

WiNLP 2021

The 5th Widening NLP Workshop

Proceedings of the Workshop

November 11, 2021

©2021 WiNLP

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

Message from the Organizing Committee

Welcome to 2021 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 2021 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 workshoping feature, where authors themselves opted in to give feedback to their fellow authors before the papers were sent out for review.

The years 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. In 2021 we are taking advantage of the hybrid format of the workshop to widen our audience, while also supporting authors who intended to attend in-person. Our travel funds have been additionally 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 2021, we received 52 submissions, accepting 44 of them, most of them from authors from under-represented countries and women authors. In particular, we will present papers from 19 unique countries across 5 continents!

We are proud to have great invited speakers - Dr. Jasmijn Bastings, Prof. Manuel Montes, and Dra. Adriana Lorena Iñiguez Carrillo, as well as four excellent panelists - Drs. Anna Rogers, Cecilia Superchi, Bahar Mehmani, and Yang Feng.

We thus proudly present the 2021 Widening NLP workshop, and hope that you will enjoy the wonderful work of the authors, speakers, and panelists of this edition of the workshop.

Erika, Ryan, Alicia, Antonios, Khyathi, Tirthankar, Haley, Xanda, and Surangika.

Platinum Sponsors



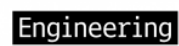
Gold Sponsors



Silver Sponsors



Bronze Sponsors



In-Kind Sponsors



Organizing Committee

Erika Varis, Walt Disney Studios.
Ryan Georgi, University of Washington.
Antonios Anastasopoulos, George Mason University.
Khyathi Raghavi Chandu, Carnegie Mellon University.
Alicia Tsai, UC Berkeley.
Tirthankar Ghosal, Institute of Formal and Applied Linguistics, Charles University, Czech Republic
Haley M. Lepp, Educational Testing Service.
Xanda Schofield, Harvey Mudd College.
Surangika Ranathunga, University of Moratuwa.

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
Rossana Cunha, Federal University of Minas Gerais.
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
Samira Shaikh, University of North Carolina at Charlotte.
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, Heriot Watt University, United Kingdom
Supunmali Ahangama, University of Moratuwa, Sri Lanka
Maria Barrett, IT University of Copenhagen, Denmark
Emily M. Bender, University of Washington, United States
Samuel R. Bowman, New York University, United States
Zhuo Cheng, University of North Carolina at Charlotte, United States
Colin Cherry, Google, Canada
Kenneth Church, Baidu, USA, United States
Hal Daumé III, UMD, United States
Piyumal Demotte, University of Moratuwa, Sri Lanka
Jacob Eisenstein, Google, United States
Jason Eisner, Johns Hopkins University, United States
George Foster, Google, Canada
Liye Fu, Cornell University, United States
Hardy Hardy, The University of Sheffield, United Kingdom

Julia Hockenmaier, University of Illinois Urbana-Champaign, United States
Kristen Howell, LivePerson Inc., United States
Naoya Inoue, Stony Brook University / RIKEN, United States
Hitoshi Isahara, Toyohashi University of Technology, Japan
Cassandra L. Jacobs, University at Buffalo, United States
Sneha Jha, Imperial College, United Kingdom
Rishemjit Kaur, CSIR-Central Scientific Instruments Organisation, India
Rebecca Knowles, National Research Council Canada, Canada
Lun-Wei Ku, Academia Sinica, Taiwan
Sadao Kurohashi, Kyoto University, Japan
Eric Laporte, Université Gustave Eiffel, France
Annie Lee, University of Toronto, Canada
Els Lefever, LT3, Ghent University, Belgium
Lucy Lin, University of Washington, United States
Christopher Manning, Stanford University, United States
Mitchell Marcus, University of Pennsylvania, United States
Aakanksha Naik, Carnegie Mellon University, United States
Graham Neubig, Carnegie Mellon University, United States
Bill Noble, University of Gothenburg, Sweden
Kemal Oflazer, Carnegie Mellon University in Qatar, Qatar
Miles Osborne, Bloomberg, United Kingdom
Sebastian Padó, Stuttgart University, Germany
Yannick Parmentier, University of Lorraine, France
Siyao Peng, Georgetown University, United States
Dragomir Radev, Yale University, United States
Gabriela Ramirez-de-la-Rosa, Universidad Autónoma Metropolitana, Mexico
Philip Resnik, University of Maryland, United States
Michael Riley, Google, United States
Ellen Riloff, University of Utah, United States
Brian Roark, Google Inc., United States
Horacio Saggion, Universitat Pompeu Fabra, Spain
Sashank Santhanam, University of North Carolina at Charlotte, United States
Ravi Shekhar, Queen Mary University of London, United Kingdom
Karin Sim Smith, Iconic, United Kingdom
Ry Smith, Metropolitan State University of Denver, United States
Noah A. Smith, University of Washington, United States
Richard Sproat, Google, Japan, Japan
Mihai Surdeanu, University of Arizona, United States
Zeerak Talat, University of Sheffield, United Kingdom
Joel Tetreault, Dataminr, United States
Mokanarangan Thayaparan, University of Manchester, United Kingdom
Christoph Tillmann, IBM Research, United States
Sara Tonelli, FBK, Italy
Bonnie Webber, University of Edinburgh, United Kingdom
Olga Zamaraeva, University of Washington, United States
Luke Zettlemoyer, University of Washington; Facebook, United States
Justine Zhang, Cornell University, United States

Conference Program

Thursday, November 11, 2021

8:00–9:00 **Poster Session A**

9:00–9:10 **Opening Remarks**

9:10–9:50 **Joint Keynote speaker with Queer In AI: Dr. Jasmijn Bastings (English with Spanish translation)**

9:50–10:30 **Keynote speaker 2: Prof. Manuel Montes (Spanish with English translation)**

10:30–11:00 **Coffee Break**

11:00–12:00 **Tutorial: How to write and respond to a review**

12:00–13:00 **Lunch**

13:00–14:20 **Panel: The Peer Review Process and Widening NLP**

14:20–14:45 **Coffee break**

Thursday, November 11, 2021 (continued)

14:45–15:25 Keynote speaker 3: Dra. Adriana Lorena Iiguez Carrillo

15:25–16:25 Poster Session B

OkwuGbé: End-to-End Speech Recognition for Fon and Igbo

Bonaventure F. P. Dossou and Chris Chinenye Emezue

TEET! Tunisian Dataset for Toxic Speech Detection

Slim Gharbi, Hatem Haddad, Mayssa Kchaou and Heger Arfaoui

Ara-Women-Hate: The first Arabic Hate Speech corpus regarding Women

Imane Guellil, Ahsan Adeel, Faical Azouaou, Mohamed Boubred, Yousra Houichi and Akram Abdelhaq Moumna

Multi-label Emotion Classification on Code-Mixed Text: Data and Methods

Iqra Ameer, Grigori Sidorov, Helena Gómez-Adorno and Rao Muhammad Adeel Nawab

Behavioral Testing of Knowledge Graph Embedding Models for Link Prediction

Wiem Ben Rim, Carolin Lawrence, Kiril Gashteovski, Mathias Niepert and Naoaki Okazaki

Developing Language Technology and NLP tools for endangered languages: Torwali

Naeem Uddin Hadi

How to Make Virtual Conferences Queer-Friendly: A Guide

Organizers of QueerInAI, A Pranav, MaryLena Bleile, Arjun Subramonian, Luca Soldaini, Danica Sutherland, Sabine Weber and Pan Xu

Neutralizing Gender Bias in Neural Machine Translation by Introducing Linguistic Knowledge

Ksenia Kharitonova, Marta R. Costa-jussà, Carlos Escolano, Christine Basta and Jordi Armengol-Estapé

Developing Keyboards for the Endangered Livonian Language

Mika Hämäläinen and Khalid Alnajjar

Nuanced Queerphobic Bias in Popular Sentiment Analysis Tools: A Data Set and Evaluation

Eddie Ungless, Björn Ross and Vaishak Belle

Thursday, November 11, 2021 (continued)

Coral: An Approach for Conversational Agents in Mental Health Applications
Harsh Sakhrani, Saloni Parekh and Shubham Mahajan

EM ALBERT: a step towards equipping Manipuri for NLP
Rudali Huidrom and Yves Lepage

Elementary-Level Math Word Problem Generation using Pre-Trained Transformers
Kashyapa Niyarepola, Savindu Ekanayake and Dineth Athapaththu

Towards the Early Detection of Child Predators in Chat Rooms: A BERT-based Approach
Sinchana Kumbale and Smriti Singh

One-Shot Lexicon Learning for Low-Resource Machine Translation
Anjali Kantharuban and Jacob Andreas

Sinhala-English Code-mixed and Code-switched Data Classification
Janani Sumanapala, Himashi Rathnayake and Raveesha Rukshani

"I don't know who she is": Discourse and Knowledge Driven Coreference Resolution

Angela Ramirez, Cecilia Li, Phillip Lee, Eduardo Zamora, Jeshwanth Bheemampally, Marilyn Walker and Adwait Ratnaparkhi

Idiom Extraction Method with Fine-tuning of Pre-trained Transformers Model for Named Entity Recognition
Nao Yamato

Occupational Gender stereotypes in Indic Languages
Neeraja Kirtane and Tanvi Anand

#WhyDidTheyStay: An NLP-driven approach to analyzing the factors that affect domestic violence victims
Marthala Kavya and Smriti Singh

Exploring Transfer Learning Pathways for Neural Machine Back Translation of Eskimo-Aleut, Chicham, and Classical Languages
Aaron Serianni and Daniel Whitenack

An interpretable representation that visually grounds dialog history
Mauricio Mazuecos, Franco M. Luque, Jorge Sánchez, Hernán Maina, Thomas Vadora and Luciana Benotti

Thursday, November 11, 2021 (continued)

Automated Template Paraphrasing for Conversational Assistants

Liane Vogel and Lucie Flek

Discovering changes in birthing narratives during COVID-19

Daphna Spira, Noreen Mayat, Caitlin Dreisbach and Adam Poliak

Explorations in Transfer Learning for OCR Post-Correction

Lindia Tjuatja, Shruti Rijhwani and Graham Neubig

A Prototype Free/Open-Source Morphological Analyser and Generator for Sakha

Sardana Ivanova, Francis Tyers and Jonathan Washington

Towards Text Simplification for Sinhala Language

Rumesh Sirithunga, Lahiru De Silva and Thamindu Aluthwala

Natural language processing as a tool to identify the Reddit particularities of cancer survivors around the time of diagnosis and remission: A pilot study

Ioana R. Podină, Ana-Maria Bucur, Diana Todea, Liviu Fodor, Andreea Luca, Liviu P. Dinu and Rareş Boian

Identifying Significant Citations via Mining Paper Full-Text

Muskaan Singh and Tirthankar Ghosal

SciBERT-based Multitasking Deep Neural Architecture to identify Contribution Statements from Scientific articles

Komal Gupta and Tirthankar Ghosal

Characterizing Test Anxiety on Social Media

Esha Julka, Olivia Kowalishin, Jalisha B. Jenifer and Adam Poliak

Sample Selection Guided by Domain and Task for Cross-Domain Targeted Sentiment Analysis

Kasturi Bhattacharjee, Rashmi Gangadharaiah and Smaranda Muresan

Bengali Parallel Universal Dependency Treebank

Pritha Majumdar

The Development of Pre-processing Tools and Pre-trained Embedding Models for Amharic

Tadesse Destaw, Abinew Ayele and Seid Muhie Yimam

Thursday, November 11, 2021 (continued)

Towards Syntax-Aware Dialogue Summarization using Multi-task Learning
Seolhwa Lee, Kisu Yang, Chanjun Park, João Sedoc and Heuseok Lim

Detoxifying Language Models with Proximal Policy Optimization
Taaha Kazi

Towards Personalized Descriptions of Scientific Concepts
Sonia Murthy, Daniel King, Tom Hope, Daniel Weld and Doug Downey

Building Prosody Labeled Corpus in Hindi
Esha Banerjee, Atul Kr. Ojha and Girish Jha

How Well Can an Agent Understand Different Accents?
Divya Tadimeti, Kallirroi Georgila and David Traum

Monolingual Pre-Trained Language Models for Tigrinya
Fitsum Gaim, Wonsuk Yang and Jong C. Park

ASQ: Automatically Generating Question-Answer Pairs using AMRs
Geetanjali Rakshit and Jeffrey Flanigan

Adverse Drug Reaction Classification of Tweets with Fusion of Text and Drug Representations
Andrey Sakhovskiy and Elena Tutubalina

Detecting Gender Bias using Explainability
Gauri Gupta, Supriti Vijay and Krithika Ramesh

Leveraging ultradense embeddings to analyze gender-oriented extremist recruitment targeting
Jatin Khilnani, Rasika Bhalerao, Tatenda Ndambakuwa and Maxwell Aliapoulios

Thursday, November 11, 2021 (continued)

16:25–16:35 Closing Remarks

16:35–17:30 Social

Coffee