

LAW 2026

The 20th Linguistic Annotation Workshop

Proceedings of the Workshop

July 3, 2026

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



©2026 Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
317 Sidney Baker St. S
Suite 400 - 134
Kerrville, TX 78028
USA
Tel: +1-855-225-1962
acl@aclweb.org

ISBN 979-8-89176-404-0

Preface

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, the LAW will provide 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 20th Linguistic Annotation Workshop (LAW-XX), co-located with ACL 2026 in San Diego, California, United States, on July 3, 2026.

This edition of the workshop is the twentieth meeting of the ACL and ELRA Special Interest Group for Annotation. The first LAW 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). This sustained engagement provides evidence that the LAW’s overall focus on annotation and data quality 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 41 direct submissions, out of which 20 papers have been accepted to be presented at the workshop, as long or short papers, or as posters.

In addition, LAW-XX features two invited talks by Marie-Catherine de Marneffe (UCLouvain) and Rachel Rudinger (University of Maryland), as well as a roundtable discussion that brings together diverse perspectives on open questions and future directions relevant to the workshop and its special theme.

The special theme of LAW-XX is *Errors in Annotation*. The special oral session highlights research on identifying and addressing annotation errors arising from objectively incorrect labels, annotations, or omissions, as distinct from cases of annotator disagreement. It also examines the role of error analysis in improving the quality of both human-annotated and LLM-generated datasets. As the field increasingly relies on large-scale annotated data, ensuring data quality has become a critical concern. Consequently, there is a growing need for methods, metrics, and tools that can systematically assess, quantify, and improve the quality of annotations across diverse data sources.

We extend our thanks to SIGANN for its financial support, and to the organizing committee for its continued dedication to the LAW workshops. We are also grateful to the ACL Workshop Office, Zeerak Talat, and the ACL 2026 Workshop Co-Chairs, Loic Barrault and Yang Zhao, for their guidance and support throughout the organization of the workshop. Above all, we would like to thank all the authors for submitting their papers to the workshop and our program committee members for their time, expertise, and thoughtful reviews, which made this workshop possible.

The LAW-XX Program Co-Chairs:
Yang Janet Liu and Luke Gessler

Organizing Committee

Program Chairs

Yang Janet Liu, University of Pittsburgh
Luke Gessler, Indiana University

Organizing Committee

Amir Zeldes, Georgetown University
Ines Rehbein, Mannheim University

Keynote Talk

Marie-Catherine de Marneffe
FNRS - UCLouvain - CENTAL
2026-07-03 09:00:00 – Room: **La Jolla**

Bio: Marie-Catherine de Marneffe is a FNRS research associate and professor at UCLouvain. She obtained her PhD under the supervision of Chris Manning at Stanford University and worked 10 years in the Linguistics department at The Ohio State University as assistant then associate professor. Her main research interests are in computational pragmatics, building models that capture what people infer “between the lines”. Her 2012 dissertation highlighted the need to capture variation in human annotations. She is also one of the developers of the Universal Dependencies framework. Her research work has been funded by Google Inc., the National Science Foundation and the FNRS.

Keynote Talk

Rachel Rudinger

University of Maryland

2026-07-03 16:00:00 – Room: La Jolla

Bio: Rachel Rudinger is an Assistant Professor of Computer Science at the University of Maryland. Her research interests lie in the areas of natural language understanding, commonsense reasoning, computational semantics/pragmatics, and issues of sociocultural fairness in NLP systems. She is a recipient of the National Science Foundation CAREER Award and holds a Ph.D. in Computer Science from Johns Hopkins University.

Table of Contents

<i>Annotating Clinical Risk and Variation in Haitian Creole Medical Translation</i> Ludovic Mompelat, David Tézil and Rose Flaure Accilien	1
<i>Parser agreement and disagreement in L2 Korean UD: Implications for human-in-the-loop annotation</i> Hakyung Sung and Gyu-Ho Shin	12
<i>Rules-based system for Czech legal text readability</i> Kateřina Motalík Hodková, Ivan Kraus and Barbora Hladká	22
<i>Human-AI Annotation Error Auditing for Hebrew Diacritization with Frontier LLMs</i> Hillel Gershuni and Avi Shmidman	33
<i>Beyond Annotator Disagreement: Guideline-Induced Errors in Arabic Hate Speech Annotation</i> Wajdi Zaghouani	47
<i>When LLMs Disagree with Human Experts: Understanding LLM Annotation Failures in Nutrition Misinformation through Hierarchical Error Analysis using Seed Oil Narratives</i> Vishwaa Shah, Indika Kahanda and Andrea Arikawa	59
<i>Math-DB: A Discourse Framework for Mathematical Word Problems to Enhance LLM Reasoning</i> Mustafa Erolcan Er	75
<i>Cross-Linguistic Situation Entity Segmentation for Discourse Analysis in Diachronic English and German Text</i> Hanna Schmück, Veronika Urban, Xaver Krückl, Sonja Zeman, Claudia Claridge and Annemarie Friedrich	95
<i>UD-CHILDES-BG: a dependency treebank of Bulgarian child and child-directed speech</i> Mila Marcheva-Nash, Yasena Chantova, Tsvetina Kirilova, Ivelina Pavlova, Tsvetelina Stefanova, Yoana Vasileva and Weiwei Sun	113
<i>IndiAnn: A Web-based Annotation Platform for Indic Languages</i> Bandaru Lavadeep, Ritwik Raghav and Abhik Jana	130
<i>Designing Annotation Guidelines for Trait-Based Arabic Automated Essay Scoring: A Systematic Methodology</i> Walid Massoud, Houda Bouamor, Abdelrahman Abdel Latif Hussein and Abdullah Mohamed Mohamed Zekri	146
<i>Revisiting Faithfulness Annotations for Long-form Summaries</i> Yang Zhong, Yang Janet Liu and Diane Litman	158
<i>Completing and Validating the Re-Aligned Switchboard Dialog Act Corpus</i> Run Chen, Zihao Tao, John Prado, Ignazio LaManna, Ryan Puterbaugh, Mim Datta and Julia Hirschberg	173
<i>Not Worth Mentioning? A Pilot Study on Salient Proposition Annotation</i> Amir Zeldes, Katherine Conhaim and Lauren Levine	178
<i>LLMs as annotators of credibility assessment in Danish asylum decisions: evaluating classification performance and errors beyond aggregated metrics</i> Galadrielle Humblot-Renaux, Mohammad N. S. Jahromi, Rohat Bakuri-Jørgensen, Marieke Anne Heyl, Asta S. Stage Jarlner, Maria Vlachou, Anna Murphy Høgenhaug, Desmond Elliott, Thomas Gammeltoft-Hansen and Thomas B. Moeslund	187

<i>Cracks in the Bridge—or A Bridge Too Far? Comparing Human and LLM Errors in the Annotation of Bridging Anaphora</i>	
Lauren Levine and Amir Zeldes	219
<i>Clustering Analysis for Error Detection in Named Entity Recognition Datasets</i>	
Matthew Flynn, Timothy Obiso, Sam Newman and Constantine Lignos	229
<i>When Ground Truth Disagrees: A Human-in-the-Loop Audit of Annotation Errors in High-Stakes Crash Narratives</i>	
Md Sajjad Hossain, Lin Li, Judy A. Perkins, John Clary and Joel Meyer	241
<i>Prompts in the Wild: A Large Analyzed Collection of Transactional Prompts in Code</i>	
Victoria Basmov, Yoav Goldberg and Reut Tsarfaty	257
<i>TalkTag: Fine-Grained Morphosyntactic Error Annotation for Transcribed Speech</i>	
Shamira Venturini, Oliver Hennhöfer, Steffen Kinkel and Jannik Strötgen	309

Program

Friday, July 3, 2026

08:45 - 09:00 *Opening Remarks*

09:00 - 09:45 *Keynote 1: Marie-Catherine de Marneffe*

09:45 - 10:30 *Talks*

Human-AI Annotation Error Auditing for Hebrew Diacritization with Frontier LLMs

Hillel Gershuni and Avi Shmidman

LLMs as annotators of credibility assessment in Danish asylum decisions: evaluating classification performance and errors beyond aggregated metrics

Galadrielle Humblot-Renaux, Mohammad N. S. Jahromi, Rohat Bakuri-Jørgensen, Marieke Anne Heyl, Asta S. Stage Jarlner, Maria Vlachou, Anna Murphy Høgenhaug, Desmond Elliott, Thomas Gammeltoft-Hansen and Thomas B. Moeslund

Revisiting Faithfulness Annotations for Long-form Summaries

Yang Zhong, Yang Janet Liu and Diane Litman

10:30 - 11:15 *Coffee Break*

11:15 - 12:00 *Talks*

UD-CHILDES-BG: a dependency treebank of Bulgarian child and child-directed speech

Mila Marcheva-Nash, Yasena Chantova, Tsvetina Kirilova, Ivelina Pavlova, Tsvetelina Stefanova, Yoana Vasileva and Weiwei Sun

When Ground Truth Disagrees: A Human-in-the-Loop Audit of Annotation Errors in High-Stakes Crash Narratives

Md Sajjad Hossain, Lin Li, Judy A. Perkins, John Clary and Joel Meyer

Rules-based system for Czech legal text readability

Kateřina Motalík Hodková, Ivan Kraus and Barbora Hladká

12:00 - 13:30 *Lunch*

13:30 - 14:15 *Talks*

Friday, July 3, 2026 (continued)

TalkTag: Fine-Grained Morphosyntactic Error Annotation for Transcribed Speech

Shamira Venturini, Oliver Hennhöfer, Steffen Kinkel and Jannik Strötgen

Semantic-pragmatic Annotations in the Prague Dependency Treebank

Marie Mikulová, Eva Hajičová, Jiří Mírovský, Anna Nedoluzhko, Michal Novák, Pavlína Synková, Jan Štěpánek, Barbora Štěpánková and Jan Hajič

Not Worth Mentioning? A Pilot Study on Salient Proposition Annotation

Amir Zeldes, Katherine Conhaim and Lauren Levine

14:15 - 14:30 *Short Break*

14:30 - 15:30 *Posters*

Designing Annotation Guidelines for Trait-Based Arabic Automated Essay Scoring: A Systematic Methodology

Walid Massoud, Houda Bouamor, Abdelrahman Abdel Latif Hussein and Abdullah Mohamed Mohamed Zekri

Cracks in the Bridge—or A Bridge Too Far? Comparing Human and LLM Errors in the Annotation of Bridging Anaphora

Lauren Levine and Amir Zeldes

Prompts in the Wild: A Large Analyzed Collection of Transactional Prompts in Code

Victoria Basmov, Yoav Goldberg and Reut Tsarfaty

When LLMs Disagree with Human Experts: Understanding LLM Annotation Failures in Nutrition Misinformation through Hierarchical Error Analysis using Seed Oil Narratives

Vishwaa Shah, Indika Kahanda and Andrea Arikawa

Math-DB: A Discourse Framework for Mathematical Word Problems to Enhance LLM Reasoning

Mustafa Erolcan Er

Parser agreement and disagreement in L2 Korean UD: Implications for human-in-the-loop annotation

Hakyung Sung and Gyu-Ho Shin

Completing and Validating the Re-Aligned Switchboard Dialog Act Corpus

Run Chen, Zihao Tao, John Prado, Ignazio LaManna, Ryan Puterbaugh, Mim Datta and Julia Hirschberg

Friday, July 3, 2026 (continued)

Clustering Analysis for Error Detection in Named Entity Recognition Datasets

Matthew Flynn, Timothy Obiso, Sam Newman and Constantine Lignos

Annotating Clinical Risk and Variation in Haitian Creole Medical Translation

Ludovic Mompelat, David Tézil and Rose Flaure Accilien

EVADE: LLM-Based Explanation Generation and Validation for Error Detection in NLI

Longfei Zuo, Barbara Plank and Siyao Peng

IndiAnn: A Web-based Annotation Platform for Indic Languages

Bandaru Lavadeep, Ritwik Raghav and Abhik Jana

Beyond Annotator Disagreement: Guideline-Induced Errors in Arabic Hate Speech Annotation

Wajdi Zaghouni

Cross-Linguistic Situation Entity Segmentation for Discourse Analysis in Diachronic English and German Text

Hanna Schmück, Veronika Urban, Xaver Krückl, Sonja Zeman, Claudia Claridge and Annemarie Friedrich

15:30 - 16:00 *Coffee Break*

16:00 - 16:45 *Keynote 2: Rachel Rudinger*

16:45 - 17:15 *Roundtable Discussion: Open Questions & Future Directions*

17:15 - 17:30 *Closing Remarks*