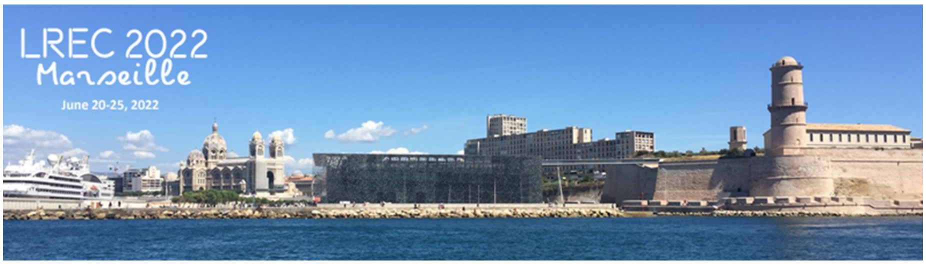

LREC 2022 Workshop
Language Resources and Evaluation Conference
20-25 June 2022

The 16th Linguistic Annotation Workshop
24 June 2022
(LAW-XVI)

PROCEEDINGS

Editors:
Sameer Pradhan
Sandra Kübler

Proceedings of the 16th Linguistic Annotation Workshop (LAW-XVI 2022)

Edited by:
Sameer Pradhan and Sandra Kübler

ISBN: 978-2-493814-08-1)
EAN: 9782493814081

For more information:
European Language Resources Association (ELRA)
9 rue des Cordelières
75013, Paris
France
http://www.elra.info
Email: lrec@elda.org

Message from the Workshop Organizers

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 XVI, held in conjunction with the 13th LREC in Marseille, France, on June 24, 2022.

The series is now in its sixteenth 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 28 submissions, out of which 20 papers have been accepted to be presented at the workshop.

In addition to oral and poster paper presentations, LAW XVI also features a panel on this year's special theme—*The Impact of Multimodal Language Understanding on Annotation Practices and Representations.* Recent years have seen rapid improvements in performance of machine learning models across multiple modalities of communication such as, text, speech, images, video, gestures, etc. Improvements in unsupervised representation and learning have resulted in state of the art models needing less manually annotated data for training. However, the need for high quality, manual annotations for capturing multiple layers of information surrogates across various signals, including linguistic, is unlikely to go away. On the contrary, annotation practices, guidelines and representations will need to be adapted, extended, to address the challenges brought about by a richer landscape of phenomena. Historically these communities have existed as separate islands, and have crafted solutions that satisfy local research and application needs. The evolution of next generation, situated language understanding systems is likely to create a greater demand on the availability, and ease of use of such multimodal annotations and frameworks.

Our thanks go to SIGANN, our organizing committee, for its continuing organization of the LAW workshops, and to the LREC 2022 workshop chairs 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 thoughtful reviews.

— Sandra Kübler and Sameer Pradhan

Organizers

Sameer Pradhan (University of Pennsylvania and `cemantix.org`, USA)
Sandra Kübler (Indiana University, USA)
Ines Rehbein (University of Mannheim, Germany)
Amir Zeldes (Georgetown University, USA)

Program Committee:

Julia Bonn (University of Colorado, Boulder, USA)
Santiago Arróniz (Indiana University, USA)
Emmanuele Chersoni (Hong Kong Polytechnic University)
Jonathan Dunn (University of Canterbury, New Zealand)
Kilian Evang (Heinrich-Heine University Düsseldorf, Germany)
Annemarie Friedrich (Bosch, Germany)
Kim Gerdes (Université Paris-Saclay, France)
Jena D. Hwang (Allen Institute for AI, USA)
Nancy Ide (Vassar College, USA)
Mikel Iruskieta (University of the Basque Country)
John Lee (City University of Hong Kong)
Adam Meyers (New York University, USA)
Jiří Mírovský (Charles University, Czech Republic)
Philippe Muller (Institut de Recherche en Informatique de Toulouse, France)
Skatje Myers (University of Colorado, Boulder, USA)
Kemal Oflazer (Carnegie Mellon University, Qatar)
Maciej Ogrodniczuk (Polish Academy of Sciences, Poland)
Antonio Pareja-Lora (Universidad de Alcalá de Henares, Spain)
Miriam R.L. Petruck (ICSI, USA)
Michael Roth (University of Stuttgart, Germany)
Manfred Stede (University of Potsdam, Germany)
Daniel Swanson (Indiana University, USA)
Bonnie Webber (University of Edinburgh, USA)
Michael Wiegand (Alpen-Adria-Universität Klagenfurt, Austria)
Fei Xia (University of Washington, USA)
Nianwen Xue (Brandeis University, USA)
Deniz Zeyrek (Middle East Technical University, Turkey)
He Zhou (Indiana University, USA)
Heike Zinsmeister (University of Hamburg, Germany)
Yilun Zhu (Georgetown University, USA)

Table of Contents

Workshop Program

Friday, June 24, 2022

8:45–9:00 ***Opening Remarks***

9:00–10:30 ***Session I—Paper Presentations***

9:00–9:15 *Automatic Approach for Building Dataset of Citation Functions for COVID-19 Academic Papers*
Setio Basuki and Masatoshi Tsuchiya

9:15–9:30 *The Development of a Comprehensive Spanish Dictionary for Phonetic and Lexical Tagging in Socio-phonetic Research (ESPADA)*
Simon Gonzalez

9:30–9:50 *Extending the SSJ Universal Dependencies Treebank for Slovenian: Was it Worth it?*
Kaja Dobrovoljc and Nikola Ljubešić

9:50–10:10 *Converting the Sinica Treebank of Mandarin Chinese to Universal Dependencies*
Yu-Ming Hsieh, Yueh-Yin Shih and Wei-Yun Ma

10:10–10:30 *Desiderata for the Annotation of Information Structure in Complex Sentences*
Hannah Booth

10:30–11:00 ***Coffee Break***

11:00–11:40 ***Session II—Paper Presentations***

11:00–11:20 *The Sensitivity of Annotator Bias to Task Definitions in Argument Mining*
Terne Sasha Thorn Jakobsen, Maria Barrett, Anders Søgaard and David Lassen

11:20–11:40 *NLP in Human Rights Research: Extracting Knowledge Graphs About Police and Army Units and Their Commanders*
Daniel Bauer, Tom Longley, Yueen Ma and Tony Wilson

Friday, June 24, 2022 (continued)

11:40–12:40 ***Session III—Posters***

Advantages of a complex multilayer annotation scheme: The case of the Prague Dependency Treebank
Eva Hajicova, Marie Mikulová, Barbora Štěpánková and Jiří Mírovský

Introducing StarDust: A UD-based Dependency Annotation Tool
Arife B. Yenice, Neslihan Cesur, Aslı Kuzgun and Olcay Taner Yıldız

Annotation of Messages from Social Media for Influencer Detection
Kevin Deturck, Damien Nouvel, Namrata Patel and Frédérique Segond

Charon: a FrameNet Annotation Tool for Multimodal Corpora
Frederico Belcavello, Marcelo Viridiano, Ely Matos and Tiago Timponi Torrent

Effect of Source Language on AMR Structure
Shira Wein, Wai Ching Leung, Yifu Mu and Nathan Schneider

12:40–14:00 ***Lunch Break***

14:00–15:00 ***Session IV—Panel on Annotating Multimodality***

15:00–16:00 ***Session V—Paper Presentations (Long; in Person)***

15:00–15:20 *Midas Loop: A Prioritized Human-in-the-Loop Annotation for Large Scale Multilayer Data*
Luke Gessler, Lauren Levine and Amir Zeldes

15:20–15:40 *How "Loco” is the LOCO Corpus?Annotating the Language of Conspiracy Theories*
Ludovic Mompelat, Zuoyu Tian, Amanda Kessler, Matthew Luettgen, Aaryana Rajanala, Sandra Kübler and Michelle Seelig

15:40–16:00 *Putting Context in SNACS: A 5-Way Classification of Adpositional Pragmatic Markers*
Yang Janet Liu, Jena D. Hwang, Nathan Schneider and Vivek Srikumar

16:00–16:30 ***Coffee Break***

16:30–18:10 ***Session VI—Paper Presentations (Long; Virtual)***

16:30–16:50 *Building a Biomedical Full-Text Part-of-Speech Corpus Semi-Automatically*
Nicholas Elder, Robert E. Mercer and Sudipta Singha Roy

16:50–17:10 *Human Schema Curation via Causal Association Rule Mining*
Noah Weber, Anton Belyy, Nils Holzenberger, Rachel Rudinger and Benjamin Van Durme

17:10–17:30 *A Cognitive Approach to Annotating Causal Constructions in a Cross-Genre Corpus*
Angela Cao, Gregor Williamson and Jinho D. Choi

17:30–17:50 *Automatic Enrichment of Abstract Meaning Representations*
Yuxin Ji, Gregor Williamson and Jinho D. Choi

17:50–18:10 *GRAIL—Generalized Representation and Aggregation of Information Layers*
Sameer Pradhan and Mark Liberman