

LAW XIII

## **The 13th Linguistic Annotation Workshop**



### **Proceedings of the Workshop**

August 1, 2019  
Florence, Italy

©2019 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](mailto:acl@aclweb.org)

ISBN 978-1-950737-38-3

## **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 harmonisation 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 XIII, held in conjunction with the annual meeting of the Association for Computational Linguistics (ACL) in Florence, Italy, on August 1, 2019.

The series is now in its thirteenth 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.

This year's LAW has received 52 submissions, out of which 28 papers have been accepted to be presented at the workshop, 10 as talks and 18 as posters. In addition to oral and poster paper presentations, LAW XIII also features an invited talk by Rebecca Passonneau and a discussion session.

Our thanks go to SIGANN, our organizing committee, for its continuing organization of the LAW workshops, and to the ACL 2019 workshop chairs for their support. Also, we thank Jet Hoek, the LAW XIII publications chair, for her 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: Marking of information quality in discourse**

This special theme considers the marking of information quality in discourse, i.e., annotations that mark how the speaker/writer expresses assessments. These assessments may be explicit and/or implicit in discourse, and may reflect positions, beliefs, opinions, appraisals and/or assessments about written or spoken propositions, for example, how a politician shows in discourse the degree of truthfulness in one of his/her electoral promises, or how a reporter shows his/her degree of belief in what the politician stated. This might include the annotation of devices such as hedges (“Donald claims that the crowd size, if you can really trust him to measure it, was enormous.”), committed belief (“The winners of the contest will be announced tomorrow.”) or attitudes (“It is with great sadness that we have learnt about the death of 6 people in the accident.”).

**Annemarie Friedrich and Deniz Zeyrek**  
Workshop chairs



**Organizers:**

Annemarie Friedrich, Bosch Research  
Deniz Zeyrek, Middle East Technical University, Ankara

**Publications chair:**

Jet Hoek, The University of Edinburgh

**Invited Speaker:**

Rebecca Passonneau, Penn State University

**Organizing Committee:**

Stefanie Dipper, Ruhr University Bochum  
Annemarie Friedrich, Bosch Research  
Chu-Ren Huang, The Hong Kong Polytechnic University  
Nancy Ide, Vassar College  
Lori Levin, Carnegie Mellon University  
Adam Meyers, New York University  
Antonio Pareja-Lora, Universidad Complutense de Madrid / ATLAS, UNED  
Massimo Poesio, Queen Mary University of London  
Sameer Pradhan, Boulder Learning, Inc.  
Ines Rehbein, Leibniz Science Campus, Institute for German Language and Heidelberg University  
Manfred Stede, University of Potsdam  
Katrín Tomanek, Google  
Fei Xia, University of Washington  
Heike Zinsmeister, University of Hamburg

**Program Committee:**

Adam Meyers, New York University  
Alexis Palmer, University of North Texas  
Amir Zeldes, Georgetown University  
Amália Mendes, University of Lisbon  
Andrea Horbach, University of Duisburg-Essen  
Ann Bies, Linguistic Data Consortium  
Anna Nedoluzhko, Charles University Prague  
Antonio Pareja-Lora, Universidad Complutense de Madrid  
Archana Bhatia, Florida Institute for Human and Machine Cognition  
Arndt Riestler, University of Stuttgart  
Bonnie Webber, The University of Edinburgh  
Caroline Sporleder, University of Göttingen  
Chloé Braud, Loria  
Christian Chiarcos, Goethe University Frankfurt  
Djamé Seddah, University Paris-Sorbonne  
Ellen Dodge, International Computer Science Institute  
Els Lefever, Ghent University

Emmanuele Chersoni, The Hong Kong Polytechnic University  
Éva Mújdricza-Maydt, Heidelberg University  
Federico Fancellu, Samsung Artificial Intelligence Center  
Fei Xia, University of Washington  
Heike Zinsmeister, Hamburg University  
Ines Rehbein, Leibniz Science Campus, Institute for German Language and Heidelberg University  
Jena D. Hwang, Institute for Human & Machine Cognition  
Jinghang Gu, Soochow University  
Jiří Mírovský, Charles University Prague  
John Lee, City University of Hong Kong  
Jonathan Dunn, University of Canterbury  
Kathryn Conger, University of Colorado Boulder  
Kemal Oflazer, Carnegie Mellon University in Qatar  
Kilian Evang, University of Düsseldorf  
Kim Gerdes, University Paris-Sorbonne  
Kristen Wright-Bettner, University of Colorado Boulder  
Lilja Øvrelid, University of Oslo  
Lori Levin, Carnegie Mellon University  
Maciej Ogrodniczuk, Institute of Computer Science, Polish Academy of Sciences  
Manfred Stede, University of Potsdam  
Marie-Catherine de Marneffe, The Ohio State University  
Massimo Poesio, Queen Mary University of London  
Michael Wiegand, Heidelberg University  
Michael Roth, University of Stuttgart  
Mingyu WAN, City University of Hong Kong  
Miriam R.L. Petruck, International Computer Science Institute  
Nancy Ide, Vassar College  
Nathan Schneider, Georgetown University  
Nianwen Xue, Brandeis University  
Nicoletta Calzolari, Italian National Research Council  
Omri Abend, The Hebrew University of Jerusalem  
Özlem Çetinoğlu, University of Stuttgart  
Pablo Picasso Feliciano de Faria, University of Campinas  
Phillippe Muller, IRIT, Université de Toulouse  
Reut Tsarfaty, Open University of Israel  
Ron Artstein, USC Institute for Creative Technologies  
Sandra Kübler, Indiana University  
Simon Ostermann, Saarland University  
Stefanie Dipper, Ruhr University Bochum  
Susan Windisch Brown, University of Colorado Boulder  
Tim O’Gorman, University of Colorado Boulder  
Udo Hahn, University of Jena

## Table of Contents

<i>Crowdsourced Hedge Term Disambiguation</i> Morgan Ulinski and Julia Hirschberg .....	1
<i>WiRe57 : A Fine-Grained Benchmark for Open Information Extraction</i> William Lechelle, Fabrizio Gotti and Phillippe Langlais .....	6
<i>Crowdsourcing Discourse Relation Annotations by a Two-Step Connective Insertion Task</i> Frances Yung, Vera Demberg and Merel Scholman .....	16
<i>Annotating and Analyzing the Interactions between Meaning Relations</i> Darina Gold, Venelin Kovatchev and Torsten Zesch .....	26
<i>CCGweb: a New Annotation Tool and a First Quadrilingual CCG Treebank</i> Kilian Evang, Lasha Abzianidze and Johan Bos .....	37
<i>The Making of the Litkey Corpus, a Richly Annotated Longitudinal Corpus of German Texts Written by Primary School Children</i> Ronja Laarmann-Quante, Stefanie Dipper and Eva Belke .....	43
<i>The Materials Science Procedural Text Corpus: Annotating Materials Synthesis Procedures with Shallow Semantic Structures</i> Sheshera Mysore, Zachary Jensen, Edward Kim, Kevin Huang, Haw-Shiuan Chang, Emma Strubell, Jeffrey Flanigan, Andrew McCallum and Elsa Olivetti .....	56
<i>Tagging Modality in Oceanic Languages of Melanesia</i> Annika Tjuka, Lena Weißmann and Kilu von Prince .....	65
<i>Harmonizing Different Lemmatization Strategies for Building a Knowledge Base of Linguistic Resources for Latin</i> Francesco Mambrini and Marco Passarotti .....	71
<i>Assessing Back-Translation as a Corpus Generation Strategy for non-English Tasks: A Study in Reading Comprehension and Word Sense Disambiguation</i> Fabricio Monsalve, Kervy Rivas Rojas, Marco Antonio Sobrevilla Cabezudo and Arturo Oncevay .....	81
<i>A Framework for Annotating ‘Related Works’ to Support Feedback to Novice Writers</i> Arlene Casey, Bonnie Webber and Dorota Glowacka .....	90
<i>An Online Annotation Assistant for Argument Schemes</i> John Lawrence, Jacky Visser and Chris Reed .....	100
<i>Annotating Formulaic Sequences in Spoken Slovenian: Structure, Function and Relevance</i> Kaja Dobrovoljc .....	108
<i>Annotating Information Structure in Italian: Characteristics and Cross-Linguistic Applicability of a QUD-Based Approach</i> Kordula De Kuthy, Lisa Brunetti and Marta Berardi .....	113
<i>DEFT: A Corpus for Definition Extraction in Free- and Semi-Structured Text</i> Sasha Spala, Nicholas A. Miller, Yiming Yang, Franck Dernoncourt and Carl Dockhorn .....	125

<i>Explaining Simple Natural Language Inference</i>	
Aikaterini-Lida Kalouli, Annebeth Buis, Livy Real, Martha Palmer and Valeria dePaiva . . . . .	133
<i>On the Role of Discourse Relations in Persuasive Texts</i>	
Ines Rehbein . . . . .	145
<i>One Format to Rule Them All – The Emtsv Pipeline for Hungarian</i>	
Balázs Indig, Bálint Sass, Eszter Simon, Iván Mittelholcz, Noémi Vadász and Márton Makrai .	156
<i>Turkish Treebanking: Unifying and Constructing Efforts</i>	
Utku Türk, Furkan Atmaca, Şaziye Betül Özateş, Abdullatif Köksal, Balkiz Ozturk Basaran, Tunga Gungor and Arzucan Özgür . . . . .	167
<i>A Dataset for Semantic Role Labelling of Hindi-English Code-Mixed Tweets</i>	
Riya Pal and Dipti Sharma . . . . .	179
<i>A Multi-Platform Annotation Ecosystem for Domain Adaptation</i>	
Richard Eckart de Castilho, Nancy Ide, Jin-Dong Kim, Jan-Christoph Klie and Keith Suderman	190
<i>A New Annotation Scheme for the Sejong Part-of-speech Tagged Corpus</i>	
Jungyeul Park and Francis Tyers . . . . .	196
<i>A Turkish Dataset for Gender Identification of Twitter Users</i>	
Erhan Sezerer, Ozan Polatbilek and Selma Tekir . . . . .	204
<i>Comparative Judgments are More Consistent than Binary Classification for Labelling Word Complexity</i>	
Sian Gooding, Ekaterina Kochmar, Advait Sarkar and Alan Blackwell . . . . .	209
<i>Continuous Quality Control and Advanced Text Segment Annotation with WAT-SL 2.0</i>	
Christina Lohr, Johannes Kiesel, Stephanie Luther, Johannes Hellrich, Tobias Kolditz, Benno Stein and Udo Hahn . . . . .	216
<i>Creation of a Corpus with Semantic Role Labels for Hungarian</i>	
Attila Novák, László Laki, Borbála Novák, Andrea Dömötör, Noémi Ligeti-Nagy and Ágnes Kalivoda . . . . .	221
<i>Toward Dialogue Modeling: A Semantic Annotation Scheme for Questions and Answers</i>	
María Andrea Cruz Blandón, Gosse Minnema, Aria Nourbakhsh, Maria Boritchev and Maxime Amblard . . . . .	231
<i>Towards a General Abstract Meaning Representation Corpus for Brazilian Portuguese</i>	
Marco Antonio Sobrevilla Cabezuado and Thiago Pardo . . . . .	237



# Conference Program

## Thursday 1 August 2019

*Breakfast is served from 7.30am to 9.00am.*

### **08:30–10:00 Session 1**

08:30–08:45 *Introduction by workshop chairs*

08:45–09:45 *Invited talk: Rebecca Passonneau*

09:45–10:00 *Crowdsourced Hedge Term Disambiguation*  
Morgan Ulinski and Julia Hirschberg

### **10:00–11:00 Poster session 1**

*Coffee is served from 10.30-11.00*

### **11:00–12:15 Session 2**

11:00–11:20 *WiRe57 : A Fine-Grained Benchmark for Open Information Extraction*  
William Lechelle, Fabrizio Gotti and Phillippe Langlais

11:20–11:40 *Crowdsourcing Discourse Relation Annotations by a Two-Step Connective Insertion Task*  
Frances Yung, Vera Demberg and Merel Scholman

11:40–12:00 *Annotating and Analyzing the Interactions between Meaning Relations*  
Darina Gold, Venelin Kovatchev and Torsten Zesch

12:00–12:15 *CCGweb: a New Annotation Tool and a First Quadrilingual CCG Treebank*  
Kilian Evang, Lasha Abzianidze and Johan Bos

12:15–14:00 *Lunch break*

**Thursday 1 August 2019 (continued)**

**14:00–15:00 Session 3**

14:00–14:20 *The Making of the Litkey Corpus, a Richly Annotated Longitudinal Corpus of German Texts Written by Primary School Children*  
Ronja Laarmann-Quante, Stefanie Dipper and Eva Belke

14:20–14:40 *The Materials Science Procedural Text Corpus: Annotating Materials Synthesis Procedures with Shallow Semantic Structures*  
Sheshera Mysore, Zachary Jensen, Edward Kim, Kevin Huang, Haw-Shiuan Chang, Emma Strubell, Jeffrey Flanigan, Andrew McCallum and Elsa Olivetti

14:40–14:55 *Tagging Modality in Oceanic Languages of Melanesia*  
Annika Tjuka, Lena Weißmann and Kilu von Prince

**15:00–16:00 Poster session 2**

*Coffee & snacks are served from 15.30-16.00*

**16:00–16:40 Session 4**

16:00–16:20 *Harmonizing Different Lemmatization Strategies for Building a Knowledge Base of Linguistic Resources for Latin*  
Francesco Mambrini and Marco Passarotti

16:20–16:40 *Assessing Back-Translation as a Corpus Generation Strategy for Non-English Tasks: A Study in Reading Comprehension and Word Sense Disambiguation*  
Fabricio Monsalve, Kervy Rivas Rojas, Marco Antonio Sobrevilla Cabezudo and Arturo Oncevay

**16:40–17:40 Discussion / Panel Session**

**Thursday 1 August 2019 (continued)**

**Poster session 1 (10.00-11.00)**

*A Framework for Annotating ‘Related Works’ to Support Feedback to Novice Writers*

Arlene Casey, Bonnie Webber and Dorota Glowacka

*An Online Annotation Assistant for Argument Schemes*

John Lawrence, Jacky Visser and Chris Reed

*Annotating Formulaic Sequences in Spoken Slovenian: Structure, Function and Relevance*

Kaja Dobrovoljc

*Annotating Information Structure in Italian: Characteristics and Cross-Linguistic Applicability of a QUD-Based Approach*

Kordula De Kuthy, Lisa Brunetti and Marta Berardi

*DEFT: A Corpus for Definition Extraction in Free- and Semi-Structured Text*

Sasha Spala, Nicholas A. Miller, Yiming Yang, Franck Dérnoncourt and Carl Dockhorn

*Explaining Simple Natural Language Inference*

Aikaterini-Lida Kalouli, Annebeth Buis, Livy Real, Martha Palmer and Valeria de-Paiva

*On the Role of Discourse Relations in Persuasive Texts*

Ines Rehbein

*One Format to Rule Them All – The Emtsv Pipeline for Hungarian*

Balázs Indig, Bálint Sass, Eszter Simon, Iván Mittelholcz, Noémi Vadász and Márton Makrai

*Turkish Treebanking: Unifying and Constructing Efforts*

Utku Türk, Furkan Atmaca, Şaziye Betül Özateş, Abdullatif Köksal, Balkiz Ozturk Basaran, Tunga Gungor and Arzucan Özgür

**Thursday 1 August 2019 (continued)**

**Poster session 2 (15.00-16.00)**

*A Dataset for Semantic Role Labelling of Hindi-English Code-Mixed Tweets*

Riya Pal and Dipti Sharma

*A Multi-Platform Annotation Ecosystem for Domain Adaptation*

Richard Eckart de Castilho, Nancy Ide, Jin-Dong Kim, Jan-Christoph Klie and Keith Suderman

*A New Annotation Scheme for the Sejong Part-of-speech Tagged Corpus*

Jungyeul Park and Francis Tyers

*A Turkish Dataset for Gender Identification of Twitter Users*

Erhan Sezerer, Ozan Polatbilek and Selma Tekir

*Comparative Judgments are More Consistent than Binary Classification for Labelling Word Complexity*

Sian Gooding, Ekaterina Kochmar, Advait Sarkar and Alan Blackwell

*Continuous Quality Control and Advanced Text Segment Annotation with WAT-SL 2.0*

Christina Lohr, Johannes Kiesel, Stephanie Luther, Johannes Hellrich, Tobias Kolditz, Benno Stein and Udo Hahn

*Creation of a Corpus with Semantic Role Labels for Hungarian*

Attila Novák, László Laki, Borbála Novák, Andrea Dömötör, Noémi Ligeti-Nagy and Ágnes Kalivoda

*Toward Dialogue Modeling: A Semantic Annotation Scheme for Questions and Answers*

María Andrea Cruz Blandón, Gosse Minnema, Aria Nourbakhsh, Maria Boritchev and Maxime Amblard

*Towards a General Abstract Meaning Representation Corpus for Brazilian Portuguese*

Marco Antonio Sobrevilla Cabezudo and Thiago Pardo