alan Akbik 21

Disagreements in analyses of rhetorical text structure: A new dataset and first analyses
Freya Hewett and Manfred Stede 35

Subjectivity in the Annotation of Bridging Anaphora
Lauren Levine and Amir Zeldes 48

The revision of linguistic annotation in the Universal Dependencies framework: a look at the annotators' behavior
Magali Sanches Duran, Lucelene Lopes and Thiago Alexandre Salgueiro Pardo 60

Forbidden FRUIT is the Sweetest: An Annotated Tweets Corpus for French Unfrozen Idioms Identification
Julien Bezançon, Gaël Lejeune, Antoine Gautier, Marceau Hernandez and Félix Alié 70

Another Approach to Agreement Measurement and Prediction with Emotion Annotations
Quanqi Du and Veronique Hoste 86

Harmonizing Divergent Lemmatization and Part-of-Speech Tagging Practices for Latin Participles through the LiLa Knowledge Base
Marco Passarotti, Federica Iurescia and Paolo Ruffolo 102

UD-KSL Treebank v1.3: A semi-automated framework for aligning XPOS-extracted units with UPOS tags
Hakyung Sung, Gyu-Ho Shin, Chanyoung Lee, You Kyung Sung and Boo Kyung Jung 114

Bootstrapping UMRs from Universal Dependencies for Scalable Multilingual Annotation
Federica Gamba, Alexis Palmer and Daniel Zeman 125

Classifying TEI Encoding for DutchDraCor with Transformer Models
Florian Debaene and Veronique Hoste 136

Label Bias in Symbolic Representation of Meaning
Marie Mikulová, Jan Štěpánek and Jan Hajič 141

An Annotation Protocol for Diachronic Evaluation of Semantic Drift in Disability Sources
Nitisha Jain, Chiara Di Bonaventura, Albert Merono Penuela and Barbara McGillivray 159

Pre-annotation Matters: A Comparative Study on POS and Dependency Annotation for an Alsatian Dialect
Delphine Bernhard, Nathanaël Beiner and Barbara Hoff 172

Where it's at: Annotating Verb Placement Types in Learner Language
Josef Ruppenhofer, Annette Annette Portmann, Christine Renker, Matthias Schwendemann, Katrin Wisniewski and Torsten Zesch 186

ICLE-RC: International Corpus of Learner English for Relative Clauses
Debopam Das, Izabela Czerniak and Peter Bourgonje 200

Towards Resource-Rich Mizo and Khasi in NLP: Resource Development, Synthetic Data Generation and Model Building
Soumyadip Ghosh, Henry Lalsiam, Dorothy Marbaniang, Gracious Mary Temsen, Rahul Mishra and Parameswari Krishnamurthy 215

Creating Hierarchical Relations in a Multilingual Event-type Ontology
Zdeňka Urešová, Eva Fučíková and Jan Hajič 227

ExpLay: A new Corpus Resource for the Research on Expertise as an Influential Factor on Language Production
Carmen Schacht and Renate Delucchi Danhier 237

Visual Representations of Temporal Relations between Events and Time Expressions in News Stories
Evelin Amorim, António Leal, Nana Yu, Purificação Moura Silvano and Alipio Mario Jorge . . 249

Annotating candy speech in German YouTube comments
Yulia Clausen and Tatjana Scheffler 263

Variety delights (sometimes) - Annotation differences in morphologically annotated corpora
Andrea Dömötör, Balázs Indig and Dávid Márk Nemeskey 269

Addressing Variability in Interlinear Glossed Texts with Linguistic Linked Data
Maxim Ionov and Natalia Patiño Mazzotti 278

Illuminating Logical Fallacies with the CAMPFIRE Corpus
Austin Blodgett, Claire Bonial, Taylor A. Pellegrin, Melissa Torgbi and Harish Tayyar Madabushi 284

Cheap Annotation of Complex Information: A Study on the Annotation of Information Status in German TEDx Talks
Carmen Schacht, Tobias Nischk, Oleksandra Yazdanfar and Stefanie Dipper 296

Annotating Spatial Descriptions in Literary and Non-Literary Text
Emilie Sitter, Omar Momen, Florian Steig, J. Berenike Herrmann and Sina Zarrieß 307

A GitHub-based Workflow for Annotated Resource Development
Brandon Waldon and Nathan Schneider 325

Enhancing an Annotation Scheme for Clinical Narratives in Portuguese through Human Variation Analysis
Ana Luisa Fernandes, Purificação Silvano, António Leal, Nuno Guimarães, Rita Rb-Silva, Luís Filipe Cunha and Alípio Jorge 331

Expanding the UNSC Conflicts Corpus by Incorporating Domain Expert Annotations and LLM Experiments
Karolina Zaczynska 343

Guidelines for Fine-grained Sentence-level Arabic Readability Annotation
Nizar Habash, Hanada Taha-Thomure, Khalid Elmadani, Zeina Zeino and Abdallah Abushmaes 358

Workshop Program

Thursday, July 31, 2025

8:45–9:00 **Welcome**

9:00–9:45 **Keynote 1**

09:45–10:30 **Oral Session 1: Disagreement and ambiguity in highly subjective tasks**

09:45–10:00 *Understanding Disagreement: An Annotation Study of Sentiment and Emotional Language in Environmental Communication*
Christina Barz, Melanie Siegel, Daniel Hanss and Michael Wiegand

10:00–10:15 *Measuring Label Ambiguity in Subjective Tasks using Predictive Uncertainty Estimation*
Richard Alies, Elena Merdjanovska and Alan Akbik

10:15–10:30 *Disagreements in analyses of rhetorical text structure: A new dataset and first analyses*
Freya Hewett and Manfred Stede

10:30–11:00 ***Coffee***

11:00–11:30 **Oral Session 2: Subjectivity in linguistic annotation**

11:00–11:15 *Subjectivity in the Annotation of Bridging Anaphora*
Lauren Levine and Amir Zeldes

11:15–11:30 *The revision of linguistic annotation in the Universal Dependencies framework: a look at the annotators' behavior*
Magali Sanches Duran, Lucelene Lopes and Thiago Alexandre Salgueiro Pardo

Thursday, July 31, 2025 (continued)

11:30–12:30 Poster Session 1

Forbidden FRUIT is the Sweetest: An Annotated Tweets Corpus for French Unfrozen Idioms Identification
Julien Bezançon, Gaël Lejeune, Antoine Gautier, Marceau Hernandez and Félix Alié

Another Approach to Agreement Measurement and Prediction with Emotion Annotations
Quanqi Du and Veronique Hoste

Harmonizing Divergent Lemmatization and Part-of-Speech Tagging Practices for Latin Participles through the LiLa Knowledge Base
Marco Passarotti, Federica Iurescia and Paolo Ruffolo

UD-KSL Treebank v1.3: A semi-automated framework for aligning XPOS-extracted units with UPOS tags
Hakyung Sung, Gyu-Ho Shin, Chanyoung Lee, You Kyung Sung and Boo Kyung Jung

Bootstrapping UMRs from Universal Dependencies for Scalable Multilingual Annotation
Federica Gamba, Alexis Palmer and Daniel Zeman

Classifying TEI Encoding for DutchDraCor with Transformer Models
Florian Debaene and Veronique Hoste

Label Bias in Symbolic Representation of Meaning
Marie Mikulová, Jan Štěpánek and Jan Hajič

An Annotation Protocol for Diachronic Evaluation of Semantic Drift in Disability Sources
Nitisha Jain, Chiara Di Bonaventura, Albert Merono Penuela and Barbara McGillivray

Pre-annotation Matters: A Comparative Study on POS and Dependency Annotation for an Alsatian Dialect
Delphine Bernhard, Nathanaël Beiner and Barbara Hoff

Where it's at: Annotating Verb Placement Types in Learner Language
Josef Ruppenhofer, Annette Annette Portmann, Christine Renker, Matthias Schwendemann, Katrin Wisniewski and Torsten Zesch

ICLE-RC: International Corpus of Learner English for Relative Clauses
Debopam Das, Izabela Czerniak and Peter Bourgonje

Thursday, July 31, 2025 (continued)

12:30–14:00 ***Lunch***

14:00–15:00 **Poster Session 2**

Towards Resource-Rich Mizo and Khasi in NLP: Resource Development, Synthetic Data Generation and Model Building
Soumyadip Ghosh, Henry Lalsiam, Dorothy Marbaniang, Gracious Mary Temsen, Rahul Mishra and Parameswari Krishnamurthy

Creating Hierarchical Relations in a Multilingual Event-type Ontology
Zdeňka Urešová, Eva Fučíková and Jan Hajič

ExpLay: A new Corpus Resource for the Research on Expertise as an Influential Factor on Language Production
Carmen Schacht and Renate Delucchi Danhier

Visual Representations of Temporal Relations between Events and Time Expressions in News Stories
Evelin Amorim, António Leal, Nana Yu, Purificação Moura Silvano and Alipio Mario Jorge

Annotating candy speech in German YouTube comments
Yulia Clausen and Tatjana Scheffler

Variety delights (sometimes) - Annotation differences in morphologically annotated corpora
Andrea Dömötör, Balázs Indig and Dávid Márk Nemeskey

Addressing Variability in Interlinear Glossed Texts with Linguistic Linked Data
Maxim Ionov and Natalia Patiño Mazzotti

Illuminating Logical Fallacies with the CAMPFIRE Corpus
Austin Blodgett, Claire Bonial, Taylor A. Pellegrin, Melissa Torgbi and Harish Tayyar Madabushi

Cheap Annotation of Complex Information: A Study on the Annotation of Information Status in German TEDx Talks
Carmen Schacht, Tobias Nischk, Oleksandra Yazdanfar and Stefanie Dipper

Annotating Spatial Descriptions in Literary and Non-Literary Text
Emilie Sitter, Omar Momen, Florian Steig, J. Berenike Herrmann and Sina Zarrieß

Thursday, July 31, 2025 (continued)

A GitHub-based Workflow for Annotated Resource Development
Brandon Waldon and Nathan Schneider

Enhancing an Annotation Scheme for Clinical Narratives in Portuguese through Human Variation Analysis
Ana Luisa Fernandes, Purificação Silvano, António Leal, Nuno Guimarães, Rita Rb-Silva, Luís Filipe Cunha and Alípio Jorge

15:00–15:30 Oral Session 3: Datasets

15:00–15:15 *Expanding the UNSC Conflicts Corpus by Incorporating Domain Expert Annotations and LLM Experiments*
Karolina Zaczynska

15:15–15:30 *Guidelines for Fine-grained Sentence-level Arabic Readability Annotation*
Nizar Habash, Hanada Taha-Thomure, Khalid Elmadani, Zeina Zeino and Abdallah Abushmaes

15:30–16:00 ***Coffee***

16:00–16:45 Keynote 2

16:45–17:00 Wrap-up