

NAACL HLT 2018

**The 2018 Conference of the  
North American Chapter of the  
Association for Computational Linguistics:  
Human Language Technologies**

**Proceedings of the Student Research Workshop**

June 2 - June 4, 2018  
New Orleans, Louisiana



This material is based upon work supported by the National Science Foundation under Grants No. 1803423 and 1714855. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

©2018 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 978-1-948087-26-1

## Introduction

Welcome to the NAACL-HLT 2018 Student Research Workshop!

This year's submissions were organized in two tracks: research papers and thesis proposals.

- Research papers may 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 are geared towards PhD students who have decided on a thesis topic and wish to get feedback on their proposal and broader ideas for their continuing work.

This year, we received a total of 51 submissions: 43 of these were research papers, and 8 were thesis proposals. We accepted 16 research papers and 4 thesis proposals, resulting in an overall acceptance rate of 39%. The main author of the accepted papers represent a variety of countries: Canada, Estonia, France, Germany, India, Japan (2 papers), Switzerland, UK, USA (11 papers).

Accepted research papers will be presented as posters within the NAACL main conference poster sessions. We will have oral presentations at our SRW oral session on June 2 for all 4 thesis papers to allow students to get feedback about their thesis work, along with 2 research papers selected as outstanding papers.

Following previous editions of the Student Research Workshop, we have offered students the opportunity to get mentoring feedback before submitting their work for review. Each student that requested pre-submission mentorship was assigned to an experienced researcher who read the paper and provided some comments on how to improve the quality of writing and presentation of the student's work. A total of 21 students participated in the mentorship program. During the workshop itself, we will also provide an in-site mentorship program. Each mentor will meet with their assigned students to provide feedback on their poster or oral presentation, and to discuss their research careers.

We would like to express our gratitude for the financial support from the National Science Foundation (NSF) and the Computing Research Association Computing Community Consortium (CRA-CCC). Thanks to their support, this year's SRW is able to assist students with their registration, travel, and lodging expenses.

We would like to thank the mentors for dedicating their time to help students improve their papers prior to submission, and we thank the members of the program committee for the constructive feedback they have provided for each submitted paper.

This workshop would not have been possible without the help from our faculty advisors, and we thank them for their guidance along this year of workshop preparation. We also thank the organizers of NAACL-HLT 2018 for their continuous support.

Finally, we would like to thank all students who have submitted their work to this edition of the Student Research Workshop. We hope our collective effort will be rewarded in the form of an excellent workshop!



**Organizers:**

Silvio Ricardo Cordeiro (Aix-Marseille Université)  
Shereen Oraby (University of California, Santa Cruz)  
Umashanthi Pavalanathan (Georgia Institute of Technology)  
Kyeongmin Rim (Brandeis University)

**Faculty Advisors:**

Swapna Somasundaran (ETS Princeton)  
Sam Bowman (New York University)

**Program Committee:**

Bharat Ram Ambati (Apple Inc)  
Rachel Bawden (Université Paris-Sud)  
Yonatan Bisk (University of Washington)  
Dallas Card (Carnegie Mellon University)  
Wanxiang Che (Harbin Institute of Technology)  
Monojit Choudhury (Microsoft Research)  
Mona Diab (George Washington University)  
Jesse Dodge (Carnegie Mellon University)  
Bonnie Dorr (Florida Institute for Human-Machine Cognition)  
Ahmed Elgohary (University of Maryland, College Park)  
Micha Elsner (The Ohio State University)  
Noura Farra (Columbia University)  
Thomas François (Université catholique de Louvain)  
Daniel Fried (University of California, Berkeley)  
Debanjan Ghosh (Rutgers University)  
Kevin Gimpel (Toyota Technological Institute at Chicago)  
Ralph Grishman (New York University)  
Alvin Grissom II (Ursinus College)  
Jiang Guo (Massachusetts Institute of Technology)  
Luheng He (University of Washington)  
Jack Hessel (Cornell University)  
Dirk Hovy (Bocconi University)  
Mohit Iyyer (Allen Institute for Artificial Intelligence)  
Yacine Jernite (New York University)  
Yangfeng Ji (University of Washington)  
Kristiina Jokinen (University of Helsinki)  
Daisuke Kawahara (Kyoto University)  
Chris Kedzie (Columbia University)  
Philipp Koehn (Johns Hopkins University)  
Jonathan K. Kummerfeld (University of Michigan)  
Maximilian Köper (University of Stuttgart)

Éric Laporte (Université Paris-Est Marne-la-Vallée)  
Sarah Ita Levitan (Columbia University)  
Jasy Suet Yan Liew (Universiti Sains Malaysia)  
Diane Litman (University of Pittsburgh)  
Stephanie M. Lukin (US Army Research Laboratory)  
Shervin Malmasi (Harvard Medical School)  
Diego Marcheggiani (University of Amsterdam)  
Daniel Marcu (University of Southern California)  
Mitchell Marcus (University of Pennsylvania)  
Prashant Mathur (eBay)  
Kathy McKeown (Columbia University)  
Julie Medero (Harvey Mudd College)  
Amita Misra (University of California, Santa Cruz)  
Saif M. Mohammad (National Research Council Canada)  
Taesun Moon (IBM Research)  
Smaranda Muresan (Columbia University)  
Karthik Narasimhan (Massachusetts Institute of Technology)  
Shashi Narayan (University of Edinburgh)  
Graham Neubig (Carnegie Mellon University)  
Avinesh P.V.S. (Technische Universität Darmstadt)  
Yannick Parmentier (University of Lorraine)  
Gerald Penn (University of Toronto)  
Massimo Poesio (Queen Mary University of London)  
Christopher Potts (Stanford University)  
Daniel Preotiuc-Pietro (University of Pennsylvania)  
Emily Prud'hommeaux (Boston College)  
Sudha Rao (University Of Maryland, College Park)  
Mohammad Sadegh Rasooli (Columbia University)  
Michael Roth (Saarland University)  
Markus Saers (Hong Kong University of Science and Technology)  
David Schlangen (Bielefeld University)  
Minjoon Seo (University of Washington)  
Kevin Small (Amazon)  
Sandeep Soni (Georgia Institute of Technology)  
Ian Stewart (Georgia Institute of Technology)  
Alane Suhr (Cornell University)  
Shyam Upadhyay (University of Pennsylvania)  
Sowmya Vajjala (Iowa State University)  
Bonnie Webber (University of Edinburgh)  
John Wieting (Carnegie Mellon University)  
Adina Williams (New York University)  
Diyi Yang (Carnegie Mellon University)  
Justine Zhang (Cornell University)  
Meishan Zhang (Heilongjiang University)  
Ran Zhao (Carnegie Mellon University)

## Table of Contents

<i>Alignment, Acceptance, and Rejection of Group Identities in Online Political Discourse</i> Hagyeong Shin and Gabriel Doyle .....	1
<i>Combining Abstractness and Language-specific Theoretical Indicators for Detecting Non-Literal Usage of Estonian Particle Verbs</i> Eleri Aedmaa, Maximilian Köper and Sabine Schulte im Walde .....	9
<i>Verb Alternations and Their Impact on Frame Induction</i> Esther Seyffarth .....	17
<i>A Generalized Knowledge Hunting Framework for the Winograd Schema Challenge</i> Ali Emami, Adam Trischler, Kaheer Suleman and Jackie Chi Kit Cheung .....	25
<i>Towards Qualitative Word Embeddings Evaluation: Measuring Neighbors Variation</i> Benedicte Pierrejean and Ludovic Tanguy .....	32
<i>A Deeper Look into Dependency-Based Word Embeddings</i> Sean MacAvaney and Amir Zeldes .....	40
<i>Learning Word Embeddings for Data Sparse and Sentiment Rich Data Sets</i> Prathusha Kameswara Sarma .....	46
<i>Igbo Diacritic Restoration using Embedding Models</i> Ignatius Ezeani, Mark Hepple, Ikechukwu Onyenwe and Enemouh Chioma .....	54
<i>Using Classifier Features to Determine Language Transfer on Morphemes</i> Alexandra Lavrentovich .....	61
<i>End-to-End Learning of Task-Oriented Dialogs</i> Bing Liu and Ian Lane .....	67
<i>Towards Generating Personalized Hospitalization Summaries</i> Sabita Acharya, Barbara Di Eugenio, Andrew Boyd, Richard Cameron, Karen Dunn Lopez, Pamela Martyn-Nemeth, Carolyn Dickens and Amer Ardati .....	74
<i>Read and Comprehend by Gated-Attention Reader with More Belief</i> Haohui Deng and Yik-Cheung Tam .....	83
<i>ListOps: A Diagnostic Dataset for Latent Tree Learning</i> Nikita Nangia and Samuel Bowman .....	92
<i>Japanese Predicate Conjugation for Neural Machine Translation</i> Michiki Kurosawa, Yukio Matsumura, Hayahide Yamagishi and Mamoru Komachi .....	100
<i>Metric for Automatic Machine Translation Evaluation based on Universal Sentence Representations</i> Hiroki Shimanaka, Tomoyuki Kajiwara and Mamoru Komachi .....	106
<i>Neural Machine Translation for Low Resource Languages using Bilingual Lexicon Induced from Comparable Corpora</i> Sree Harsha Ramesh and Krishna Prasad Sankaranarayanan .....	112
<i>Training a Ranking Function for Open-Domain Question Answering</i> Phu Mon Htut, Samuel Bowman and Kyunghyun Cho .....	120

<i>Corpus Creation and Emotion Prediction for Hindi-English Code-Mixed Social Media Text</i>	
Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar and Manish Shrivastava . .	128
<i>Sensing and Learning Human Annotators Engaged in Narrative Sensemaking</i>	
McKenna Tornblad, Luke Lapresi, Christopher Homan, Raymond Ptucha and Cecilia Ovesdotter Alm.....	136
<i>Generating Image Captions in Arabic using Root-Word Based Recurrent Neural Networks and Deep Neural Networks</i>	
Vasu Jindal .....	144



# Conference Program

**Saturday, June 2**

**10:30–12:00 Posters and Demos: Discourse and Pragmatics**

*Alignment, Acceptance, and Rejection of Group Identities in Online Political Discourse*

Hagyeong Shin and Gabriel Doyle

**15:30–17:00 Posters and Demos: Semantics**

*Combining Abstractness and Language-specific Theoretical Indicators for Detecting Non-Literal Usage of Estonian Particle Verbs*

Eleri Aedmaa, Maximilian Köper and Sabine Schulte im Walde

*Verb Alternations and Their Impact on Frame Induction*

Esther Seyffarth

*A Generalized Knowledge Hunting Framework for the Winograd Schema Challenge*

Ali Emami, Adam Trischler, Kaheer Suleman and Jackie Chi Kit Cheung

*Towards Qualitative Word Embeddings Evaluation: Measuring Neighbors Variation*

Benedicte Pierrejean and Ludovic Tanguy

*A Deeper Look into Dependency-Based Word Embeddings*

Sean MacAvaney and Amir Zeldes

**15:30–17:00 Posters and Demos: Sentiment Analysis**

*Learning Word Embeddings for Data Sparse and Sentiment Rich Data Sets*

Prathusha Kameswara Sarma

**Saturday, June 2 (continued)**

**17:00–18:30 Oral Presentations: Student Research Workshop Highlight Papers**

***17:00–17:14***

*Igbo Diacritic Restoration using Embedding Models*

Ignatius Ezeani, Mark Hepple, Ikechukwu Onyenwe and Enemouh Chioma

***17:15–17:29***

*Using Classifier Features to Determine Language Transfer on Morphemes*

Alexandra Lavrentovich

***17:30–17:44***

*End-to-End Learning of Task-Oriented Dialogs*

Bing Liu

***17:45–17:59***

*Towards Generating Personalized Hospitalization Summaries*

Sabita Acharya, Barbara Di Eugenio, Andrew Boyd, Richard Cameron, Karen Dunn Lopez, Pamela Martyn-Nemeth, Carolyn Dickens and Amer Ardati

***18:00–18:14***

*A Generalized Knowledge Hunting Framework for the Winograd Schema Challenge*

Ali Emami, Adam Trischler, Kaheer Suleman and Jackie Chi Kit Cheung

***18:15–18:30***

*Alignment, Acceptance, and Rejection of Group Identities in Online Political Discourse*

Hagyeong Shin and Gabriel Doyle

**Sunday, June 3**

**10:30–12:00 Posters and Demos: Information Extraction**

*Read and Comprehend by Gated-Attention Reader with More Belief*

Haohui Deng and Yik-Cheung Tam

**15:30–17:00 Posters and Demos: Machine Learning**

*ListOps: A Diagnostic Dataset for Latent Tree Learning*

Nikita Nangia and Samuel Bowman

**15:30–17:00 Posters and Demos: Machine Translation**

*Japanese Predicate Conjugation for Neural Machine Translation*

Michiki Kurosawa, Yukio Matsumura, Hayahide Yamagishi and Mamoru Komachi

*Metric for Automatic Machine Translation Evaluation based on Universal Sentence Representations*

Hiroki Shimanaka, Tomoyuki Kajiwara and Mamoru Komachi

*Neural Machine Translation for Low Resource Languages using Bilingual Lexicon Induced from Comparable Corpora*

Sree Harsha Ramesh and Krishna Prasad Sankaranarayanan

**Monday, June 4**

**10:30–12:00 Posters and Demos: Question Answering**

*Training a Ranking Function for Open-Domain Question Answering*

Phu Mon Htut, Samuel Bowman and Kyunghyun Cho

**10:30–12:00 Posters and Demos: Social Media and Computational Social Science**

*Corpus Creation and Emotion Prediction for Hindi-English Code-Mixed Social Media Text*

Deepanshu Vijay, Aditya Bohra, Vinay Singh, Syed Sarfaraz Akhtar and Manish Shrivastava

**14:00–15:30 Posters and Demos: Cognitive Modeling and Psycholinguistics**

*Sensing and Learning Human Annotators Engaged in Narrative Sensemaking*

McKenna Tornblad, Luke Lapresi, Christopher Homan, Raymond Ptucha and Cecilia Ovesdotter Alm

**14:00–15:30 Posters and Demos: Vision, Robotics and Other Grounding**

*Generating Image Captions in Arabic using Root-Word Based Recurrent Neural Networks and Deep Neural Networks*

Vasu Jindal