

AAACL-IJCNLP 2020

**The 1st Conference of the Asia-Pacific Chapter of the  
Association for Computational Linguistics and the 10th  
International Joint Conference on Natural Language  
Processing**

**Proceedings of the Student Research Workshop**

December 4 - December 7, 2020  
Suzhou, China

©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](mailto:acl@aclweb.org)

ISBN 978-1-952148-93-4

## Introduction

Welcome to the ACL-IJCNLP 2020 Student Research Workshop (SRW)!

Held in conjunction with The 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (ACL) and the 10th International Joint Conference on Natural Language Processing conference (IJCNLP), the ACL-IJCNLP 2020 SRW is the first SRW offered in the Asia-Pacific area, and is joining its older siblings: the ACL SRW, the NAACL SRW and the EACL SRW.

SRW is a workshop for student researchers in computational linguistics and natural language processing. The SRW aims to provide student researchers the opportunity to present their work and receive constructive feedback and mentorship by experienced members of the ACL community.

Following the tradition established by previous SRWs, this year's submissions were organized into two tracks: research papers and thesis proposals.

- **Research papers:** Papers in this category can describe completed work, or work in progress with preliminary results. For these papers, the first author must be a current graduate or undergraduate student.
- **Thesis proposals:** This category is appropriate for advanced students who have decided on a thesis topic and wish to get feedback on their proposal and broader ideas for their continuing work.

We received a total of 47 submissions: 38 research papers and 9 thesis proposals. We accepted 22 research papers and two thesis proposals, resulting in an overall acceptance rate of 51%. We were delighted that submissions were diverse not only in topics but also in terms of the student demographics.

Following previous SRWs, we provided two mentoring programs for participants. The first program was the pre-submission mentoring program which offered students the opportunity to get feedback by a mentor prior to submitting their work for review. Eleven papers participated in the pre-submission mentoring program. In addition, we offered a post-accept mentoring program for accepted papers. In the post-accept mentoring program, a mentor was assigned to each accepted paper to help the student authors with the preparation of camera-ready submission and the presentation materials.

In addition to keeping up with SRW traditions, we also introduced new program features to the SRW: the **SRW Keynote** and an **SRW Best Paper Award**. We believe that these features add value to the workshop and we hope that they will be carried over to all future SRWs.

We would like to thank the 11 pre-submission mentors and the 24 post-acceptance mentors for dedicating their time and effort to help the student researchers with their paper in various stages. We would also like to thank the members of the program committee for their in-depth review and constructive feedback for each submitted paper.

Many thanks to our faculty advisors, Lun-Wei Ku and Vincent Ng, who provided tremendous help and guidance through the preparation of this workshop. Special thanks to Shruti Rijhwani for her support and great advice. We are also indebted to Emily M. Bender who accepted our invitation to give the first SRW keynote.

We would also like to thank the organizers of the ACL-IJCNLP conference for their support. Finally, we would like to thank all the student participants who have submitted their work to the workshop.

We hope you enjoy the ACL-IJCNLP 2020 SRW!



**Student Chairs:**

Boaz Shmueli, Academia Sinica (Taiwan)  
Yin Jou Huang, Kyoto University (Japan)

**Faculty Advisors:**

Lun-Wei Ku, Academia Sinica (Taiwan)  
Vincent Ng, University of Texas at Dallas (USA)

**Pre-submission Mentors:**

Valerio Basile, University of Turin  
Rishi Bommasani, Stanford University  
David Chiang, University of Notre Dame  
Parisa Kordjamshidi, Michigan State University  
Ellie Pavlick, Brown University  
Sai Krishna Rallabandi, Carnegie Mellon University  
Paul Rayson, Lancaster University  
Arijit Sehanobish, Yale University  
Anders Søgaard, University of Copenhagen  
Chen-Tse Tsai, Bloomberg  
Alakananda Vempala, Bloomberg  
Wei Yu, Carnegie Mellon University

**Post-accept Mentors:**

Reinald Kim Amplayo, University of Edinburgh  
Daniel Beck, University of Melbourne  
Eduardo Blanco, University of North Texas  
Ronald Cardenas, Charles University  
Lucia Donatelli, Saarland University  
Micha Elsner, The Ohio State University  
Kilian Evang, DFKI  
Ji He, Citadel LLC  
Junjie Hu, Carnegie Mellon University  
Divyansh Kaushik, Carnegie Mellon University  
Yang Liu, Tsinghua University  
Kenton Murray, Johns Hopkins University  
Denis Newman-Griffis, University of Pittsburgh  
Mah Parsa, University of Toronto  
Sai Krishna Rallabandi, Carnegie Mellon University  
Sonit Sehanobish, Yale  
Kurita Shuhei, RIKEN AIP  
Singh, Macquarie University  
Haiyue Song, Kyoto University  
Arseny Tolmachev, Fujitsu Laboratories Ltd.  
Bonnie Webber, University of Edinburgh

Pei Zhou, University of Southern California  
Hao Zhu, Carnegie Mellon University  
Arkaitz Zubiaga, Queen Mary University of London

**Program Committee:**

Piush Aggarwal, University of Duisburg-Essen, Language Technology Lab  
Afroz Ahamad, BITS Pilani Hyderabad Campus  
Miguel A. Alonso, Universidade da Coruña  
Reem Alqifari, King Saud University & University of York  
Rami Aly, University of Cambridge  
Evelin Amorim, UFMG  
Maria Antoniak, Cornell University  
Anusha Balakrishnan, Microsoft Semantic Machines  
Jorge Balazs, University of Tokyo  
Esma Balkir, University of Edinburgh  
Anjali Bhavan, Delhi Technological University  
Tatiana Bladier, Heinrich Heine University of Düsseldorf  
Arlene Casey, University of Edinburgh  
Lisa Andreevna Chalaguine, UCL  
Jonathan P. Chang, Cornell University  
Aditi Chaudhary, Carnegie Mellon University  
Jifan Chen, UT Austin  
Xinchi Chen, Amazon AWS  
Xiang Dai, University of Sydney  
Siddharth Dalmia, Carnegie Mellon University  
Samvit Dammalapati, Indian Institute of Technology Delhi  
Alok Debnath, International Institute of Information Technology  
Pieter Delobelle, KU Leuven  
Louise Deléger, INRAE - Université Paris-Saclay  
Chris Develder, Ghent University  
Anne Dirkson, Leiden University  
Zi-Yi Dou, Carnegie Mellon University  
Hicham El Boukkouri, LIMSI, CNRS, Université Paris-Saclay  
Carlos Escolano, Universitat Politècnica de Catalunya  
Tina Fang, University of Waterloo  
Murhaf Fares, University of Oslo  
Amir Feder, Technion - Israel Institute of Technology  
Dayne Freitag, SRI International  
Yoshinari Fujinuma, University of Colorado Boulder  
Diana Galvan-Sosa, Tohoku University  
Marcos Garcia, Universidade de Santiago de Compostela  
Tirthankar Ghosal, Indian Institute of Technology Patna  
Arijit Ghosh Chowdhury, Manipal Institute of Technology  
Abhinav Gupta, Mila  
Sarah Gupta, University of Washington  
Vivek Gupta, University of Utah  
Barry Haddow, University of Edinburgh  
Hardy Hardy, The University of Sheffield  
Barbora Hladka, Charles University  
Christopher Homan, Rochester Institute of Technology

Phu Mon Htut, New York University  
Jeff Jacobs, Columbia University  
Labiba Jahan, Florida International University  
Jyoti Jha, IIIT Hyderabad  
Vasu Jindal, University of Texas at Dallas  
Jad Kabbara, McGill University  
Tomoyuki Kajiwara, Osaka University  
Zara Kancheva, IICT-BAS  
Sudipta Kar, University of Houston  
Alina Karakanta, University of Trento  
Taiwo Kolajo, Federal University Lokoja  
Mamoru Komachi, Tokyo Metropolitan University  
Mandy Korpusik, Loyola Marymount University  
Kalpesh Krishna, University of Massachusetts Amherst  
Kemal Kurniawan, University of Melbourne  
Yash Kumar Lal, Stony Brook University  
Alexandra Lavrentovich, Amazon Alexa  
Bowen Li, University of Edinburgh  
Lei Li, Peking University  
Yiyuan Li, Carnegie Mellon University  
Jasy Suet Yan Liew, Syracuse University  
Fangyu Liu, University of Cambridge  
Di Lu, Dataminr  
Sean MacAvaney, Georgetown University  
Valentin Malykh, Huawei Noah's Ark Lab / Kazan Federal University  
Emma Manning, Georgetown University  
Zita Marinho, Priberam Labs / Institute of Systems and Robotics, IST  
Bruno Martins, IST and INESC-ID  
Pedro Henrique Martins, Instituto de Telecomunicações / Instituto Superior Técnico  
Kate McCurdy, University of Edinburgh  
Shikib Mehri, Carnegie Mellon University  
Rui Meng, University of Pittsburgh  
Antonio Valerio Miceli Barone, The University of Edinburgh  
Tsvetomila Mihaylova, Instituto de Telecomunicações  
Farjana Sultana Mim, Tohoku University  
Sewon Min, University of Washington  
Gosse Minnema, University of Groningen  
Rohan Mishra, Delhi Technological University  
Nora Muheim, University of Zürich  
Masaaki Nagata, NTT Corporation  
Nihal V. Nayak, Brown University  
Vincent Nguyen, Australian National University & CSIRO Data61  
Arturo Oncevay, The University of Edinburgh  
Yasumasa Onoe, The University of Texas at Austin  
Naoki Otani, Carnegie Mellon University  
Endang Wahyu Pamungkas, University of Turin  
Sheena Panthaplackel, The University of Texas at Austin  
Chan Young Park, Carnegie Mellon University  
Archita Pathak, University at Buffalo  
Siyao Peng, Georgetown University  
Pranav A, Dayta AI

Yusu Qian, New York University  
Ivaylo Radev, IICT-BAS  
Sree Harsha Ramesh, UMass Amherst  
Surangika Ranathunga, University of Moratuwa  
Vikas Raunak, CMU  
Abhilasha Ravichander, Carnegie Mellon University  
Steffen Remus, Hamburg University  
Shruti Rijhwani, Carnegie Mellon University  
Maria Ryskina, Carnegie Mellon University  
Sepideh Sadeghi, Uber Technologies Inc.  
Niloofer Safi Samghabadi, University of Houston  
Jin Sakuma, University of Tokyo  
Jainisha Sankhavara, DA-IICT  
Ramit Sawhney, Netaji Subhas Institute of Technology  
Sebastian Schuster, Stanford University  
Olga Seminck, CNRS  
Gautam Kishore Shahi, University of Duisburg-Essen  
A.B. Siddique, University of California, Riverside  
Pradyumna Sinha, Delhi Technological University  
Sunayana Sitaram, Microsoft Research India  
Kevin Small, Amazon  
Sergey Smetanin, National Research University Higher School of Economics  
Richard Sproat, Google, Japan  
Shabnam Tafreshi, The George Washington University  
Antonio Toral, University of Groningen  
Elena Tutubalina, Kazan Federal University  
Sowmya Vajjala, National Research Council  
Emiel Van Miltenburg, Tilburg University  
Adina Williams, Facebook, Inc.  
Yumo Xu, University of Edinburgh  
Rongtian Ye, Aalto University  
Da Yin, Peking University  
Michael Yoder, Carnegie Mellon University  
Omnia Zayed, National University of Ireland Galway  
Meishan Zhang, Tianjin University  
Ivan Šimko, University of Zurich

## **Keynote Talk**

### **Societal impacts of NLP: How and when to integrate them into your research (and how to make time for that)**

**Emily M. Bender**

Professor, University of Washington

#### **Abstract**

In March 2020, the ACL adopted the ACM's Code of Ethics and, beginning with EMNLP 2020, ACL conferences have included ethical considerations in the review process. In this talk, I will put these developments into historical perspective, talk about the kinds of risks that give rise to them, and explore positive steps that all NLP researchers can take to increase the chances that our time, effort and creativity are put towards work that is helpful for people, especially those who have been subject to marginalization.

#### **Bio**

Dr. Emily M. Bender is a professor of linguistics at the University of Washington where she is the faculty director of the professional MS program in computational linguistics. Her research interests include the interaction of linguistics and NLP and the societal impact of language technology and how transparent documentation can help mitigate the effects of bias and the potential for trained systems to perpetuate systems of oppression. She is also actively working on how to best incorporate training on ethics and societal impact into NLP curricula.



## Table of Contents

<i>Text Classification through Glyph-aware Disentangled Character Embedding and Semantic Sub-character Augmentation</i>	
Takumi Aoki, Shunsuke Kitada and Hitoshi Iyatomi . . . . .	1
<i>Two-Headed Monster and Crossed Co-Attention Networks</i>	
Yaoyiran Li and Jing Jiang . . . . .	8
<i>Towards a Task-Agnostic Model of Difficulty Estimation for Supervised Learning Tasks</i>	
Antonio Laverghetta Jr., Jamshidbek Mirzakhlov and John Licato . . . . .	16
<i>A Siamese CNN Architecture for Learning Chinese Sentence Similarity</i>	
Haoxiang Shi, Cen Wang and Tetsuya Sakai . . . . .	24
<i>Automatic Classification of Students on Twitter Using Simple Profile Information</i>	
Lili-Michal Wilson and Christopher Wun . . . . .	30
<i>Towards Code-switched Classification Exploiting Constituent Language Resources</i>	
Kartikey Pant and Tanvi Dadu . . . . .	37
<i>Hindi History Note Generation with Unsupervised Extractive Summarization</i>	
Aayush Shah, Dhineshkumar Ramasubbu, Dhruv Mathew and Meet Chetan Gadoya . . . . .	44
<i>Unbiasing Review Ratings with Tendency Based Collaborative Filtering</i>	
Pranshi Yadav, Priya Yadav, Pegah Nokhiz and Vivek Gupta . . . . .	50
<i>Building a Part-of-Speech Tagged Corpus for Drenjongke (Bhutia)</i>	
Mana Ashida, Seunghun Lee and Kunzang Namgyal . . . . .	57
<i>Towards a Standardized Dataset on Indonesian Named Entity Recognition</i>	
Siti Oryza Khairunnisa, Aizhan Imankulova and Mamoru Komachi . . . . .	64
<i>Formal Sanskrit Syntax: A Specification for Programming Language</i>	
K. Kabi Khangamba and Girish Jha . . . . .	72
<i>Resource Creation and Evaluation of Aspect Based Sentiment Analysis in Urdu</i>	
Sadaf Rani and Muhammad Waqas Anwar . . . . .	79
<i>Making a Point: Pointer-Generator Transformers for Disjoint Vocabularies</i>	
Nikhil Prabhu and Katharina Kann . . . . .	85
<i>Training with Adversaries to Improve Faithfulness of Attention in Neural Machine Translation</i>	
Pooya Moradi, Nishant Kambhatla and Anoop Sarkar . . . . .	93
<i>Document-Level Neural Machine Translation Using BERT as Context Encoder</i>	
Zhiyu Guo and Minh Le Nguyen . . . . .	101
<i>A Review of Cross-Domain Text-to-SQL Models</i>	
Yujian Gan, Matthew Purver and John R. Woodward . . . . .	108
<i>Multi-task Learning for Automated Essay Scoring with Sentiment Analysis</i>	
Panitan Muangkammuen and Fumiyo Fukumoto . . . . .	116

<i>Aspect Extraction Using Coreference Resolution and Unsupervised Filtering</i> Deon Mai and Wei Emma Zhang .....	124
<i>GRUBERT: A GRU-Based Method to Fuse BERT Hidden Layers for Twitter Sentiment Analysis</i> Leo Horne, Matthias Matti, Pouya Pourjafar and Zuowen Wang .....	130
<i>Exploring Statistical and Neural Models for Noun Ellipsis Detection and Resolution in English</i> Payal Khullar .....	139
<i>MRC Examples Answerable by BERT without a Question Are Less Effective in MRC Model Training</i> Hongyu Li, Tengyang Chen, Shuting Bai, Takehito Utsuro and Yasuhide Kawada .....	146
<i>Text Simplification with Reinforcement Learning Using Supervised Rewards on Grammaticality, Meaning Preservation, and Simplicity</i> Akifumi Nakamachi, Tomoyuki Kajiwara and Yuki Arase .....	153
<i>Label Representations in Modeling Classification as Text Generation</i> Xinyi Chen, Jingxian Xu and Alex Wang .....	160
<i>Generating Inflectional Errors for Grammatical Error Correction in Hindi</i> Ankur Sonawane, Sujeet Kumar Vishwakarma, Bhavana Srivastava and Anil Kumar Singh .....	165

# Conference Program

Dec 7, 2020 (GMT+8)

**08:45–09:00** *Welcome and Opening Remarks*

**09:00–10:00** **Keynote Address**

*Societal impacts of NLP: How and when to integrate them into your research (and how to make time for that)*

Emily M. Bender, University of Washington

**10:00–11:00** **Session SRW1: Advanced Models**

*Text Classification through Glyph-aware Disentangled Character Embedding and Semantic Sub-character Augmentation*

Takumi Aoki, Shunsuke Kitada and Hitoshi Iyatomi

*Two-Headed Monster and Crossed Co-Attention Networks*

Yaoyiran Li and Jing Jiang

*Towards a Task-Agnostic Model of Difficulty Estimation for Supervised Learning Tasks*

Antonio Laverghetta Jr., Jamshidbek Mirzakhlov and John Licato

*A Siamese CNN Architecture for Learning Chinese Sentence Similarity*

Haoxiang Shi, Cen Wang and Tetsuya Sakai

Dec 7, 2020 (GMT+8) (continued)

**11:00–12:00 Session SRW2: Social Media and Applications**

*Automatic Classification of Students on Twitter Using Simple Profile Information*

Lili-Michal Wilson and Christopher Wun

*Towards Code-switched Classification Exploiting Constituent Language Resources*

Kartikey Pant and Tanvi Dadu

*Hindi History Note Generation with Unsupervised Extractive Summarization*

Aayush Shah, Dhineshkumar Ramasubbu, Dhruv Mathew and Meet Chetan Gadoya

*Unbiasing Review Ratings with Tendency Based Collaborative Filtering*

Pranshi Yadav, Priya Yadav, Pegah Nokhiz and Vivek Gupta

**12:00–14:00 Break**

**14:00–15:00 Session SRW3: Low-Resource Languages**

*Building a Part-of-Speech Tagged Corpus for Drenjongke (Bhutia)*

Mana Ashida, Seunghun Lee and Kunzang Namgyal

*Towards a Standardized Dataset on Indonesian Named Entity Recognition*

Siti Oryza Khairunnisa, Aizhan Imankulova and Mamoru Komachi

*Formal Sanskrit Syntax: A Specification for Programming Language*

K. Kabi Khangamba and Girish Jha

*Resource Creation and Evaluation of Aspect Based Sentiment Analysis in Urdu*

Sadaf Rani and Muhammad Waqas Anwar

Dec 7, 2020 (GMT+8) (continued)

**15:00–16:00 Session SRW4: Translation and Transformation**

*Making a Point: Pointer-Generator Transformers for Disjoint Vocabularies*

Nikhil Prabhu and Katharina Kann

*Training with Adversaries to Improve Faithfulness of Attention in Neural Machine Translation*

Pooya Moradi, Nishant Kambhatla and Anoop Sarkar

*Document-Level Neural Machine Translation Using BERT as Context Encoder*

Zhiyu Guo and Minh Le Nguyen

*A Review of Cross-Domain Text-to-SQL Models*

Yujian Gan, Matthew Purver and John R. Woodward

**16:00–17:00 Session SRW5: Semantics and Pragmatics**

*Multi-task Learning for Automated Essay Scoring with Sentiment Analysis*

Panitan Muangkammuen and Fumiyo Fukumoto

*Aspect Extraction Using Coreference Resolution and Unsupervised Filtering*

Deon Mai and Wei Emma Zhang

*GRUBERT: A GRU-Based Method to Fuse BERT Hidden Layers for Twitter Sentiment Analysis*

Leo Horne, Matthias Matti, Pouya Pourjafar and Zuowen Wang

*Exploring Statistical and Neural Models for Noun Ellipsis Detection and Resolution in English*

Payal Khullar

Dec 7, 2020 (GMT+8) (continued)

**17:00–18:00 Session SRW6: Applications and Methods**

*MRC Examples Answerable by BERT without a Question Are Less Effective in MRC Model Training*

Hongyu Li, Tengyang Chen, Shuting Bai, Takehito Utsuro and Yasuhide Kawada

*Text Simplification with Reinforcement Learning Using Supervised Rewards on Grammaticality, Meaning Preservation, and Simplicity*

Akifumi Nakamachi, Tomoyuki Kajiwara and Yuki Arase

*Label Representations in Modeling Classification as Text Generation*

Xinyi Chen, Jingxian Xu and Alex Wang

*Generating Inflectional Errors for Grammatical Error Correction in Hindi*

Ankur Sonawane, Sujeet Kumar Vishwakarma, Bhavana Srivastava and Anil Kumar Singh

**18:00–18:30 Best Paper Award and Closing Remarks**