

ACL 2020

**The Fourth Widening Natural Language Processing
Workshop**

Program of the Workshop

July 5, 2020
Seattle, USA

Platinum Sponsors



Google Research

Gold Sponsors

amazon | science

Silver Sponsors



Bronze Sponsors

Bloomberg

Engineering

©2020 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 978-1-952148-06-4

Introduction

Welcome to 2020 Widening NLP Workshop!

In 2016, a select group of people met at ACL in Berlin to discuss the underrepresentation of women and other minorities in the Natural Language Processing community. The outcome of this was the first Workshop for Women and underrepresented minorities in NLP at ACL 2017: a workshop specifically dedicated to highlighting the work of groups whose work is often overlooked. We are proud to continue this work with the 2020 Widening NLP workshop.

Every year, we strive to make the workshop more inclusive. After the 2017 iteration of WiNLP, we set up two submission deadlines, one for early submissions to accommodate those who need more time for visa applications and a later deadline for those who do not have such concerns.

Following the success of the 2018 workshop, in 2019 we sought to also highlight diversity in scientific background, discipline, training, obtained degrees, and seniority. In addition, we introduced a workshopping feature, where authors themselves opted in to give feedback to their fellow authors before the papers were sent out for review.

The year 2020 brought new challenges with a global pandemic canceling all in-person conferences and workshops but has also given us the opportunity to widen our audience and authorship further. Our travel funds have been repurposed to assist with higher speed internet connections and registration fees – since those are lower cost items it means we can offer assistance to even more people to join our virtual workshop.

Despite the rapid changes in 2020, we received 66 submissions! Of these submissions 54% were by women as first authors, 73% were by people of color, 54% of all papers came from outside of North America and Europe. Turning our attention to the papers which were accepted, equally encouraging numbers come up: 57% of all accepted submissions had women as their first authors, 75% of all accepted papers were by people of color, and 52% of the accepted papers came from outside of Europe and North America.

We are proud to have great invited speakers - Drs. Helena Caseli, Jose Eduardo Ochoa Luna, Verena Riesner, Rachel Tatman, as well as four excellent panelists - Drs. Anima Anandkumar, Luciana Benotti, Xanda Schofield, and MSc Jade Abbott.

We thus proudly present the 2020 Widening NLP workshop, and hope that you will enjoy the wonderful work of the authors, speakers, and panelists of this edition of the workshop.

Rossana, Samira, Erika, Ryan, Alicia, Antonios, and Khyathi.

Organizers:

Rossana Cunha, Federal University of Minas Gerais
Samira Shaikh, University of North Carolina at Charlotte
Erika Varis, Walt Disney Studios
Ryan Georgi, University of Washington
Antonios Anastasopoulos, Carnegie Mellon University
Khyathi Raghavi Chandu, Carnegie Mellon University
Alicia Tsai, UC Berkeley

Advisory Board:

Isabelle Augenstein, University of Copenhagen
Amittai Axelrod, University of Maryland
Libby Barak, Rutgers University
Daniel Beck, University of Melbourne
Chloé Braud, CNRS – LORIA
Hal Daumé III, Associate Professor, University of Maryland
Mourad Gridach, Ibn Zohr University, Agadir
Valia Kordoni, Humboldt University Berlin
Cisca Odalipo, Kampala International University
Martha Palmer, University of Colorado
Natalie Schluter, University of Copenhagen
Tiago Torrent, Federal University of Juiz de Fora,
Aline Villavicencio, Federal University of Rio Grande do Sul, University of Essex
Hanna Wallach, Microsoft Research
Zeeraq Waseem, University of Sheffield

Program Committee:

Gavin Abercrombie, University of Manchester
Fernando Alva-Manchego, University of Sheffield
Antonios Anastasopoulos, Carnegie Mellon University
Gabor Angeli, Square Inc
Miguel Ballesteros, Amazon
Libby Barak, Rutgers University
Marco Baroni, ICREA
Maria Barrett, University of Copenhagen
Chloé Braud, IRIT - CNRS
Jose Camacho-Collados, Cardiff University
Khyathi Raghavi Chandu, Carnegie Mellon University
Colin Cherry, Google
George Chrysostomou, The University of Sheffield
Kenneth Church, Baidu, USA
Marta R. Costa-jussà, Universitat Politècnica de Catalunya
Rossana Cunha, Federal University of Minas Gerais
Walter Daelemans, University of Antwerp, CLiPS
Dipanjan Das, Google AI Language
Erika Doggett, Walt Disney Studios
Luis Espinosa Anke, Cardiff University
Jennifer Foster, Dublin City University
Ryan Georgi, University of Washington

Kevin Gimpel, Toyota Technological Institute at Chicago
Julia Hirschberg, Columbia University in the City of New York
Dirk Hovy, Bocconi University
Kristen Howell, University of Washington
Naoya Inoue, Tohoku University / RIKEN AIP
Cassandra L. Jacobs, University of Wisconsin, Madison
Huda Khayrallah, Johns Hopkins University
Kevin Knight, DiDi Labs
Rebecca Knowles, National Research Council Canada
Philipp Koehn, Johns Hopkins University
Rik Koncel-Kedziorski, University of Washington
Valia Kordoni, Humboldt-Universität zu Berlin
Ophélie Lacroix, Siteimprove
Els Lefever, LT3, Ghent University
Lucy H. Lin, University of Washington
Nitin Madnani, Educational Testing Service
Alex Marin, Microsoft Corporation
Héctor Martínez Alonso, Apple Inc
Arya D. McCarthy, Johns Hopkins University
Kourosh Meshgi, Riken AIP/ Kyoto University
Lisa Michaud, Interactions LLC
Sabrina J. Mielke, Johns Hopkins University
Aakanksha Naik, Carnegie Mellon University
Vincent Ng, University of Texas at Dallas
Debora Nozza, Bocconi University
Eda Okur, Intel Labs
Miles Osborne, Bloomberg
Gustavo Henrique Paetzold, Federal University of Technology - Paraná
Alexis Palmer, University of North Texas
Patrick Pantel, Facebook
Yannick Parmentier, University of Lorraine
Panupong Pasupat, Google
Roma Patel, Brown University
Viviana Patti, University of Turin, Dipartimento di Informatica
Amandalynne Paullada, University of Washington
Siyao Peng, Georgetown University
Hoifung Poon, Microsoft Research
Matthew Purver, Queen Mary University of London
Dragomir Radev, Yale University
Gabriela Ramirez-de-la-Rosa, Universidad Autónoma Metropolitana
Michael Roth, University of Stuttgart
Niloofer Safi Samghabadi, University of Houston
Sashank Santhanam, University of North Carolina at Charlotte
Naomi Saphra, University of Edinburgh
David Schlangen, University of Potsdam
Alexandra Schofield, Harvey Mudd College
Brandon Sepulvado, NORC at the University of Chicago
Samira Shaikh, UNC - Charlotte
Karin Sim Smith, University of Sheffield
Noah A. Smith, University of Washington
Marco Antonio Sobrevilla Cabezudo, University of São Paulo

Vivek Srikumar, University of Utah
Mark Steedman, University of Edinburgh
Madhumita Sushil, University of Antwerp
Jeniya Tabassum, The Ohio State University
Christoph Teichmann, Bloomberg LP
Elena Tutubalina, Kazan Federal University
Aline Villavicencio, University of Sheffield, UK
Bonnie Webber, University of Edinburgh
Wei Xu, The Ohio State University
Olga Zamaraeva, University of Washington

Invited Speakers:

Verena Rieser, Heriot-Watt University
Helena Caseli, Universidade Federal de São Carlos
José Eduardo Ochoa, Universidad Católica San Pablo
Rachael Tatman, Rasa

Panelists:

Anima Anandkumar, California Institute of Technology, NVIDIA
Jade Abbott, Retro Rabbit
Luciana Benotti, Universidad Nacional de Córdoba
Xanda Schofield, Harvey Mudd College

Workshop Program

Sunday, July 5, 2020

8:00–9:00 **Sponsored "breakfast"**

9:00–10:30 **Session 1: Opening, Keynote, and Posters**

9:00–9:10 *Opening Remarks*

9:10–9:50 *Keynote 1: Verena Rieser*

9:50–10:30 *Poster Session A*

Corpus based Amharic sentiment lexicon generation

Girma Neshir Alemneh, Andreas Rauber and Solomon Atnafu

Negation handling for Amharic sentiment classification

Girma Neshir Alemneh, Andreas Rauber and Solomon Atnafu

Embedding Oriented Adaptable Semantic Annotation Framework for Amharic Web Documents

Kidane Woldemariyam and Dr. Fekade Getahun

Similarity and Farness Based Bidirectional Neural Co-Attention for Amharic Natural Language Inference

Abebawu Eshetu, Getenesh Teshome and Ribka Alemayehu

Large Vocabulary Read Speech Corpora for Four Ethiopian Languages: Amharic, Tigrigna, Oromo, and Wolaytta

Solomon Teferra Abate, Martha Yifiru Tachbelie, Michael Melese, Hafte Abera, Tewodros Gebreselassie, Wondwossen Mulugeta, Yaregal Assabie, Million Meshe-sha Beyene, Solomon Atnafu and Binyam Ephrem Seyoum

SIMPLEX-PB 2.0: A Reliable Dataset for Lexical Simplification in Brazilian Portuguese

Nathan Hartmann, Gustavo Henrique Paetzold and Sandra Aluísio

Token Level Identification of Multiword Expressions Using Contextual Information

REYHANEH HASHEMPOUR and Aline Villavicencio

Sunday, July 5, 2020 (continued)

Bi-directional Answer-to-Answer Co-attention for Short Answer Grading using Deep Learning

Abebawu Eshetu, Getenesh Teshome and Ribka Alemahu

Effective questions in referential visual dialogue

Mauricio Mazuecos, Alberto Testoni, Raffaella Bernardi and Luciana Benotti

A Translation-Based Approach to Morphology Learning for Low Resource Languages

Tewodros Gebreselassie, Amanuel Mersha and Michael Gasser

Monolingual corpus creation and evaluation of truly low-resource languages from Peru

Gina Bustamante and Arturo Oncevay

Tigrinya Automatic Speech recognition with Morpheme based recognition units

Hafta Abera and sebsibe hailemariam

Variants of Vector Space Reductions for Predicting the Compositionality of English Noun Compounds

Pegah Alipoormolabashi and Sabine Schulte im Walde

SentiTel: TABSA for Twitter reviews on Uganda Telecoms

David Kabiito and Joyce Nakatumba Nabende

An Assessment of Language Identification Methods on Tweets and Wikipedia Articles

Pedro Verneti and Larissa Freitas

A Comparison of Identification Methods of Brazilian Music Styles by Lyrics

Patrick Guimarães, Jader Froes, Douglas Costa and Larissa Freitas

Enabling fast and correct typing in 'Leichte Sprache' (Easy Language)

Ina Steinmetz and Karin Harbusch

AI4D - African Language Dataset Challenge

Kathleen Siminyu and Sackey Freshia

Can Wikipedia Categories Improve Masked Language Model Pretraining?

Diksha Meghwal, Katharina Kann, Iacer Calixto and Stanislaw Jastrzebski

Sunday, July 5, 2020 (continued)

Adversarial Evaluation of BERT for Biomedical Named Entity Recognition

Vladimir Araujo, Andrés Carvallo and Denis Parra

FFR v1.1: Fon-French Neural Machine Translation

Chris Chinenye Emezue and Femi Pancrace Bonaventure Dossou

Classification and Analysis of Neologisms Produced by Learners of Spanish: Effects of Proficiency and Task

Shira Wein

10:30–10:50 *Break (sponsored)*

10:50–12:30 **Session 2: Keynote and Mentoring**

10:50–11:30 *Keynote 2: Helena Caseli*

11:30–12:30 *Mentoring and Sponsor Roundtables*

12:30–13:10 **Lunch**

13:10–15:10 **Session 3: Panel, Keynote, and Posters**

13:10–13:50 *Panel Discussion: Anima Anandkumar, Jade Abbott, Luciana Benotti, and Xanda Schofield*

13:50–14:30 *Keynote 3: José Eduardo Ochoa*

14:30–15:10 *Poster Session B*

Developing a Monolingual Sentence Simplification Corpus for Urdu

Yusra Anees, Sadaf Abdul Rauf, Nauman Iqbal and Abdul Basit Siddiqi

Sunday, July 5, 2020 (continued)

Translating Natural Language Instructions for Behavioral Robot Navigation with a Multi-Head Attention Mechanism

Patricio Cerda-Mardini, Vladimir Araujo and Álvaro Soto

Towards Mitigating Gender Bias in a decoder-based Neural Machine Translation model by Adding Contextual Information

Christine Basta, Marta R. Costa-jussà and José A. R. Fonollosa

Predicting and Analyzing Law-Making in Kenya

Oyinlola Babafemi and Adewale Akinfaderin

Defining and Evaluating Fair Natural Language Generation

Catherine Yeo and Alyssa Chen

Political Advertising Dataset: the use case of the Polish 2020 Presidential Elections

Lukasz Augustyniak, Krzysztof Rajda, Tomasz Kajdanowicz and Michał Bernaczyk

The human unlikeness of neural language models in next-word prediction

Cassandra L. Jacobs and Arya D. McCarthy

A Study on the Influence of Architecture Complexity of RNNs for Intent Classification in E-Commerce Chats in Bahasa Indonesia

Renny Pradina Kusumawardani and Muhammad Azzam

Long-Tail Predictions with Continuous-Output Language Models

Shiran Dudy and Steven Bedrick

Analyzing the Framing of 2020 Presidential Candidates in the News

Audrey Acken and Dorottya Demszky

Understanding the Impact of Experiment Design for Evaluating Dialogue System Output

Sashank Santhanam and Samira Shaikh

Studying The Effect of Emotional and Moral Language on Information Contagion during the Charlottesville Event

Khyati Mahajan and Samira Shaikh

Mapping of Narrative Text Fields To ICD-10 Codes Using Natural Language Processing and Machine Learning

Risuna Nkolele

Sunday, July 5, 2020 (continued)

Multitask Models for Controlling the Complexity of Neural Machine Translation

Sweta Agrawal and Marine Carpuat

Using Social Media For Bitcoin Day Trading Behavior Prediction

Anna Paula Pawlicka Maule and Kristen Johnson

HausaMT v1.0: Towards English–Hausa Neural Machine Translation

Adewale Akinfaderin

Outcomes of coming out: Analyzing stories of LGBTQ+

Krithika Ramesh and Tanvi Anand

An Evaluation of Subword Segmentation Strategies for Neural Machine Translation of Morphologically Rich Languages

Aquia Richburg, Ramy Eskander, Smaranda Muresan and Marine Carpuat

Enhanced Urdu Word Segmentation using Conditional Random Fields and Morphological Context Features

Aamir Farhan, Mashrukh Islam and Dipti Misra Sharma

Flexible Non-Autoregressive Neural Machine Translation via Repositioning Edit Operations

Weijia Xu and Marine Carpuat

15:10–15:25 Break (sponsored)

Sunday, July 5, 2020 (continued)

15:25–16:15 Session 4: Keynote and Closing

15:25–16:05 *Keynote 4: Rachael Tatman*

16:05–16:15 *Closing Remarks*