The NAACL organizers gratefully acknowledge the support from the following sponsors.

**Diamond**

![Amazon Science](image1)

Bloomberg

Google Research

LivePerson

**Platinum**

ByteDance

Kenshó

Grammarly

Megagon Labs

Microsoft

Reveal Brainspace

**Gold**

Co:here

Gresearch.co.uk

Relativity

Servicenow
Logotype: Modernizing a classic
Primary logo
Our logo is made up of two elements: the window shape and the type. These should not be altered or recreated. Send requests for the .eps artwork to email@tiaa.org.

Silver

Bronze

D&I Champions

D&I Contributors

Magic Data provides high quality training datasets for ML and customized AI training data labeling services to enterprises and academic institutions engaged in artificial intelligence R&D and application research to natural language processing (NLP), voice recognition (ASR), speech synthesis (TTS), and computer vision (CV). We provide data total solutions which cover automobile, finance, social networks, smart home automation, and end-user device, involving smart customer service, virtual assistant, machine translation, and many other AI scenarios.

business@magicdatatech.com
www.magicdatatech.com
Contact Us

G-Research is Europe's leading quantitative finance research firm. We hire the brightest minds in the world to tackle Solve the world's toughest financial problems. Research vacancies today.
CREATE TODAY.
PREDICT TOMORROW.
IMAGINE DISCOVERING SOMETHING UNDISCOVERED.
gresearch.co.uk
Message from the General Chair

Welcome to the 2022 meeting of the North American Association for Computational Linguistics! Due to the COVID-19 pandemic, NAACL-2021 was held virtually, and NAACL-2022 is the first major NLP conference that is run as an hybrid conference in North America. It is my pleasure to welcome many of you who are joining us in Seattle, as well as those who chose to participate in the conference virtually.

COVID safety is important to us and we will do whatever we can to help you enjoy the in-person conference despite the difficulties we all experience coming back to normality. At the same time, thanks to the virtual conference platform put together by Underline, we hope that our virtual attendees will experience the conference almost as if they are in Seattle and enjoy the conference.

NAACL-2022 decided, along with ACL-2022, to experiment with a new reviewing process, based on “rolling review” (ARR). While we believe that, eventually, this process will converge to an efficient review process that would benefit our community, pioneering such a process is not without difficulties. This would not have been possible without the incredible effort, devotion, thoughtfulness, patience, and many work hours put by our program chairs, Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, and the help from the ACL-2022 program chairs, Smaranda Muresan, Preslav Nakov, Aline Villavicencio. This process necessitated developing a new software package to support our publication, an effort that was done in collaboration with ACL-2022, and I am thankful to Ryan Cotterell who led this effort.

Among other innovations we installed in NAACL-2022 is a reproducibility track, where we attempted to incentivize authors to release models, code, and other information necessary to reproduce the main results and findings of their papers. We hope that this effort, led by Niranjan Balasubramanian, Jesse Dodge, Annie Louis, Daniel Deutsch and Yash Kumar Lal, will be followed in future conferences. Other initiatives include incorporating a “Responsible NLP Research” checklist into the submission process, a new special theme on “Human-Centered Natural Language Processing”, and many innovative activities led by our very active and thoughtful Diversity and Inclusion Committee, led by Diana Galván, Snigdha Chaturvedi and Yonatan Bisk, with Pranav A and Luciana Benotti as advisors.

Organizing a conference as large as NAACL, especially under the constraints of the times we live in, requires the support of a large number of volunteers who care deeply about our community and are willing to spend a lot of time and effort in this long process. It is an honor to coordinate such a team. I would like to thank the members of the organizing committee for their dedication, creativity, and hard work.

First, it is hard to imagine the amount of thought, care, and time, our program chairs Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz put into all aspects of organizing this conference – resulting in an exciting and high quality scientific program.

Many other volunteers have worked hard to make this conference a success and it would not be possible to name all of them here. I will only list the chairs of the main committees whose dedication, creativity, hard work and lively communication contributed to making NAACL-2022 a successful event:

- The diversity & inclusion committee chaired by Diana Galván, Snigdha Chaturvedi and Yonatan Bisk, with advisors Pranav A and Luciana Benotti.
- The industry track chairs, Rashmi Gangadharaiah, Anastassia Loukina and Bonan Min, and advisors Owen Rambow and Yunyao Li.
- The tutorial chairs, Cecilia Alm, Miguel Ballesteros and Yulia Tsvetkov.
- The demonstration chairs, Hannaneh Hajishirzi, Qiang Ning and Avi Sil.
• The workshop chairs, Dan Goldwasser, Yunyao Li and Ashish Sabharwal.

• The student research workshop chairs, Daphne Ippolito, Liunian Harold Li, Maria Leonor Pacheco and advisors, Danqi Chen and Nianwen Xue.

• The publication chairs, Ryan Cotterell, Danilo Croce and Jordan Zhang.

• The reproducibility chairs, Niranjan Balasubramanian, Jesse Dodge, Annie Louis, Daniel Deutsch and Yash Kumar Lal.

• The sponsorship chair, Byron Wallace.

• The volunteer chair, Daniel Khashabi.

• The publicity chairs, Nanyun (Violet) Peng, Emily Sheng, Sameer Singh.

• The virtual infrastructure chairs, Deepak Ramachandran, Martín Villalba, and Rishita Anubhai.

• The website chairs, Ice Pasupat and Vered Shwartz.

Many thanks to Chris Callison-Burch, the ACL Sponsorship Director, for helping the NAACL-2022 Sponsorship chair, Byron Wallace, managing the relations between the sponsors and NAACL-2022.

I am also very grateful to the chairs of previous years’ conferences, who were always ready to help and share their experience, and to the members of the ACL and NAACL Executive Committees for their support, feedback and advice.

As usual, special thanks go to Priscilla Rasmussen and to Jennifer Rachford who has stepped into the role of the ACL business manager just in time to help us with NAACL-22. They have been our local organizers and have dealt with all aspects of organizing and managing the conference, from room assignment, to food, to COVID tests.

Finally, I would like to thank all authors, invited speakers and panelists, area chairs and reviewers, the volunteers organizing and chairing sessions, and all attendees, in-person and virtual, for making this a scientifically exciting and socially engaging conference.

Welcome and hope you all enjoy the conference!

Dan Roth
University of Pennsylvania and AWS AI Labs
NAACL-2022 General Chair
June 2022
Welcome to the 2022 Annual Conference of the North American Association for Computational Linguistics! For the first time, NAACL-HLT 2022 is a hybrid conference. After two years of exclusively virtual conferences due to the COVID-19 pandemic, we are pleased that attendees and presenters will be able to join us in person in Seattle and from all over the world online for this year’s edition.

**Review Process** NAACL 2022 invited the submission of long and short papers featuring substantial, original, and unpublished research in all aspects of Computational Linguistics and Natural Language Processing (NLP). Our paper review process was organized in a hierarchical structure similar to recent years. We recruited 62 senior area chairs (SACs) for 26 areas, following the areas defined for NAACL 2022. There were two paths for submitting papers: special theme papers were directly submitted to the NAACL OpenReview site, and other main conference papers were reviewed through a new ACL-wide centralized reviewing process. In coordination with the ACL 2022 organizers, we experimented with the ACL Rolling Review (ARR) introduced as part of an initiative to improve efficiency and turnaround of reviewing for ACL conferences. Within this system, reviewing and acceptance of papers to publication venues was done in a two-step process: (1) centralized rolling review via ARR, where submissions receive reviews and meta-reviews from ARR reviewers and action editors; (2) commitment to a publication venue (e.g., NAACL 2022), so that Senior Area Chairs and Program Chairs make acceptance decisions for a submission using the ARR reviews and meta-reviews. During the first phase of the review process, we served as guest Editors in Chief for the ACL Rolling Review and worked to ensure that all papers submitted received at least three review and one meta-review, while balancing the reviewing load for reviewers and action editors. NAACL SACs acted as guest senior area chairs in the ARR system, by helping monitor review progress and supporting the 408 action editors and 3379 reviewers in their work. While the new reviewing mechanism was not as smooth as one could have hoped for, all papers submitted to ARR received at least three reviews and a meta-review, so that authors could decide to commit it to NAACL 2022 if they wanted to. The ACL Executive Committee, based on feedback from the community, will decide whether the advantages of a centralized rolling review system outweigh the disadvantages, taking into account the fast growth of our research field. Once papers were committed to the NAACL OpenReview site, SACs were in charge of making acceptance recommendation per area, taking into account the submission itself, (meta-)reviews, as well as comments to SACs provided by the authors and ethics reviews when applicable.

In coordination with Jesse Dodge, Anna Rogers, Margot Mieskes, Amanda Stent, and the ACL Ethics Committee, we incorporated a “Responsible NLP Research” checklist into the submission process, designed to encourage best research practices in our field, from an ethics and reproducibility perspective. The ARR Responsible NLP Research checklist is largely based on the NeurIPS 2021 paper checklist, the reproducible data checklist from Rogers, Baldwin, Leins’s paper “Just What do You Think You’re Doing, Dave? A Checklist for Responsible Data Use in NLP”, and the NLP Reproducibility checklist introduced by Dodge, Gururangan, Card, Schwartz and Smith in “Show Your Work: Improved Reporting of Experimental Results”. Authors were asked to follow the ACL code of ethics and to fill the checklist to ensure that best practices are put in place. Reviewers were asked to consult the checklist when deciding whether the paper requires ethics review. Based on input from reviewers and action editors, SACs flagged papers that required an in-depth ethics review, which was handled by a committee of 11 ethics reviewers. The ethics chairs provided guidance and office hours to help SACs decide when ethics review was required. The ethics reviews were integrated in the final acceptance recommendation by SACs and decisions by PCs.

**Special Theme** We highlighted “Human-Centered Natural Language Processing” as the special theme for the conference. As NLP applications increasingly mediate people’s lives, it is crucial to understand how the design decisions made throughout the NLP research and development lifecycle impact people,
whether there are users, developers, data providers or other stakeholders. For NAACL 2022, we invited submissions that address research questions that meaningfully incorporate stakeholders in the design, development, and evaluation of NLP resources, models and systems. We particularly encouraged submissions that bring together perspectives and methods from NLP and Human-Computer Interaction. Given their interdisciplinary nature, theme papers were reviewed through a dedicated process by reviewers with expertise in NLP and in Human-Computer Interaction. We received 52 submissions to the special theme, of which 14 have been accepted to appear at the conference.

Submission Statistics  The ACL Rolling Review received 196 submissions in December and 1897 in January, which were the two submission deadlines between the ACL and NAACL commitment deadlines. Of these 2103 submissions, 56% (1073) were committed to NAACL 2022 for the senior program committee to make an acceptance decision. We accepted a total of 442 papers (358 long papers and 84 short papers), representing 21.96% of papers submitted to ARR in December and January and to the NAACL special theme, and 41.19% of papers committed to NAACL (including the special theme papers). As a reference point, NAACL-HLT 2021 received 1797 submissions and accepted 477 papers, including 350 long and 127 short, for an overall acceptance rate of 26%. From the accepted papers, and based on the nominations from SACs, the best paper committee selected best papers, as well as a small number of outstanding papers with the goal of recognizing diverse types of contributions (including contributions to the special theme on human-centered NLP; innovation in model design, training or evaluation; resource or dataset contribution).

Additionally, 209 submissions (183 long and 26 short) were accepted for publications in the “Findings of ACL: NAACL 2022” (or Findings for short), an online companion publication for papers that are not accepted for publication in the main conference, but nonetheless have been assessed by the program committee as solid work with sufficient substance. A total of 5 accepted Findings papers were withdrawn. Findings paper were given the option to be presented as posters during the main conference: 183 took this opportunity and will be presented either in person or virtually.

NAACL 2022 will also feature 15 papers that were published at Transactions of the Association for Computational Linguistics (TACL) and 3 papers from the journal of Computational Linguistics (CL).

Program Format  The conference program was designed to allow for presentation and attendance in person in Seattle and virtually from all over the world. Oral sessions will consist of presentations done either in person or virtually. The Q&A session for each paper will alternate between in-person and online questions, with a volunteer helping monitor the online questions. All oral sessions will be live-streamed and recorded. All main conference posters will be presented with a 5-minute video pitch available online and with a virtual Q&A session, where papers will be grouped by topic to foster discussion. In addition, authors who attend the conference in Seattle will present their poster in person during traditional poster sessions. Finally, asynchronous interaction between authors and attendees will be made possible before, during and after the conference on the Underline platform. We also chose to start the conference early in the morning to overlap with normal waking hours in distant time zones.

The program includes several plenary sessions, which we hope will provide thought-provoking perspectives that will enrich discussions during the conference and beyond. In addition to a session for best paper awards, we are delighted to have keynote talks by Batya Friedman (University of Washington) and Manuel Montes-y-Gómez (National Institute of Astrophysics, Optics and Electronics of Mexico). Dan Roth (University of Pennsylvania and Amazon) will moderate a discussion on the role of linguistics and symbolic representations in NLP, with panelists Chitta Baral (Arizona State University), Emily Bender (University of Washington), Dilek Hakkani-tur (Amazon), and Christopher D. Manning (Stanford University). The industry track, demonstrations track and the student research workshop will have dedicated sessions during the main conference to round up the program, including a plenary panel on careers in NLP organized by the industry track chairs.
Gratitude  NAACL would not have been possible without the hard work of many volunteers. We are very grateful to all who contributed to make the conference possible, especially given the ongoing challenges raised by the COVID-19 pandemic.

We would like to start by thanking all the authors who submitted their work to the ACL Rolling Review and NAACL 2022. We could only accept a small fraction of submissions but hope that most papers received valuable feedback.

Next, we would like to thank all members of the Program Committee – who are too numerous to be all named here, but are listed elsewhere in the proceedings:

- The senior area chairs, who were incredibly responsive throughout the reviewing process and patiently helped improve the new reviewing infrastructure.
- ARR action editors and reviewers. Special thanks to those who stepped in at the last minute to serve as emergency reviewers. This was tremendously appreciated!
- The special theme area chair, Jeff Bigham, and all reviewers, with a special note of appreciation for those who contributed their time and expertise even though they do not usually publish in NLP conferences.
- The ethics Chairs, Kai-Wei Chang, Dirk Hovy and Diyi Yang, for designing a process to encourage consistent evaluation of ethical considerations during the review process, and their timely input to ensure the integration of ethics review in acceptance recommendations and decisions.
- The ethics reviewers: Yonatan Bisk, Kevin Bretonnel Cohen, Francien Dechesne, Jack Hessel, Jin-Dong Kim, Anne Lauscher, Dave Lewis, Margot Mieskes, Xanda Schofield, Lyle Ungar, Jingbo Xia.
- The outstanding reviewers and action editors who were nominated by the senior area chairs for writing reviews that were particularly helpful in the decision making process. They are recognized by name later in this volume.

Experimenting with a new reviewing system on the large scale required by NAACL would not have been possible without the following people:

- Amanda Stent and Goran Glavaš, as ARR Editors-in-Chiefs, for their tireless work in support of the ARR December and January cycles.
- Graham Neubig, Dhruv Naik and Nils Dycke, as ARR Tech Team for these two cycles.
- Celeste Martinez Gomez, Melisa Bok, and Nadia L’Bahy, as OpenReview Tech Team.
- Elijah Rippeth for his help coordinating the special theme submissions.

The following committees helped shape the conference in countless ways:

- The best paper committee: Thamar Solorio (Chair), Isabelle Augenstein, Gemma Bel Enguix, Alona Fyshe, Shafiq Joty and Emily Prud’hommeaux who enthusiastically read and recommended papers for awards in a short time frame.
- The diversity & inclusion committee chaired by Diana Galván, Snigdha Chaturvedi and Yonatan Bisk, with advisors Pranav A and Luciana Benotti.
- The industry track chairs, Rashmi Gangadharaiyah, Anastassia Loukina and Bonan Min, and advisors Owen Rambow and Yunyao Li.
- The tutorial chairs, Cecilia Alm, Miguel Ballesteros and Yulia Tsvetkov.
- The demonstration chairs, Hannaneh Hajishirzi, Qiang Ning and Avi Sil.
• The workshop chairs, Dan Goldwasser, Yunyao Li and Ashish Sabharwal.

• The student research workshop chairs, Daphne Ippolito, Liunian Harold Li, Maria Leonor Pacheco and advisors, Danqi Chen and Nianwen Xue.

• The publication chairs, Ryan Cotterell, Danilo Croce and Jordan Zhang.

• The reproducibility chairs, Niranjan Balasubramanian, Jesse Dodge, Annie Louis, Daniel Deutsch and Yash Kumar Lal.

• The sponsorship chair, Byron Wallace.

• The volunteer chair, Daniel Khashabi.

• The publicity chairs, Nanyun (Violet) Peng, Emily Sheng, Sameer Singh.

• The virtual infrastructure chairs, Deepak Ramachandran, Martín Villalba, and Rishita Anubhai.

• The website chairs, Ice Pasupat and Vered Shwartz for their exceptional reactivity and thorough checks of the conference schedule.

Finally, we would not have been able to organize this conference without the guidance, advice and cooperation of the following people:

• Damira Mrsic, Jernej Masnec, and Sol Rosenberg from Underline, who have been very prompt at answering all our questions and very helpful in setting up the virtual platform.

• Priscilla Rasmussen and Jenn Rachford who make all the logistics of the conference possible.

• Smaranda Muresan, Preslav Nakov, Aline Villavicencio, the Program co-Chairs of ACL 2022 who shared with us their materials and recent experience, and provided moral support.

• Anna Rumshisky, Thamar Solorio and Luke Zettlemoyer, as previous Program co-Chairs of NAACL-CL, who answered many questions and provided invaluable guidance.

• TACL Editorial Assistant Cindy Robinson, and CL Editor-in-Chief Hwee Tou Ng for coordinating TACL and CL presentations with us.

• And last but not least, our General Chair, Dan Roth, whose guidance and support throughout the process were truly invaluable: his quiet strength, big picture thinking, and respect for all the parties involved were a soothing balm and an inspiration.

We hope you will enjoy the NAACL 2022 conference!

Marie-Catherine de Marneffe, Marine Carpuat and Ivan Vladimir Meza Ruiz
NAACL 2022 Program Committee Co-Chairs
June 2022
Organizing Committee

General Chair
Dan Roth, University of Pennsylvania and AWS AI Labs

Program Chairs
Marine Carpuat, University of Maryland
Marie-Catherine de Marneffe, The Ohio State University
Ivan Vladimir Meza Ruiz, National Autonomous University of Mexico

Local Arrangements
Priscilla Rasmussen, ACL
Jennifer Rachford, ACL

Industry Track Chairs
Rashmi Gangadharaiah, Amazon
Anastassia Loukina, Grammarly
Bonan Min, Raytheon BBN Technologies

Advisors for the Industry Track
Owen Rambow, Stony Brook University
Yunyao Li, IBM Research

Tutorial Chairs
Cecilia Alm, Rochester Institute of Technology
Miguel Ballesteros, Amazon
Yulia Tsvetkov, University of Washington

Demonstration Chairs
Hannaneh Hajishirzi, University of Washington
Qiang Ning, Amazon
Avi Sil, IBM Research

Workshops Chairs
Dan Goldwasser, Purdue University
Yunyao Li, IBM Research
Ashish Sabharwal, Allen Institute for AI

Student Research Workshop Chairs
Daphne Ippolito, University of Pennsylvania
Liumin Harold Li, University of California Los Angeles
Maria Leonor Pacheco, Purdue University

Advisors for the Student Research Workshop
Danqi Chen, Princeton University
Nianwen Xue, Brandeis University

Publication Chairs
Ryan Cotterell, ETH Zürich
Danilo Croce, Tor Vergata University of Rome
Jordan Zhang,

Ethics Chairs
Kai-Wei Chang, University of California Los Angeles
Diyi Yang, Georgia Institute of Technology
Dirk Hovy, Bocconi University

Reproducibility Chairs
Niranjan Balasubramanian, Stony Brook University
Jesse Dodge, Allen Institute for AI
Annie Louis, Google
Daniel Deutsch, University of Pennsylvania
Yash Kumar Lal, Stony Brook University

Sponsorship Chair
Byron Wallace, Northeastern University

Diversity and Inclusion Chairs
Diana Galván, Tohoku University
Snigdha Chaturvedi, University of North Carolina Chapel Hill
Yonatan Bisk, Carnegie Mellon University

Academic Inclusion Chairs
Parisa Kordjamshidi, Michigan State University
Aakanksha Naik, National University of Córdoba
Khyati Chandu, National University of Córdoba

Childcare Chairs
Fatemehsadat Mireshghallah, Univeristy of California, San Diego
Abhilasha Ravichander, Carnegie Mellon University
Saadia Gabriel, Carnegie Mellon University

xiii
Accessibility Chairs
Arjun Subramonian, University of California, Los Angeles
Divyansh Kaushik, Carnegie Mellon University

Childcare Chairs
Yuntian Deng, Harvard University
Shirley Hayati, Georgia Institute of Technology

Financial Accessibility Chairs
Nedjma Ousidhoum, University of Cambridge

Advisory for the Diversity and Inclusion Committees
Pranav A, Dayta AI
Luciana Benotti, National University of Córdoba

Volunteers Chair
Daniel Khashabi, Allen Institute for AI

Publicity Chairs
Nanyun (Violet) Peng, Microsoft Research
Emily Sheng, University of North Carolina Chapel Hill
Sameer Singh, University of California Irvine

Virtual Infrastructure Chair
Deepak Ramachandran, Google
Martín Villalba, Saarland University
Rishita Anubhai, Amazon

Website Chairs
Ice Pasupat, Google
Vered Shwartz, University of British Columbia
Program Committee

Computational Social Science and Cultural Analytics
Svitlana Volkova, Pacific Northwest National Laboratory
David Bamman, University of California Berkeley

Dialogue and Interactive systems
Michel Galley, Microsoft
Kallirroi Georgila, University of Southern California
Nina Dethlefs, University of Hull
Heriberto Cuayáhuitl, University of Lincoln

Discourse and Pragmatics
Viviane Moreira, Universidade Federal do Rio Grande do Sul
Nafise Moosavi, University of Shefield

Ethics Bias and Fairness
Vinodkumar Prabhakaran, Google
Svetlana Kiritchenko, National Research Council Canada

Efficient methods in NLP
Alexandra Luccioni, Hugging Face
Roy Schwartz, Hebrew University Hebrew University of Jerusalem

Language Generation
Michael White, Ohio State University
Snigdha Chaturvedi, Department of Computer Science University of North Carolina Chapel Hill
Shashi Narayan, Google

Information Extraction
Muhao Chen, University of Southern California
Timothy Miller, Harvard University
Deepak Ramachandran, Google
Ruihong Huang, Texas A&M University

Information Retrieval and Text Mining
Sophia Ananiadou, University of Manchester
Luca Soldaini, Allen Institute for Artificial Intelligence
Interpretability and Analysis of Models for NLP
   Sebastian Gehrmann, Google Research
   Sameer Singh, University of California Irvine

Language Grounding to Vision Robotics and Beyond
   Parisa Kordjamshidi, Michigan State University
   Peter Anderson, Google

Language Resources and Evaluation
   Annemarie Friedrich, Bosch Center for Artificial Intelligence
   Sebastian Schuster, New York University
   Pradeep Dasigi, Allen Institute for Artificial Intelligence

Linguistic Theories Cognitive Modeling and Psycholinguistics
   Allyson Ettinger, University of Chicago
   Raquel Fernández, University of Amsterdam

Machine Learning for NLP - Classification and Structured Prediction Models
   He He, New York University
   Wei Xu, Georgia Institute of Technology

Machine Learning for NLP - Language Modeling and Sequence to Sequence Models
   Colin Raffel, Hugging Face
   Miguel Ballesteros, Amazon

Machine Translation
   Kevin Duh, Johns Hopkins University
   Gholamreza Haffari, Monash University
   Rachel Bawden, Inria

Multilinguality
   Dan Garrette, Google Research
   Avirup Sil, International Business Machines

NLP Applications
   Vukosi Marivate, University of Pretoria
   Helena Gomez Adorno, Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas - UNAM
   Wei Lu, Singapore University of Technology and Design
   Dan Goldwasser, Purdue University
Phonology Morphology and Word Segmentation
KMicha Elsner, The Ohio State University
Reut Tsarfaty, Bar-Ilan University

Question Answering
Mihaela Bornea, IBM International Business Machines
Jordan Boyd-Graber, University of Maryland College Park
Siva Reddy, Mila McGill University

Semantics - Lexical Semantics
Adina Williams, Facebook AI Research (Meta Platforms Inc.)
Mohammad Taher Pilehvar, Tehran Institute for Advanced Studies

Semantics - Sentence-level Semantics and Textual Inference
Xiaodan Zhu, Queen’s University
Eduardo Blanco, Arizona State University
Rachel Rudinger, University of Maryland College Park

Sentiment Analysis and Stylistic Analysis
Davide Buscaldi, Ecole polytechnique
Sara Rosenthal, International Business Machines

Speech
Adam Stiff, Infoscitex Corporation
Rolando Coto Solano, Dartmouth College

Summarization
Greg Durrett, University of Texas Austin
Jackie Cheung, Microsoft

Syntax - Tagging Chunking and Parsing
Marie Candito, Université de Paris
Lilja Øvrelid, Dept. of Informatics University of Oslo

Special Theme
Jeffrey Bigham, Carnegie Mellon University

Action Editors
Zeljko Agic, Md Shad Akhtar, Malihe Alikhani, Cecilia Alm, Mark Anderson, Jacob Andreas, Xiang Ao, Marianna Apidianaki, Yuki Arase, Mikel Artetxe, Ehsaneddin Asgari, Giuseppe Attardi, Niranjan Balasubramanian, Timothy Baldwin, Miguel Ballesteros, Mohamad Hardyman Barawi,


Ehsan Qasemi, Fanchao Qi, Jianzhong Qi, Peng Qi, Tao Qi, Dong Qian, Jing Qian, Kun Qian, Yujie Qian, Bowen Qin, Lianhui Qin, Libo Qin, Wenda Qin, Yujia Qin, Long Qiu, Xipeng Qiu, Yunqi Qiu, Ziming Qiu, Chen Qu, Xiaoye Qu


Masoud Sabet, Ashish Sabharwal, Mrinmaya Sachan, Fatiha Sadat, Arka Sadhu, Nilofar Safi Samghabadi, Benoit Sagot, Monjoy Saha, Swarnadeep Saha, Tulika Saha, Saurav Sahay, Gaurav
Tseng, Yuen-Hsien Tseng, Masaaki Tsuchida, Yoshimasa Tsuruoka, Mei Tu, Zhaopeng Tu, Nicolas Turenne, Martin Tutek, Francis Tyers

Ana Uban, Takuma Udagawa, Dennis Ulmer, Shyam Upadhyay, Tanguy Urvoy

Raúl Vázquez, Sowmya Vajjala, Josef Valvoda, Jannis Vamvas, Tim Van de Cruys, Vincent Vandehistine, Lucy Vanderwende, David Vandyke, Natalia Vanetik, Daniel Varab, Siddharth Vashishtha, Julien Velcin, Srimat Venkatapathy, Suzan Verberne, Gaurav Verma, Rakesh Verma, Giorgos Vernikos, Prashanth Vijayaraghavan, David Vilar, Jesús Vilares, Martin Villalba, Veronika Vincec, Krishnapriya Vishnubhotla, Rob Voigt, Elena Voita, Pooya Voladoddi, Soroush Vosoughi, Son Vu, Thang Vu, Thuy-Trang Vu, Tu Vu, Yogarshi Vyas, Carel van Niekerk, Rik van Noord, Jan-Willem van de Meent


Ai Xi, Congying Xia, Jingbo Xia, Patrick Xia, Qingrong Xia, Rui Xia, Yandi Xia, Yikun Xian, Jiannan Xiang, Rong Xiang, Bo Xiao, Chaojun Xiao, Chunyang Xiao, Jinghui Xiao, Lin Xiao, Liqiang Xiao, Min Xiao, Wen Xiao, Yanghua Xiao, Jun Xie, Qianqian Xie, Ruobing Xie, Tianbao Xie, Yuqiang Xie, Linzi Xing, Wenhan Xiong, Benfeng Xu, Boyan Xu, Can Xu, Canwen Xu, Chen Xu, Frank Xu, Jia Xu, Jiacheng Xu, Jing Xu, Jingjing Xu, Jiatao Xu, Jun Xu, Kun Xu, Lin Xu, Lian Xu, Peng Xu, Peng Xu, Qiantong Xu, Qiongkai Xu, Ruiying Xu, Runxin Xu, Ruochen Xu, Shusheng Xu, Wang Xu, Weijia Xu, Weiran Xu, Wenduan Xu, Xinruo Xu, Yan Xu, Yige Xu, Yumo Xu, Zenglin Xu, Zhen Xu, Zhiyang Xu

Mohit Yadav, Prateek Yadav, Yadollah Yaghoobzadeh, Ivan Yamshchikov, An Yan, Jun Yan, Rui Yan, Xu Yan, Yu Yan, Yuanmeng Yan, Baosong Yang, Changbing Yang, Fan Yang, Haiqin Yang, Jingfeng Yang, Jun Yang, Junjie Yang, Liner Yang, Linyi Yang, Min Yang, Ruosong Yang, Sen Yang, Songlin Yang, Tsung-Yen Yang, Wei Yang, Wennian Yang, Xiaoyu Yang, Yilin Yang, Yin-
Siyuan Wang, Ye Wang, Yue Wang, Zhen Wang, Christian Wartena, Guillaume Wisniewski, Jian Wu, Qingyang Wu, Xixin Wu, Zhengxuan Wu

Weijia Xu, Wenduan Xu, Yan Xu

Yadollah Yaghoobzadeh, Jun Yang, Linyi Yang, Jin-Ge Yao, Wenlin Yao, Wengpeng Yin, Bowen Yu, Weizhe Yuan

Guanhua Zhang, Jiayao Zhang, Qiang Zhang, Xiang Zhang, Xiaoying Zhang, Yichi Zhang, Zeyu Zhang, Zhixuan Zhou, Shi Zong

Outstanding Action Editors

Outstanding Reviewers
Abstract: Tools and technologies are fundamental to the human condition. They do no less than create and structure the conditions in which we live, express ourselves, enact society, and experience what it means to be human. They are also the result of our moral and technical imaginations. Yet, with our limited view, it is not at all obvious how to design and engineer tools and technology so that they are more likely to support the actions, relationships, institutions, and experiences that human beings care deeply about – a life and society of human flourishing.

Value Sensitive Design (VSD) was developed as an approach to address this challenge from within technical design processes. Drawing on over three decades of work, in this plenary talk I will provide an introduction to value sensitive design foregrounding human values in the technical design process. My remarks will present some of value sensitive design’s core theoretical constructs. Along the way, I’ll provide some examples of applying value sensitive design to robots for healthcare and to bias in computing systems as well as demonstrate one toolkit—The Envisioning Cards—in the context of a design activity.

As time permits, I will turn to a discussion of structure, scale and time: we act within existing structure in the now, from which futures unfold across time and scale. I will unpack these observations and their implications for artificial intelligence and machine learning technologies. Thinking longer-term and systematically, I will bring forward a range of potential challenges and offer some constructive ways forward. My comments will engage individual lives, society writ large, what it means to be human, the planet and beyond.

Please have scratch paper and a pencil handy for the design activity.

Bio: Batya Friedman is a Professor in the Information School and holds adjunct appointments in the Paul G. Allen School of Computer Science & Engineering, the School of Law, and the Department of Human Centered Design and Engineering at the University of Washington where she co-founded the Value Sensitive Design Lab and the UW Tech Policy Lab. Dr. Friedman pioneered value sensitive design (VSD), an
established approach to account for human values in the design of technical systems. Her work in value sensitive design has resulted in robust theoretical constructs, dozens of innovative methods, and practical toolkits such as the Envisioning Cards. Value sensitive design has been widely adopted nationally and internationally where it has been used in architecture, biomedical health informatics, civil engineering, computer security, energy, global health, human-computer interaction, human-robotic interaction, information management, legal theory, moral philosophy, tech policy, transportation, and urban planning, among others. Additionally, value sensitive design is emerging in higher education, government, and industry as a key approach to address computing ethics and responsible innovation. Today, Dr. Friedman is working on open questions in value sensitive design including multi-lifespan design, and designing for and with non-human stakeholders – questions critical for the wellbeing of human societies and the planet.

Dr. Friedman’s 2019 MIT Press book co-authored with David Hendry, Value Sensitive Design: Shaping Technology with Moral Imagination, provides a comprehensive account of value sensitive design. In 2012 Dr. Friedman received the ACM-SIGCHI Social Impact Award and the University Faculty Lecturer award at the University of Washington, in 2019 she was inducted into the CHI Academy, in 2020 she received an honorary doctorate from Delft University of Technology, and in 2021 she was recognized as an ACM Fellow. She is also a stone sculptor and mixed media artist. Dr. Friedman received both her B.A. and Ph.D. from the University of California at Berkeley.
Keynote Talk: NLP in Mexican Spanish: One of many stories
Manuel Montes-y-Gómez
National Institute of Astrophysics, Optics and Electronics (INAOE)

Abstract:

Wednesday, July 13, 2022 - Room: Columbia A/C/D & 302 Beckler - Time: 16:15-17:15

Abstract: Spanish is one of the most widely spoken languages in the world, however, the development of language technologies for it has not been in the same proportion. This is particularly true for some of its Latin American variants, such as the Mexican Spanish. This talk will focus on presenting the development of NLP for Mexican Spanish, emphasizing one of its many research stories related to the analysis of social media content.

This talk will present some data on the languages spoken in Mexico and on the development of the area of Natural Language Processing in our country, and will describe a research project that combined the efforts of several groups: the identification of abusive language in Mexican tweets. The talk will conclude by exposing some calls for collaboration, with the intention of increasing and improving the research in Mexican Spanish as well as in the many indigenous languages spoken in Mexico.

Bio: Manuel Montes-y-Gómez is Full Professor at the National Institute of Astrophysics, Optics and Electronics (INAOE) of Mexico. His research is on automatic text processing. He is author of more than 250 journal and conference papers in the fields of information retrieval, text mining and authorship analysis.

He has been visiting professor at the Polytechnic University of Valencia (Spain), and the University of Alabama (USA). He is also a member of the Mexican Academy of Sciences (AMC), and founding member of the Mexican Academy of Computer Science (AMEXCOMP), the Mexican Association of Natural Language Processing (AMNLP), and of the Language Technology Network of CONACYT. In the context of them, he has been the organizer of the National Workshop on Language Technologies (from 2004 to 2016), the Mexican Workshop on Plagiarism Detection and Authorship Analysis (2016-2020), the Mexican Autumn School on Language Technologies (2015 and 2016), and a shared task on author profiling, aggressiveness analysis and fake news detection in Mexican Spanish at IberLEF (2018-2021).
Panel: “The Place of Linguistics and Symbolic Structures”

Tuesday, July 12, 2022 - Room: Columbia A/C/D & 302 Beckler - Time: 9:15-10:15

The widespread adoption of neural models in NLP research and the fact that NLP applications increasingly mediate people’s lives have prompted many discussions about what productive research directions might look like for our community. Since NAACL is a meeting of a chapter of the Association for Computational Linguistics, we would like to highlight specifically the role that linguistics and symbolic structures can play (or not) in shaping these research directions.

Moderator: Dan Roth, University of Pennsylvania & AWS AI Labs

Bio: Dan Roth is the Eduardo D. Glandt Distinguished Professor at the Department of CIS, UPenn, the NLP Lead at AWS AI, and a Fellow of the AAAS, ACM, AAAI, and ACL. In 2017 Roth received the John McCarthy Award. Roth has published broadly in ML, NLP, KRR, and learning theory, and has given keynote talks and tutorials in all ACL and AAAI major conferences. Roth was the Editor-in-Chief of JAIR until 2017, and the program chair of AAAI’11, ACL’03 and CoNLL’02.

Emily M. Bender, University of Washington

Bio: Emily M. Bender is a Professor of Linguistics at the University of Washington and the Faculty Director of UW’s Professional Master’s in Computational Linguistics. Her research interests include computational semantics, multilingual grammar engineering, the interplay between linguistics and NLP, and societal impacts of language technology. She is the author of two books which present linguistic concepts in a manner accessible to NLP practitioners: Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax (2013) and Linguistic Fundamentals for Natural Language Processing II: 100 Essentials from Semantics and Pragmatics (2019; with Alex Lascarides), as well as the co-author of recent influential papers such as Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data (ACL 2020) and On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? (FAcct 2021).

Dilek Hakkani-Tür, Amazon Alexa AI

Bio: Dilek Hakkani-Tür is a senior principal scientist at Amazon Alexa AI, focusing on enabling natural dialogues with machines. Prior to joining Amazon, she was a researcher at Google, Microsoft Research, International Computer Science Institute at UC Berkeley and AT&T Labs-Research. Her research interests include conversational AI, natural language and speech processing, spoken dialogue systems, and machine learning for language processing. She received best paper awards for publications she co-authored on conversational systems from IEEE Signal Processing Society, ISCA and EURASIP. Recently, she served as a program chair for NAACL 2020, the editor-in-chief of IEEE Transactions on Audio, Speech, and Language Processing and an IEEE Distinguished Industry Speaker. She is a fellow of ISCA and IEEE.

Chitta Baral, Arizona State University
Bio: Chitta Baral is a Professor in the School of Computing and AI at Arizona State University. His research interests include Knowledge Representation and Reasoning (KR & R), Natural Language Understanding (NLU), Image/Video Understanding; and their applications to Molecular Biology, Health Informatics and Robotics. Chitta is the author of the book “Knowledge Representation, Reasoning and Declarative Problem Solving” and a past President of KR Inc. His current research focus is on leveraging decades of research in KR & R for better understanding of natural language and images/videos. Towards that end he has worked on a framework for translating natural language to formal representations (NL2KR); abducting missing knowledge and knowledge hunting; exploring NLU challenges where reasoning with knowledge, reasoning about actions, and commonsense reasoning are crucial; exploring the use of natural language as a knowledge representation and instructional formalism; and exploring the role of reasoning and knowledge in enhancing generalizability, robustness, and few-shot learning.

Christopher D. Manning, Stanford University
Bio: Christopher Manning is a professor of linguistics and computer science at Stanford University, Director of the Stanford Artificial Intelligence Lab (SAIL), and an Associate Director of the Stanford Institute for Human-Centered AI (HAI). He is a leader in applying deep neural networks to natural language processing (NLP), including work on neural machine translation, tree-recursive models, natural language inference, summarization, parsing, question answering, and the GloVe word vectors. Manning founded the Stanford NLP group (@stanfordnlp), teaches and has co-written textbooks for NLP (CS 224N) and information retrieval (CS 276), co-developed Stanford Dependencies and Universal Dependencies, manages development of the Stanford CoreNLP and Stanza software, is the most-cited researcher in NLP, and is an ACM, AAAI, and ACL Fellow and a Past President of ACL.
The Careers in NLP Panel is a standing feature of NAACL Industry Track. The panel is addressed to graduate students and junior researchers as well as their supervisors and mentors, although all NAACL participants are welcomed. The panellists will discuss the diversity of career paths in NLP: from more research-oriented NLP scientist roles to careers in product.

**Moderator: Yunyao Li, Apple Knowledge Platform**

**Bio:** Yunyao Li is the Head of Machine Learning, Apple Knowledge Platform, where her team builds the next-generation machine learning solutions to help power features such as Siri and Spotlight. Previously she was a Distinguished Research Staff Member and Senior Research Manager at IBM Research - Almaden. She is particularly known for her work in scalable NLP, enterprise search, and database usability. She has built systems, developed solutions, and delivered core technologies to over 20 IBM products under brands such as Watson, InfoSphere, and Cognos. She has published over 80 articles with multiple awards and a book. She was an IBM Master Inventor, with over 50 patents filed/granted. She is an ACM Distinguished Member. She was a member of the inaugural New Voices program of the US National Academies (1 out of 18 selected nationwide) and represented US young scientists at World Laureates Forum Young Scientists Forum in 2019 (1 of 4 selected nationwide).

**Yang Liu, Amazon, Alexa AI**

**Bio:** Yang Liu is currently a principal scientist at Amazon, Alexa AI. Her research interest is in speech and language processing. She received her BS and MS from Tsinghua University, and Ph.D. from Purdue University. Before joining Amazon, she was the head of LAIX Silicon Valley AI lab, a research scientist at Facebook, visiting scientist at Google, a faculty member at the University of Texas at Dallas, and researcher at ICSI in Berkeley. She received NSF CAREER award and Air Force Young Investigator Program award. She is currently a member of the IEEE SLTC committee, a senior area editor for IEEE/ACM Transactions on Audio, Speech and Language Processing, an action editor for TACL. She was one of the program chairs for EMNLP 2020, and has served regularly as an area chair and reviewer in the past NLP conferences. She is a fellow of IEEE and ISCA.

**Timo Mertens, Grammarly**

**Bio:** Timo Mertens is the Head of Machine Learning & NLP Products at Grammarly. In his role, he oversees the teams that design and build products that use machine learning and natural language processing. These technologies empower Grammarly to offer a digital writing assistant that helps millions of users write more clearly and effectively every day. Timo has focused on the intersection between machine learning and delivering impactful products throughout his career, spanning academia—with a Ph.D. in Speech Recognition—and industry, where he’s held product leadership positions across Microsoft, Google, and Dropbox.
Thamar Solorio, University of Houston and Bloomberg LP
Bio: Thamar Solorio is a Professor of Computer Science at the University of Houston (UH) and she is also a visiting scientist at Bloomberg LP. She holds graduate degrees in Computer Science from the Instituto Nacional de Astrofísica, Óptica y Electrónica, in Puebla, Mexico. Her research interests include information extraction from social media data, enabling technology for code-switched data, stylistic modelling of text, and more recently multimodal approaches for online content understanding. She is the director and founder of the Research in Text Understanding and Language Analysis Lab at UH. She is the recipient of an NSF CAREER award for her work on authorship attribution, and recipient of the 2014 Emerging Leader ABIE Award in Honor of Denice Denton. She is currently serving a second term as an elected board member of the North American Chapter of the Association of Computational Linguistics.

Luke Zettlemoyer, University of Washington and Meta
Bio: Luke Zettlemoyer is a Professor in the Paul G. Allen School of Computer Science & Engineering at the University of Washington, and a Research Scientist at Meta. His research focuses on empirical methods for natural language semantics, and involves designing machine learning algorithms, introducing new tasks and datasets, and, most recently, studying how to best develop self-supervision signals for pre-training. His honors include being named an ACL Fellow as well as winning a PECASE award, an Allen Distinguished Investigator award, and multiple best paper awards.
# Table of Contents

**Social Norms Guide Reference Resolution**  
Mitchell Abrams and Matthias Scheutz .................................................. 1

**Learning Natural Language Generation with Truncated Reinforcement Learning**  
Alice Martin, Guillaume Quispe, Charles Ollion, Sylvain Le Corff, Florian Strub and Olivier Pietquin ................................................................. 12

**Language Model Augmented Monotonic Attention for Simultaneous Translation**  
Sathish Reddy Indurthi, Mohd Abbas Zaidi, Beomseok Lee, Nikhil Kumar Lakumarapu and Sangha Kim .............................................................. 38

**What Makes a Good and Useful Summary? Incorporating Users in Automatic Summarization Research**  
Maartje Ter Hoeve, Julia Kiseleva and Maarten de Rijke ............................................. 46

**ErAConD: Error Annotated Conversational Dialog Dataset for Grammatical Error Correction**  
Xun Yuan, Derek Pham, Sam Davidson and Zhou Yu ........................................... 76

**Semantic Diversity in Dialogue with Natural Language Inference**  
Katherine Stasaski and Marti Hearst .............................................................. 85

**LEA: Meta Knowledge-Driven Self-Attentive Document Embedding for Few-Shot Text Classification**  
S. K. Hong and Tae Young Jang ................................................................. 99

**Enhancing Self-Attention with Knowledge-Assisted Attention Maps**  
Jiangang Bai, Yujing Wang, Hong Sun, Ruonan Wu, Tianmeng Yang, Pengfei Tang, Defu Cao, Mingliang Zhang1, Yunhai Tong, Yaming Yang, Jing Bai, Ruofei Zhang, Hao Sun and Wei Shen ......................................................... 107

**Batch-Softmax Contrastive Loss for Pairwise Sentence Scoring Tasks**  
Anton Chernyavskiy, Dmitry Ilvovsky, Pavel Kalinin and Preslav Nakov ......................... 116

**NewsEdits: A News Article Revision Dataset and a Novel Document-Level Reasoning Challenge**  
Alexander Spangher, Xiang Ren, Jonathan May and Nanyun Peng ........................................... 127

**Putting the Con in Context: Identifying Deceptive Actors in the Game of Mafia**  
Samee Omotayo Ibraheem, Gaoyue Zhou and John DeNero ........................................... 158

**SUBS: Subtree Substitution for Compositional Semantic Parsing**  
Jingfeng Yang, Le Zhang and Diyi Yang .................................................... 169

**Two Contrasting Data Annotation Paradigms for Subjective NLP Tasks**  
Paul Rottger, Bertie Vidgen, Dirk Hovy and Janet B. Pierrehumbert ................................. 175

**Do Deep Neural Nets Display Human-like Attention in Short Answer Scoring?**  
Zijie Zeng, Xinyu Li, Dragan Gasevic and Guanliang Chen ........................................... 191

**Knowledge-Grounded Dialogue Generation with a Unified Knowledge Representation**  
Yu Li, Baolin Peng, Yelong Shen, Yi Mao, Lars Liden, Zhou Yu and Jianfeng Gao ......................... 206

**CERES: Pretraining of Graph-Conditioned Transformer for Semi-Structured Session Data**  
Rui Feng, Chen Luo, Qingyu Yin, Bing Yin, Tu Zhao and Chao Zhang ................................. 219

**Political Ideology and Polarization: A Multi-dimensional Approach**  
Barea Sinno, Bernardo Oviedo, Katherine Atwell, Malihe Alikhani and Junyi Jessy Li ........... 231
Cooperative Self-training of Machine Reading Comprehension
Hongyin Luo, Shang-Wen Li, Mingye Gao, Seunghak Yu and James R. Glass ................. 244

GlobEnc: Quantifying Global Token Attribution by Incorporating the Whole Encoder Layer in Transformers
Ali Modarressi, Mohsen Fayyaz, Yadollah Yaghoobzadeh and Mohammad Taher Pilehvar ........ 258

A Robustly Optimized BMRC for Aspect Sentiment Triplet Extraction
Shu Liu, Kaiwen Li and Zuhe Li .................................................................................. 272

Seed-Guided Topic Discovery with Out-of-Vocabulary Seeds
Yu Zhang, Yu Meng, Xuan Wang, Sheng Wang and Jiawei Han ................................. 279

Towards Process-Oriented, Modular, and Versatile Question Generation that Meets Educational Needs
Xu Wang, Simin Fan, Jessica Houghton and Lu Wang ................................................. 291

SwahBERT: Language Model of Swahili
Gati L Martin, Medard Edmund Mswahili, Young-Seob Jeong and Jeong Young-Seob ....... 303

Deconstructing NLG Evaluation: Evaluation Practices, Assumptions, and Their Implications
Kaitlyn Zhou, Su Lin Blodgett, Adam Trischler, Hal Daumé III, Kaheer Suleman and Alexandra Olteanu ................................................................. 314

TSTR: Too Short to Represent, Summarize with Details! Intro-Guided Extended Summary Generation
Sajad Sotudeh and Nazli Goharian ................................................................. 325

Empathic Machines: Using Intermediate Features as Levers to Emulate Emotions in Text-To-Speech Systems
Saiteja Kosgi, Sarath Sivaprasad, Niranjan Pedanekar, Anil Kumar Nelakanti and Vineet Gandhi ........ 336

The Why and The How: A Survey on Natural Language Interaction in Visualization
Henrik Voigt, Ozge Alacam, Monique Meuschke, Kai Lawonn and Sina Zarrieß ............ 348

Understand before Answer: Improve Temporal Reading Comprehension via Precise Question Understanding
Hao Huang, Xiubo Geng, Guodong Long and Daxin Jiang ...................................... 375

User-Driven Research of Medical Note Generation Software
Tom Knoll, Francesco Moramarco, Alex Papadopoulos Korfiatis, Rachel Young, Claudia Ruffini, Mark Perera, Christian Perstl, Ehud Reiter, Anya Belz and Aleksandar Savkov 385

Ask Me Anything in Your Native Language
Nikita Sorokin, Dmitry Abulkhanov, Irina Piontkovskaya and Valentin Malykh ............ 395

Diversifying Neural Dialogue Generation via Negative Distillation
Yiwei Li, Shaoxiong Feng, Bin Sun and Kan Li ......................................................... 407

On Synthetic Data for Back Translation
Jiahao Xu, Yubin Ruan, Wei Bi, Guoping Huang, Shuming Shi, Lihui Chen and Lemao Liu . 419

Mapping the Design Space of Human-AI Interaction in Text Summarization
Ruijia Cheng, Alison Smith-Renner, Ke Zhang, Joel R. Tetreault and Alejandro Jaimes-Larrarte 431

Towards Robust and Semantically Organised Latent Representations for Unsupervised Text Style Transfer
Sharan Narasimhan, Suvodip Dey and Maunendra Sankar Desarkar .......................... 456

xxxix
An Exploration of Post-Editing Effectiveness in Text Summarization
Vivian Lai, Alison Smith-Renner, Ke Zhang, Ruijia Cheng, Wenjuan Zhang, Joel R. Tetreault and Alejandro Jaimes-Larrarte .............................................................. 475

Automatic Correction of Human Translations
Jessy Lin, Geza Kovacs, Aditya Shastry, Joern Wuebker and John DeNero ............... 494

On the Robustness of Reading Comprehension Models to Entity Renaming
Jun Yan, Yang Xiao, Sagnik Mukherjee, Bill Yuchen Lin, Robin Jia and Xiang Ren .... 508

Explaining Why: How Instructions and User Interfaces Impact Annotator Rationales When Labeling Text Data
Cynthia Sullivan, William Brackenbury, Andrew Michael McNut, Kevin Bryson, kbyllofficial@gmail.com kbyllofficial@gmail.com, Yuxin Chen, Michael Littman, Chenhao Tan and Blase Ur ............. 521

Fine-tuning Pre-trained Language Models for Few-shot Intent Detection: Supervised Pre-training and Isotropization
Haode Zhang, Haowen Liang, Yuwei Zhang, Li-Ming Zhan, Xiao-Ming Wu, Xiaolei Lu and Albert Ys. Lam ................................................................. 532

Cross-document Misinformation Detection based on Event Graph Reasoning
Xueqing Wu, Kung-Hsiang Huang, Yi Fung and Heng Ji .................................. 543

Disentangled Action Recognition with Knowledge Bases
Zhekun Luo, Shalini Ghosh, Devin Guillory, Keizo Kato, Trevor Darrell and Huijuan Xu . 559

Machine-in-the-Loop Rewriting for Creative Image Captioning
Vishakh Padmakumar and He He ................................................................. 573

A Word is Worth A Thousand Dollars: Adversarial Attack on Tweets Fools Stock Prediction
Yong Xie, Dakuo Wang, Pin-Yu Chen, Jinjun Xiong, Sijia Liu and Oluwasanmi O Koyejo . 587

Building Multilingual Machine Translation Systems That Serve Arbitrary XY Translations
Akiko Eriguchi, Shufang Xie, Tao Qin and Hany Hassan ................................. 600

Non-Autoregressive Neural Machine Translation with Consistency Regularization Optimized Variational Framework
Minghao Zhu, Junli Wang and Chungen Yan ............................................. 607

User-Centric Gender Rewriting
Bashar Alhafni, Nizar Habash and Houda Bouamor ..................................... 618

Reframing Human-AI Collaboration for Generating Free-Text Explanations
Sarah Wiegreffe, Jack Hessel, Swabha Swayamdipta, Mark Riedl and Yejin Choi . 632

EmRel: Joint Representation of Entities and Embedded Relations for Multi-triple Extraction
Benfeng Xu, Quan Wang, Yajuan Lyu, Yabing Shi, Yong Zhu, Jie Gao and Zhendong Mao . 659

Meta Learning for Natural Language Processing: A Survey
Hung-yi Lee, Shang-Wen Li and Thang Vu ............................................... 666

Analyzing Modality Robustness in Multimodal Sentiment Analysis
Devamanyu Hazarika, Yingting Li, Bo Cheng, Shuai Zhao, Roger Zimmermann and Soujanya Poria ................................................................. 685

Fuse It More Deeply! A Variational Transformer with Layer-Wise Latent Variable Inference for Text Generation
Jinyi Hu, Xiaoyuan Yi, Wenhao Li, Maosong Sun and Xing Xie .......................... 697
Easy Adaptation to Mitigate Gender Bias in Multilingual Text Classification
Xiaolei Huang ................................................................. 717

On the Use of External Data for Spoken Named Entity Recognition
Ankita Pasad, Felix Wu, Suwon Shon, Karen Livescu and Kyu Han ......................... 724

Long-term Control for Dialogue Generation: Methods and Evaluation
Ramya Ramakrishnan, Hashan Buddhika Narangodage, Mauro Schilman, Kilian Q Weinberger and Ryan McDonald ................................................................. 738

Learning Dialogue Representations from Consecutive Utterances
Zhihan Zhou, Dejiao Zhang, Wei Xiao, Nicholas Dingwall, Xiaofei Ma, Andrew Arnold and Bing Xiang ................................................................. 754

On the Machine Learning of Ethical Judgments from Natural Language
Zeerak Talat, Hagen Blix, Josef Valvoda, Maya Indira Ganesh, Ryan Cotterell and Adina Williams 769

NeuroLogic A*esque Decoding: Constrained Text Generation with Lookahead Heuristics
Ximing Lu, Sean Welleck, Peter West, Liwei Jiang, Jungo Kasai, Daniel Khashabi, Ronan Le Bras, Lianhui Qin, Youngjae Yu, Rowan Zellers, Noah Smith and Yejin Choi ......................... 780

PARADISE: Exploiting Parallel Data for Multilingual Sequence-to-Sequence Pretraining
Machel Reid and Mikel Artetxe ................................................................. 800

Explaining Toxic Text via Knowledge Enhanced Text Generation
Rohit Sridhar and Diyi Yang ................................................................. 811

Teaching BERT to Wait: Balancing Accuracy and Latency for Streaming Disfluency Detection
Angelica Chen, Vicky Zayats, Daniel David Walker and Dirk Padfield ....................... 827

GRAM: Fast Fine-tuning of Pre-trained Language Models for Content-based Collaborative Filtering
Yoonseok Yang, Kyu Seok Kim, Minsam Kim and Juneyoung Park ................................................................. 839

Generating Repetitions with Appropriate Repeated Words
Toshiki Kawamoto, Hidetaka Kamigaito, Kotaro Funakoshi and Manabu Okumura 852

Textless Speech-to-Speech Translation on Real Data
Ann Lee, Hongyu Gong, Paul-Ambroise Duquenne, Holger Schwenk, Peng-Jen Chen, Changhan Wang, Sravya Popuri, Yossi Adi, Juan Pino, Jiatao Gu and Wei-Ning Hsu ......................... 860

WALNUT: A Benchmark on Semi-weakly Supervised Learning for Natural Language Understanding
Guoqing Zheng, Giannis Karamanolakis, Kai Shu and Ahmed Hassan Awadallah 873

CoSIm: Commonsense Reasoning for Counterfactual Scene Imagination
Hyoughun Kim, Abhay Zala and Mohit Bansal ................................................................. 911

Abstraction not Memory: BERT and the English Article System
Harish Tayyar Madabushi, Dagmar Divjak and Petar Milin ................................................................. 924

Omnitab: Pretraining with Natural and Synthetic Data for Few-shot Table-based Question Answering
Zhengbao Jiang, Yi Mao, Pengcheng He, Graham Neubig and Weizhu Chen ................................................................. 932

Provably Confidential Language Modelling
Xuandong Zhao, Lei Li and Yu-Xiang Wang ................................................................. 943
KAT: A Knowledge Augmented Transformer for Vision-and-Language
Liangke Gui, Borui Wang, Qiujuan Huang, Alexander G Hauptmann, Yonatan Bisk and Jianfeng Gao ................................................................. 956

When a sentence does not introduce a discourse entity, Transformer-based models still sometimes refer to it
Sebastian Schuster and Tal Linzen .......................................................... 969

On Curriculum Learning for Commonsense Reasoning
Adyasha Maharana and Mohit Bansal ................................................... 983

DocTime: A Document-level Temporal Dependency Graph Parser
Puneet Mathur, Vlad I Morariu, Verena Kaynig-Fittkau, Jiuxiang Gu, Franck Dernoncourt, Quan Hung Tran, Ani Nenkova, Dinesh Manocha and Rajiv Jain ........................................ 993

FactPEGASUS: Factuality-Aware Pre-training and Fine-tuning for Abstractive Summarization
David Wan and Mohit Bansal ................................................................. 1010

ScAN: Suicide Attempt and Ideation Events Dataset
Bhanu Pratap Singh Rawat, Samuel Kovaly, Hong Yu and Wilfred Pigeon .............. 1029

Socially Aware Bias Measurements for Hindi Language Representations
Vijit Malik, Sunipa Dev, Akihiro Nishi, Nanyun Peng and Kai-Wei Chang ................ 1041

AmbiPun: Generating Humorous Puns with Ambiguous Context
Anirudh Mittal, Yufei Tian and Nanyun Peng ........................................... 1053

EmpHi: Generating Empathetic Responses with Human-like Intents
Mao Yan Chen, Siheng Li and Yujiu Yang .................................................. 1063

Yes, No or IDK: The Challenge of Unanswerable Yes/No Questions
Elior Sulem, Jamaal Hay and Dan Roth .................................................... 1075

Inducing and Using Alignments for Transition-based AMR Parsing
Andrew Drozdov, Jiawei Zhou, Radu Florian, Andrew McCallum, Tahira Naseem, Yoon Kim and Ramón Fernandez Astudillo ........................................... 1086

Masked Part-Of-Speech Model: Does Modeling Long Context Help Unsupervised POS-tagging?
Xiang Zhou, Shiyue Zhang and Mohit Bansal .......................................... 1099

DREAM: Improving Situational QA by First Elaborating the Situation
Yuling Gu, Bhavana Dalvi and Peter Clark ................................................ 1115

CoSe-Co: Text Conditioned Generative CommonSense Contextualizer
Rachit Bansal, Milan Aggarwal, Sumit Bhatia, Jivat Neet Kaur and Balaji Krishnamurthy . 1128

Probing via Prompting
Jiaoda Li, Ryan Cotterell and Mrinmaya Sachan ...................................... 1144

Database Search Results Disambiguation for Task-Oriented Dialog Systems
Kun Qian, Satwik Kottur, Ahmad Beirami, Shahin Shayandeh, Paul A. Crook, Alborz Geramifard, Zhou Yu and Chinnadhurai Sankar .............................................. 1158

Unsupervised Slot Schema Induction for Task-oriented Dialog
Dian Yu, Mingqiu Wang, Yuan Cao, Izhak Shafran, Laurent El Shafey and Hagen Soltau . 1174
Towards a Progression-Aware Autonomous Dialogue Agent
Abraham Sanders, Tomek Strzalkowski, Mei Si, Albert Chang, Deepanshu Dey, Jonas Braasch and Dakuo Wang ............................... 1194

Cross-Domain Detection of GPT-2-Generated Technical Text
Juan Diego Rodriguez, Todd Hay, David Gros, Zain Shamsi and Ravi Srinivasan ............ 1213

DISAPERE: A Dataset for Discourse Structure in Peer Review Discussions
Neha Nayak Kennard, Tim O’Gorman, Rajarshi Das, Akshay Sharma, Chhandak Bagchi, Matthew Clinton, Pranay Kumar Yelugam, Hamed Zamani and Andrew McCallum .................. 1234

MultiSpanQA: A Dataset for Multi-Span Question Answering
Haonan Li, Martin Tomko, Maria Vasardani and Timothy Baldwin .......................... 1250

Context-Aware Abbreviation Expansion Using Large Language Models
Shanqing Cai, Subhashini Venugopalan, Katrin Tomanek, Ajit Narayanan, Meredith Ringel Morris and Michael Brenner .................................. 1261

Theory-Grounded Measurement of U.S. Social Stereotypes in English Language Models
Yang Trista Cao, Anna Sotnikova, Hal Daumé III, Rachel Rudinger and Linda Zou ........... 1276

Sort by Structure: Language Model Ranking as Dependency Probing
Max Müllner-Eberstein, Rob Van Der Goot and Barbara Plank ................................. 1296

Quantifying Synthesis and Fusion and their Impact on Machine Translation
Arturo Oncevay, Duygu Ataman, Niels Van Berkel, Barry Haddow, Alexandra Birch and Johannes Bjerva ......................................................... 1308

Commonsense and Named Entity Aware Knowledge Grounded Dialogue Generation
Deeksha Varshney, Akshara Prabhakar and Asif Ekbal ........................................... 1322

Efficient Hierarchical Domain Adaptation for Pretrained Language Models
Alexandra Chronopoulou, Matthew E Peters and Jesse Dodge ............................... 1336

Hatemoji: A Test Suite and Adversarially-Generated Dataset for Benchmarking and Detecting Emoji-Based Hate
Hannah Rose Kirk, Bertie Vidgen, Paul Rottger, Tristan Thrush and Scott A. Hale .......... 1352

On the Economics of Multilingual Few-shot Learning: Modeling the Cost-Performance Trade-offs of Machine Translated and Manual Data
Kabir Ahuja, Monojit Choudhury and Sandipan Dandapat ....................................... 1369

Learning to Selectively Learn for Weakly Supervised Paraphrase Generation with Model-based Reinforcement Learning
Haiyan Yin, Dingcheng Li and Ping Li ................................................................. 1385

Quality-Aware Decoding for Neural Machine Translation
Patrick Fernandes, António Farinhias, Ricardo Rei, José G. C. De Souza, Perez Ogayo, Graham Neubig and Andre Martins ........................................ 1396

Pretrained Models for Multilingual Federated Learning
Orion Weller, Marc Marone, Vladimir Braverman, Dawn Lawrie and Benjamin Van Durme 1413

AcTune: Uncertainty-Based Active Self-Training for Active Fine-Tuning of Pretrained Language Models
Yue Yu, Lingkai Kong, Jieyu Zhang, Rongzhi Zhang and Chao Zhang ....................... 1422
Label Anchored Contrastive Learning for Language Understanding  
Zhenyu Zhang, Yuming Zhao, Meng Chen and Xiaodong He .................................................. 1437

Go Back in Time: Generating Flashbacks in Stories with Event Temporal Prompts  
Rujun Han, Hong Chen, Yufei Tian and Nanyun Peng .......................................................... 1450

Forecasting COVID-19 Caseloads Using Unsupervised Embedding Clusters of Social Media Posts  
Felix Drinkall, Stefan Zohren and Janet B. Pierrrehumbert .................................................. 1471

Many Hands Make Light Work: Using Essay Traits to Automatically Score Essays  
Rahul Kumar, Sandeep Mathias, Sriparna Saha and Pushpak Bhattacharyya ......................... 1485

Natural Language Inference with Self-Attention for Veracity Assessment of Pandemic Claims  
Miguel Arana-Catania, Elena Kochkina, Arkaitz Zubiaga, Maria Liakata, Robert Procter and Yulan He .......................................................... 1496

Beyond Emotion: A Multi-Modal Dataset for Human Desire Understanding  
Ao Jia, Yu He, Yazhou Zhang, Sagar Uprety, Dawei Song and Christina Lioma ................. 1512

Relation-Specific Attentions over Entity Mentions for Enhanced Document-Level Relation Extraction  
Jiaxin Yu, Deqing Yang and Shuyu Tian .......................................................... 1523

Twitter-COMMs: Detecting Climate, COVID, and Military Multimodal Misinformation  
Giscard Biamby, Grace Luo, Trevor Darrell and Anna Rohrbach ....................................... 1530

BLonDe: An Automatic Evaluation Metric for Document-level Machine Translation  
Yuchen Eleanor Jiang, Tianyu Liu, Shuming Ma, Dongdong Zhang, Jian Yang, Haoyang Huang, Rico Sennrich, Ryan Cotterell, Mrinmaya Sachan and Ming Zhou .................................................. 1550

Disentangled Learning of Stance and Aspect Topics for Vaccine Attitude Detection in Social Media  
Lixing Zhu, Zheng Fang, Gabriele Pergola, Robert Procter and Yulan He .......................... 1566

SKILL: Structured Knowledge Infusion for Large Language Models  
Fedor Moiseev, Zhe Dong, Enrique Alfonseca and Martin Jaggi ........................................ 1581

Same Neurons, Different Languages: Probing Morphosyntax in Multilingual Pre-trained Models  
Karolina Stanczak, Edosardo Ponti, Lucas Torroba Hennigen, Ryan Cotterell and Isabelle Augenstein .......................................................... 1589

Aspect Is Not You Need: No-aspect Differential Sentiment Framework for Aspect-based Sentiment Analysis  
Jiahao Cao, Rui Liu, Huailiang Peng, Lei Jiang and Xu Bai ............................................... 1599

MoEBERT: from BERT to Mixture-of-Experts via Importance-Guided Adaptation  
Simiao Zuo, Qingru Zhang, Chen Liang, Pengcheng He, Tuo Zhao and Weizhu Chen .......... 1610

Implicit n-grams Induced by Recurrence  
Xiaobing Sun and Wei Lu .................................................................................................. 1624

Guiding Visual Question Generation  
Nihir Vedd, Zixu Wang, Marek Rei, Yishu Miao and Lucia Specia .................................. 1640

OPERA: Operation-Pivoted Discrete Reasoning over Text  
Yongwei Zhou, Junwei Bao, Chaouqu Duan, Haipeng Sun, Jiahui Liang, Yifan Wang, Jing Zhao, Youzheng Wu, Xiaodong He and Tiejun Zhao .................................................. 1655

Improving Multi-Document Summarization through Referenced Flexible Extraction with Credit-Awareness  
Yun-Zhu Song, Yi-Syuan Chen and Hong-Han Shuai ....................................................... 1667

xliv
Improving Constituent Representation with Hypertree Neural Networks
Hao Zhou, Gongshen Liu and Kewei Tu .......................... 1682

Measuring Fairness with Biased Rulers: A Comparative Study on Bias Metrics for Pre-trained Language Models
Pieter Delobelle, Ewoenam Kwaku Tokpo, Toon Calders and Bettina Berendt ............... 1693

MuCPAD: A Multi-Domain Chinese Predicate-Argument Dataset
Yahui Liu, Haoping Yang, Chen Gong, Qingrong Xia, Zhenghua Li and Min Zhang ........ 1707

Representation Learning for Conversational Data using Discourse Mutual Information Maximization
Bishal Santra, Sumegh Roychowdhury, Aishik Mandal, Vasu Gurram, Atharva Naik, Manish Gupta and Pawan Goyal .................................................. 1718

ValCAT: Variable-Length Contextualized Adversarial Transformations Using Encoder-Decoder Language Model
Chuyun Deng, Mingxuan Liu, Yue Qin, Jia Zhang, Hai-Xin Duan and Donghong Sun ..... 1735

A Study of Syntactic Multi-Modality in Non-Autoregressive Machine Translation
Kexun Zhang, Rui Wang, Xu Tan, Junliang Guo, Yi Ren, Tao Qin and Tie-Yan Liu ........ 1747

ClAug: Equipping Interpolative Augmentation with Curriculum Learning
Ramit Sawhney, Ritesh Singh Soun, Shrey Pandit, Megh Thakkar, Sarvagya Malaviya and Yuval Pinter ....................................................... 1758

Proposition-Level Clustering for Multi-Document Summarization
Ori Ernst, Avi Caculatoru, Ori Shapira, Ramakanth Pasunuru, Mohit Bansal, Jacob Goldberger and Ido Dagan ........................................ 1765

Non-Autoregressive Machine Translation: It’s Not as Fast as it Seems
Jindrich Helcl, Barry Haddow and Alexandra Birch ........................................... 1780

BAD-X: Bilingual Adapters Improve Zero-Shot Cross-Lingual Transfer
Marinela Parović, Goran Glavaš, Ivan Vulić and Anna Korhonen .......................... 1791

Combining Humor and Sarcasm for Improving Political Parody Detection
Xiao Ao, Danae Sanchez Villegas, Daniel Preotiuc-Pietro and Nikolaos Aletras .......... 1800

TIE: Topological Information Enhanced Structural Reading Comprehension on Web Pages
Zihan Zhao, Lu Chen, Ruisheng Cao, Hongshen Xu, Xingyu Chen and Kai Yu .......... 1808

RSTGen: Imbuing Fine-Grained Interpretable Control into Long-Form Text Generators
Rilwan Akanni Adewoyin, Ritabrata Dutta and Yulan He .................................. 1822

Intent Detection and Discovery from User Logs via Deep Semi-Supervised Contrastive Clustering
Rajat Kumar, Mayur Patidar, Vaibhav Varshney, Lovekesh Vig and Gautam Shroff .... 1836

Extending Multi-Text Sentence Fusion Resources via Pyramid Annotations
Daniela Brook Weiss, Paul Roit, Ori Ernst and Ido Dagan .................................. 1854

The Devil is in the Details: On the Pitfalls of Vocabulary Selection in Neural Machine Translation
Tobias Domhan, Eva Hasler, Ke Tran, Sony Trenous, Bill Byrne and Felix Hieber .... 1861

MultiCite: Modeling realistic citations requires moving beyond the single-sentence single-label setting
Anne Lauscher, Brandon Ko, Bailey Kuehl, Sophie Johnson, Arman Cohan, David Jurgens and Kyle Lo ......................................................... 1875
DEGREE: A Data-Efficient Generation-Based Event Extraction Model
I-Hung Hsu, Kuan-Hao Huang, Elizabeth Boschee, Scott Miller, Prem Natarajan, Kai-Wei Chang and Nanyun Peng .......................................................... 1890

Bridging the Gap between Language Models and Cross-Lingual Sequence Labeling
Nuo Chen, Linjun Shou, Ming Gong, Jian Pei and Daxin Jiang ......................... 1909

Hero-Gang Neural Model For Named Entity Recognition
Jinpeng Hu, Yaling Shen, Yang Liu, Xiang Wan and Tsung-Hui Chang ................ 1924

MGIMN: Multi-Grained Interactive Matching Network for Few-shot Text Classification
Jianhai Zhang, Mieradiliijing Maimaiti, Gao Xing, Yuanhang Zheng and Ji Zhang ..... 1937

All You May Need for VQA are Image Captions
Soravit Changpinyo, Doron Kukliansy, Idan Szpektor, Xi Chen, Nan Ding and Radu Soricut1947

Frustratingly Easy System Combination for Grammatical Error Correction
Muhammad Reza Qorib, Seung-Hoon Na and Hwee Tou Ng .............................. 1964

Simple Local Attentions Remain Competitive for Long-Context Tasks
Wenhan Xiong, Barlas Oguz, Anicht Gupta, Xiilun Chen, Diana Liskovich, Omer Levy, Scott Yih and Yashar Mehdad ................................. 1975

Even the Simplest Baseline Needs Careful Re-investigation: A Case Study on XML-CNN
Si-An Chen, Jie-jyun Liu, Tsung-Han Yang, Hsuan-Tien Lin and Chih-Jen Lin ........ 1987

Multi-Relational Graph Transformer for Automatic Short Answer Grading
Rajat Agarwal, Varun Khurana, Karish Grover, Mukesh Mohania and Vikram Goyal. 2001

Event Schema Induction with Double Graph Autoencoders
Xiaomeng Jin, Manling Li and Heng Ji .............................................................. 2013

CS1QA: A Dataset for Assisting Code-based Question Answering in an Introductory Programming Course
Changyoon Lee, Yeon Seonwoo and Alice Oh .............................. 2026

Unsupervised Cross-Lingual Transfer of Structured Predictors without Source Data
Kemal Kurniawan, Lea Frermann, Philip Schulz and Trevor Cohn ........................ 2041

Don’t Take It Literally: An Edit-Invariant Sequence Loss for Text Generation
Guangyi Liu, Zichao Yang, Tianhua Tao, Xiaodan Liang, Junwei Bao, Zhen Li, Xiaodong He, Shuguang Cui and Zhitian Hu .................................................. 2055

Modeling Exemplification in Long-form Question Answering via Retrieval
Shufan Wang, Fangyuan Xu, Lauren Thompson, Eunsol Choi and Mohit Iyyer ........ 2079

D2U: Distance-to-Uniform Learning for Out-of-Scope Detection
Eyup Halit Yilmaz and Cagri Toraman .......................................................... 2093

Reference-free Summarization Evaluation via Semantic Correlation and Compression Ratio
Yizhu Liu, Qi Jia and Kenny Q. Zhu ................................................................. 2109

KroneckerBERT: Significant Compression of Pre-trained Language Models Through Kronecker Decomposition and Knowledge Distillation
Marzieh S. Tahaei, Ella Charlaix, Vahid Partovi Nia, Ali Ghodsi and Mehdi Rezagholizadeh2116
Building a Role Specified Open-Domain Dialogue System Leveraging Large-Scale Language Models
Sanghwan Bae, Donghyun Kwak, Sungdong Kim, Donghoon Ham, Soyoung Kang, Sang-Woo Lee and Woomyoung Park ................................................................. 2128

Sentence-Level Resampling for Named Entity Recognition
Xiaochen Wang and Yue Wang ................................................................. 2151

Word Tour: One-dimensional Word Embeddings via the Traveling Salesman Problem
Ryoma Sato .............................................................................................. 2166

On the Diversity and Limits of Human Explanations
Chenhao Tan ............................................................................................. 2173

Locally Aggregated Feature Attribution on Natural Language Model Understanding
Sheng Zhang, Jin Wang, Haitao Jiang and Rui Song .................................... 2189

Generic and Trend-aware Curriculum Learning for Relation Extraction
Nidhi Vakil and Hadi Amiri ....................................................................... 2202

On Systematic Style Differences between Unsupervised and Supervised MT and an Application for High-Resource Machine Translation
Kelly Marchisio, Markus Freitag and David Grangier ................................ 2214

Evidentiality-guided Generation for Knowledge-Intensive NLP Tasks
Akari Asai, Matt Gardner and Hannaneh Hajishirzi .................................... 2226

Modularized Transfer Learning with Multiple Knowledge Graphs for Zero-shot Commonsense Reasoning
Yu Jin Kim, Beong-woo Kwak, Youngwook Kim, Reinald Kim Amplayo, Seung-won Hwang and Jinyoung Yeo ............................................................. 2244

Learning to Express in Knowledge-Grounded Conversation
Xueliang Zhao, Tingchen Fu, Chongyang Tao, Wei Wu, Dongyan Zhao and Rui Yan .... 2258

End-to-End Chinese Speaker Identification
Dian Yu, Ben Zhou and Dong Yu .................................................................. 2274

MINION: a Large-Scale and Diverse Dataset for Multilingual Event Detection
Amir Pouran Ben Veyseh, Minh Van Nguyen, Franck Demoncort and Thien Huu Nguyen. 2286

Do Prompt-Based Models Really Understand the Meaning of Their Prompts?
Albert Webson and Ellie Pavlick ................................................................ 2300

GPL: Generative Pseudo Labeling for Unsupervised Domain Adaptation of Dense Retrieval
Kexin Wang, Nandan Thakur, Nils Reimers and Iryna Gurevych ................. 2345

Sparse Distillation: Speeding Up Text Classification by Using Bigger Student Models
Qinyuan Ye, Madian Khabsa, Mike Lewis, Sinong Wang, Xiang Ren and Aaron Jaech .... 2361

Towards Understanding Large-Scale Discourse Structures in Pre-Trained and Fine-Tuned Language Models
Patrick Huber and Giuseppe Carenini ...................................................... 2376

SAIS: Supervising and Augmenting Intermediate Steps for Document-Level Relation Extraction
Yuxin Xiao, Zecheng Zhang, Yuning Mao, Carl Yang and Jiawei Han ............. 2395
LITE: Intent-based Task Representation Learning Using Weak Supervision
Naoki Otani, Michael Gamon, Sujay Kumar Jauhar, Mei Yang, Sri Raghu Malireddi and Oriana Riva ................................................................. 2410

Does Summary Evaluation Survive Translation to Other Languages?
Spencer Braun, Oleg Vasilyev, Neslihan Iskender and John Bohannon .................. 2425

A Shoulder to Cry on: Towards a Motivational Virtual Assistant for Assuaging Mental Agony
Tulika Saha, Saichethan Miriyala Reddy, Anindya Sundar Das, Sriparna Saha and Pushpak Bhattacharyya ............................................................. 2436

Forrest Sheng Bao, Ge Luo, Hebi Li, Minghui Qiu, Yinfei Yang, Youbiao He and Cen Chen 2450

Combating the Curse of Multilinguality in Cross-Lingual WSD by Aligning Sparse Contextualized Word Representations
Gábor Berend ................................................................. 2459

Cheat Codes to Quantify Missing Source Information in Neural Machine Translation
Proyag Pal and Kenneth Heafield ........................................... 2472

WiC = TSV = WSD: On the Equivalence of Three Semantic Tasks
Bradley Hauer and Grzegorz Kondrak ........................................... 2478

What do tokens know about their characters and how do they know it?
Ayush Kaushal and Kyle Mahowald ........................................... 2487

AnswerSumm: A Manually-Curated Dataset and Pipeline for Answer Summarization
Alexander Fabbri, Xiaojian Wu, Srimati Iyer, Haoran Li and Mona T. Diab ............. 2508

Paragraph-based Transformer Pre-training for Multi-Sentence Inference
Luca Di Liello, Siddhant Garg, Luca Soldaini and Alessandro Moschitti ................. 2521

Text Style Transfer via Optimal Transport
Nasim Nouri ................................................................. 2532

Exploring the Role of Task Transferability in Large-Scale Multi-Task Learning
Vishakh Padmakumar, Leonard Lausen, Miguel Ballesteros, Sheng Zha, He He and George Karypis ................................................................. 2542

Interactive Query-Assisted Summarization via Deep Reinforcement Learning
Ori Shapira, Ramakanth Pasunuru, Mohit Bansal, Ido Dagan and Yael Amsterdamer ... 2551

Data Augmentation with Dual Training for Offensive Span Detection
Nasim Nouri ................................................................. 2569

Training Mixed-Domain Translation Models via Federated Learning
Peyman Passban, Tanya Roosta, Rahul Gupta, Ankit Chadha and Clement Chung ........ 2576

QAFactEval: Improved QA-Based Factual Consistency Evaluation for Summarization
Alexander Fabbri, Chien-Sheng Wu, Wenhao Liu and Caiming Xiong ..................... 2587

How Gender Debiasing Affects Internal Model Representations, and Why It Matters
Hadas Orgad, Seraphina Goldfarb-Tarrant and Yonatan Belinkov ........................ 2602

A Structured Span Selector
Tianyu Liu, Yuchen Eleanor Jiang, Ryan Cotterell and Mrinmaya Sachan ................ 2629
Unified Semantic Typing with Meaningful Label Inference
James Y. Huang, Bangzheng Li, Jiashu Xu and Muhao Chen ................................. 2642

Learning To Retrieve Prompts for In-Context Learning
Ohad Rubin, Jonathan Herzig and Jonathan Berant ............................... 2655

Necessity and Sufficiency for Explaining Text Classifiers: A Case Study in Hate Speech Detection
Esma Balkir, Isar Nejadgholi, Kathleen C. Fraser and Svetlana Kiritchenko .......... 2672

Learning to Retrieve Passages without Supervision
Ori Ram, Gal Shachaf, Omer Levy, Jonathan Berant and Amir Globerson ................ 2687

Re2G: Retrieve, Rerank, Generate
Michael Glass, Gaetano Rossiello, Md Faisal Mahbub Chowdhury, Ankita Naik, Pengshen Cai and Alfio Gliozzo .................................................. 2701

Don’t sweat the small stuff, classify the rest: Sample Shielding to protect text classifiers against adversarial attacks
Jonathan Rusert and Padmini Srinivasan ..................................................... 2716

Federated Learning with Noisy User Feedback
Rahul Sharma, Anil Ramakrishna, Ansel MacLaughlin, Anna Rumshisky, Jimit Majmudar, Clement Chung, Salman Avestimehr and Rahul Gupta .................. 2726

Gender Bias in Masked Language Models for Multiple Languages
Masahiro Kaneko, Aizhan Imankulova, Danushka Bollegala and Naoaki Okazaki ........ 2740

Multi-Domain Targeted Sentiment Analysis
Orith Toledo-Ronen, Matan Orbach, Yoav Katz and Noam Slonim ..................... 2751

Falsesum: Generating Document-level NLI Examples for Recognizing Factual Inconsistency in Summarization
Prasetya Ajie Utama, Joshua Bambrick, Nafise Sadat Moosavi and Iryna Gurevych ...... 2763

Dynamic Gazetteer Integration in Multilingual Models for Cross-Lingual and Cross-Domain Named Entity Recognition
Besnik Fetahu, Anjie Fang, Oleg Rokhlenko and Shervin Malmasi .................... 2777

MetaICL: Learning to Learn In Context
Sewon Min, Mike Lewis, Luke Zettlemoyer and Hannaneh Hajishirzi .................. 2791

Enhancing Knowledge Selection for Grounded Dialogues via Document Semantic Graphs
Sha Li, Mahdi Namazifar, Di Jin, Mohit Bansal, Heng Ji, Yang Liu and Dilek Hakkani-Tur. 2810

Using Natural Sentence Prompts for Understanding Biases in Language Models
Sarah Alnegheimish, Alicia Guo and Yi Sun ................................................. 2824

Robust Conversational Agents against Imperceptible Toxicity Triggers
Ninareh Mehrabi, Ahmad Beirami, Fred Morstatter and Aram Galstyan ............... 2831

Selective Differential Privacy for Language Modeling
Weiyan Shi, Aiqi Cui, Evan Li, Ruoxi Jia and Zhou Yu .................................... 2848

Do Trajectories Encode Verb Meaning?
Dylan Ebert, Chen Sun and Ellie Pavlick ...................................................... 2860

Long Context Question Answering via Supervised Contrastive Learning
Avi Caciularu, Ido Dagan, Jacob Goldberger and Arman Cohan ................................ 2872
The USMLE® Step 2 Clinical Skills Patient Note Corpus
Victoria Yaneva, Janet Mee, Le An Ha, Polina Harik, Michael Jodoin and Alex J Mechaber 2880

Learning to Borrow--Relation Representation for Without-Mention Entity-Pairs for Knowledge Graph Completion
Huda Hakami, Mona Hakami, Angrosh Mandya and Danushka Bollegala 2887

Improving Entity Disambiguation by Reasoning over a Knowledge Base
Tom Ayoola, Joseph Fisher and Andrea Pierleoni 2899

Modal Dependency Parsing via Language Model Priming
Jiarui Yao, Nianwen Xue and Bonan Min 2913

Document-Level Relation Extraction with Sentences Importance Estimation and Focusing
Wang Xu, Kehai Chen, Lili Mou and Tiejun Zhao 2920

Are All the Datasets in Benchmark Necessary? A Pilot Study of Dataset Evaluation for Text Classification
Yang Xiao, Jinlan Fu, See-Kiong Ng and Pengfei Liu 2930

Triggerless Backdoor Attack for NLP Tasks with Clean Labels
Leilei Gan, Jiwei Li, Tianwei Zhang, Xiaoya Li, Yuxian Meng, Fei Wu, Yi Yang, Shangwei Guo and Chun Fan 2942

PPL-MCTS: Constrained Textual Generation Through Discriminator-Guided MCTS Decoding
Antoine Chaffin, Vincent Claveau and Ewa Kijak 2953

Interpretable Proof Generation via Iterative Backward Reasoning
Hanbao Qu, Yu Cao, Jun Gao, Liang Ding and Ruifeng Xu 2968

Domain Confused Contrastive Learning for Unsupervised Domain Adaptation
Quanyu Long, Tianze Luo, Wenya Wang and Sinno Pan 2982

Incorporating Centering Theory into Neural Coreference Resolution
Haixia Chai and Michael Strube 2996

Progressive Class Semantic Matching for Semi-supervised Text Classification
Haiming Xu, Lingqiao Liu and Ehsan M Abbasnejad 3003

Low Resource Style Transfer via Domain Adaptive Meta Learning
Xiangyang Li, Xiang Long, Yu Xia and Sujian Li 3014

Features or Spurious Artifacts? Data-centric Baselines for Fair and Robust Hate Speech Detection
Alan Ramponi and Sara Tonelli 3027

Document-Level Event Argument Extraction by Leveraging Redundant Information and Closed Boundary Loss
Hanzhang Zhou and Kezhi Mao 3041

A Few Thousand Translations Go a Long Way! Leveraging Pre-trained Models for African News Translation
David Ifeoluwa Adelani, Jesujoba Oluwadara Alabi, Angela Fan, Julia Kreutzer, Xiaoyu Shen, Machel Reid, Dana Ruiter, Dietrich Klakow, Peter Nabende, Ernie Chang, Tajuddeen Gwadabe, Freshia Sackey, Bonaventure F. P. Dossou, Chris Chinene Emeuze, Colin Leong, Michael Beukman, Shamsudeen Hassan Muhammad, Guyo Dub Jarso, Oreen Yousuf, Andre Niyongabo Rubungo, Gilles Hacheme, Eric Peter Wairagala, Muhammad Umair Nasir, Benjamin Ayoade Ajibade, Tunde Oluwaseye Ajayi, Yvonne Wambui Gitau, Jade Abbott, Mohamed Ahmed, Millicent Ochieng, Anuoluwapo
Aremu, Perez Ogayo, Jonathan Mukiibi, Fatoumata Ouoba Kabore, Godson Koffi Kalipe, Derguene Mbaye, Allahsera Auguste Tapo, Victoire Memdjokam Koagne, Edwin Munkoh-Buabeng, Valencia Wagner, Idris Abdulmumin, Ayodele Awokoya, Happy Buzaba, Blessing Sibanda, Andiswa Bukula and Sam Manthalu. 3053

Should We Rely on Entity Mentions for Relation Extraction? Debiasing Relation Extraction with Counterfactual Analysis
Yiwei Wang, Muhao Chen, Wenzuan Zhou, Yujun Cai, Xuxuan Liang, Dayiheng Liu, Baosong Yang, Juncheng Liu and Bryan Hoo. 3071

Analyzing Encoded Concepts in Transformer Language Models
Hassan Sajjad, Nadir Durrani, Fahim Dalvi, Firoj Alam, Abdul Rafae Khan and Jia Xu. 3082

Boosted Dense Retriever
Patrick Lewis, Barlas Oguz, Wenhan Xiong, Fabio Petroni, Scott Yih and Sebastian Riedel. 3102

MuCGEC: a Multi-Reference Multi-Source Evaluation Dataset for Chinese Grammatical Error Correction
Yue Zhang, Zhenghua Li, Zuyi Bao, Jiacheng Li, Bo Zhang, Chen Li, Fei Huang and Min Zhang. 3118

NeuS: Neutral Multi-News Summarization for Mitigating Framing Bias
Nayeon Lee, Yejin Bang, Tiezheng YU, Andrea Madotto and Pascale Fung. 3131

Enhance Incomplete Utterance Restoration by Joint Learning Token Extraction and Text Generation
Shumpei Inoue, Tsungwei Liu, Son Hong Nguyen and Minh-Tien Nguyen. 3149

Efficient Constituency Tree based Encoding for Natural Language to Bash Translation
Shikhar Bharadwaj and Shirish Shevade. 3159

Privacy-Preserving Text Classification on BERT Embeddings with Homomorphic Encryption
Garam Lee, Minsoo Kim, Jai Hyun Park, Seung-won Hwang and Jung Hee Cheon. 3169

ITA: Image-Text Alignments for Multi-Modal Named Entity Recognition
Xinyu Wang, Min Gui, Yong Jiang, Zixia Jia, Nguyen Bach, Tao Wang, Zhongqiang Huang and Kewei Tu. 3176

A Dataset for N-ary Relation Extraction of Drug Combinations
Aryeh Tiktinsky, Vijay Viswanathan, Danna Niezni, Dana Meron Azagury, Yosi Shamay, Hillel Taub-Tabib, Tom Hope and Yoav Goldberg. 3190

Curriculum: A Broad-Coverage Benchmark for Linguistic Phenomena in Natural Language Understanding
Zeming Chen and Qiyue Gao. 3204

Neural Language Taskonomy: Which NLP Tasks are the most Predictive of fMRI Brain Activity?
Subba Reddy Oota, Jashn Arora, Veeral Agarwal, Mounika Marreddy, Manish Gupta and Bapi Raju Surampudi. 3220

FactGraph: Evaluating Factuality in Summarization with Semantic Graph Representations
Leonardo F. R. Ribeiro, Mengwen Liu, Iryna Gurevych, Markus Dreyer and Mohit Bansal. 3238

Unsupervised Paraphrasability Prediction for Compound Nominalizations
John Sie Yuen Lee, Ho Hung Lim and Carol Carol Webster. 3254

Global Entity Disambiguation with BERT
Ikuya Yamada, Koki Washio, Hiroyuki Shindo and Yuji Matsumoto. 3264
Clues Before Answers: Generation-Enhanced Multiple-Choice QA
Zixian Huang, Ao Wu, Jiaying Zhou, Yu Gu, Yue Zhao and Gong Cheng .......................... 3272

Towards Efficient NLP: A Standard Evaluation and A Strong Baseline
Xiangyang Liu, Tianxiang Sun, Junliang He, Jiawen Wu, Lingling Wu, Xinyu Zhang, Hao Jiang, Zhao Cao, Xuanjing Huang and Xipeng Qiu .................................................. 3288

Stylized Knowledge-Grounded Dialogue Generation via Disentangled Template Rewriting
Qingfeng Sun, Can Xu, Huang Hu, Yujing Wang, Jian Miao, Xiubo Geng, Yining Chen, Fei Xu and Daxin Jiang ................................................................. 3304

LUNA: Learning Slot-Turn Alignment for Dialogue State Tracking
Yifan Wang, Jing Zhao, Junwei Bao, Chaoqun Duan, Youzheng Wu and Xiaodong He .... 3319

Crossroads, Buildings and Neighborhoods: A Dataset for Fine-grained Location Recognition
Pei Chen, Haotian Xu, Cheng Zhang and Ruihong Huang ........................................... 3329

Tricks for Training Sparse Translation Models
Dheeru Dua, Shruti Bhosale, Vedanuj Goswami, James Cross, Mike Lewis and Angela Fan 3340

Persona-Guided Planning for Controlling the Protagonist’s Persona in Story Generation
Zhexin Zhang, Jiaxin Wen, Jian Guan and Minlie Huang ........................................... 3346

CHEF: A Pilot Chinese Dataset for Evidence-Based Fact-Checking
Xuming Hu, Zhijiang Guo, GuanYu Wu, Aiwei Liu, Lijie Wen and Philip S. Yu ............ 3362

VGNMN: Video-grounded Neural Module Networks for Video-Grounded Dialogue Systems
Hung Le, Nancy F. Chen and Steven Hoi ................................................................. 3377

Multimodal Dialogue State Tracking
Hung Le, Nancy F. Chen and Steven Hoi ................................................................. 3394

On the Use of Bert for Automated Essay Scoring: Joint Learning of Multi-Scale Essay Representation
Yongjie Wang, Chuang Wang, Ruobing Li and Hui Lin ............................................ 3416

Recognition of They/Them as Singular Personal Pronouns in Coreference Resolution
Connor Baumler and Rachel Rudinger ................................................................. 3426

TWEETSPIN: Fine-grained Propaganda Detection in Social Media Using Multi-View Representations
Prashanth Vijayaraghavan and Soroush Vosoughi ................................................... 3433

UserIdentifier: Implicit User Representations for Simple and Effective Personalized Sentiment Analysis
Fatemehsadat Mireshghallah, Vaishnavi Shrivastava, Milad Shokouhi, Taylor Berg-Kirkpatrick, Robert Sim and Dimitrios Dimitriadis ......................................................... 3449

Improving Neural Models for Radiology Report Retrieval with Lexicon-based Automated Annotation
Luyao Shi, Tanveer Syeda-mahmood and Tyler Baldwin ....................................... 3457

Transparent Human Evaluation for Image Captioning
Jungo Kasai, Keisuke Sakaguchi, Lavinia Dunagan, Jacob Daniel Morrison, Ronan Le Bras, Yejin Choi and Noah Smith ......................................................... 3464

Lifting the Curse of Multilinguality by Pre-training Modular Transformers
Jonas Pfeiffer, Naman Goyal, Xi Victoria Lin, Xian Li, James Cross, Sebastian Riedel and Mikel Artetxe ................................................................. 3479
Quantifying Language Variation Acoustically with Few Resources
Martijn Bartelds and Martijn Wieling ................................................. 3735

Adaptable Adapters
Nafise Sadat Moosavi, Quentin Delfosse, Kristian Kersting and Iryna Gurevych ........ 3742

Models in the Loop: Aiding Crowdworkers with Generative Annotation Assistants
Max Bartolo, Tristan Thrush, Sebastian Riedel, Pontus Stenetorp, Robin Jia and Douwe Kiela 3754

GMN: Generative Multi-modal Network for Practical Document Information Extraction
Haoyu Cao, Jiefeng Ma, Antai Guo, Yiqing Hu, Hao Liu, Deqiang Jiang, Yinsong Liu and Bo Ren ................................................................. 3768

One Reference Is Not Enough: Diverse Distillation with Reference Selection for Non-Autoregressive Translation
Chenze Shao, Xuanfu Wu and Yang Feng .............................................. 3779

Can Rationalization Improve Robustness?
Howard Chen, Jacqueline He, Karthik R Narasimhan and Danqi Chen ....................... 3792

On the Effectiveness of Sentence Encoding for Intent Detection Meta-Learning
Tingting Ma, Qianhui Wu, Zhiwei Yu, Tiejun Zhao and Chin-Yew Lin ......................... 3806

A Computational Acquisition Model for Multimodal Word Categorization
Uri Berger, Gabriel Stanovsky, Omri Abend and Lea Frermann .............................. 3819

Residue-Based Natural Language Adversarial Attack Detection
Vyas Raina and Mark Gales ..................................................................... 3836

Does it Really Generalize Well on Unseen Data? Systematic Evaluation of Relational Triple Extraction Methods
Juhyuk Lee, Min-Joong Lee, June Yong Yang and Eunho Yang ................................. 3849

From spoken dialogue to formal summary: An utterance rewriting for dialogue summarization
Yue Fang, Hainan Zhang, Hongshen Chen, Zhuoye Ding, Bo Long, Yanyan Lan and Yanquan Zhou ........................................................................ 3859

EASE: Entity-Aware Contrastive Learning of Sentence Embedding
Sosuke Nishikawa, Ryokan Ri, Ikuya Yamada, Yoshimasa Tsuruoka and Isao Echizen .... 3870

Is Neural Topic Modelling Better than Clustering? An Empirical Study on Clustering with Contextual Embeddings for Topics
Zihan Zhang, Meng Fang, Ling Chen and Mohammad Reza Namazi Rad .................... 3886

Dynamic Multistep Reasoning based on Video Scene Graph for Video Question Answering
Jianguo Mao, Wenbin Jiang, Xiangdong Wang, Zhifan Feng, Yajuan Lyu, Hong Liu and Yong Zhu ............................................................................ 3894

TRUE: Re-evaluating Factual Consistency Evaluation
Or Honovich, Roee Aharoni, Jonathan Herzig, Hagai Taitelbaum, Doron Kukliansy, Vered Cohen, Thomas Scialom, Idan Szpektor, Avinatan Hassidim and Yossi Matias .......................... 3905

Knowledge Inheritance for Pre-trained Language Models
Yujia Qin, Yankai Lin, Jing Yi, Jiajie Zhang, Xu Han, Zhengyan Zhang, Yusheng Su, Zhiyuan Liu, Peng Li, Maosong Sun and Jie Zhou .............................................. 3921
COGMEN: COntextualized GNN based Multimodal Emotion recognition
Abhinav Joshi, Ashwani Bhat, Ayush Jain, Atin Vikram Singh and Ashutosh Modi ........ 4148

Revisit Overconfidence for OOD Detection: Reassigned Contrastive Learning with Adaptive Class-dependent Threshold
Yanan Wu, Keqing He, Yuanmeng Yan, QiXiang Gao, Zhiyuan Zeng, Fujia Zheng, Lulu Zhao, Huixing Jiang, Wei Wu and Weiran Xu ............................................................. 4165

AISFG: Abundant Information Slot Filling Generator
Yang Yan, Junda Ye, Zhongbao Zhang and Liwen Wang ........................................ 4180

Improving negation detection with negation-focused pre-training
Thinh Hung Truong, Timothy Baldwin, Trevor Cohn and Karin Verspoor ..................... 4188

Practice Makes a Solver Perfect: Data Augmentation for Math Word Problem Solvers
Vivek Kumar, Rishabh Maheshwary and Vikram Pudi ............................................. 4194

DiffCSE: Difference-based Contrastive Learning for Sentence Embeddings
Yung-Sung Chuang, Rumen Dangovski, Hongyin Luo, Yang Zhang, Shiyu Chang, Marin Soljacic, Shang-Wen Li, Scott Yih, Yoon Kim and James R. Glass ............................................ 4207

Generative Cross-Domain Data Augmentation for Aspect and Opinion Co-Extraction
Junjie Li, Jianfei Yu and Rui Xia ........................................................................... 4219

ProQA: Structural Prompt-based Pre-training for Unified Question Answering
Wanjun Zhong, Yifan Gao, Ning Ding, Yujia Qin, Zhiyuan Liu, Ming Zhou, Jiahai Wang, Jian Yin and Nan Duan ................................................................. 4230

A Data Cartography based MixUp for Pre-trained Language Models
Seo Yeon Park and Cornelia Caragea ............................................................... 4244

Grapheme-to-Phoneme Conversion for Thai using Neural Regression Models
Tomohiro Yamasaki .......................................................................................... 4251

Generating Authentic Adversarial Examples beyond Meaning-preserving with Doubly Round-trip Translation
Siyu Lai, Zhen Yang, Fandong Meng, Xue Zhang, Yufeng Chen, Jinan Xu and Jie Zhou ... 4256

TVShowGuess: Character Comprehension in Stories as Speaker Guessing
Yisi Sang, Xiangyang Mou, Mo Yu, Shunyu Yao, Jing Li and Jeffrey Stanton ............ 4267

Causal Distillation for Language Models
Zhengxuan Wu, Atticus Geiger, Joshua Rozner, Elisa Kreiss, Hanson Lu, Thomas Icard, Christopher Potts and Noah Goodman ..................................................... 4288

FNet: Mixing Tokens with Fourier Transforms
James Lee-Thorp, Joshua Ainslie, Ilya Eckstein and Santiago Ontanon ..................... 4296

Answer Consolidation: Formulation and Benchmarking
Wenxuan Zhou, Qiang Ning, Heba Elfardy, Kevin Small and Muhao Chen ................ 4314

Informativeness and Invariance: Two Perspectives on Spurious Correlations in Natural Language
Jacob Eisenstein ............................................................................................... 4326

FOAM: A Follower-aware Speaker Model For Vision-and-Language Navigation
Zi-Yi Dou and Nanyun Peng .............................................................................. 4332
Improving Compositional Generalization with Latent Structure and Data Augmentation
Linlu Qiu, Peter Shaw, Panupong Pasupat, Pawel Krzysztof Nowak, Tal Linzen, Fei Sha and Kristina Toutanova .......................................................... 4341

Joint Extraction of Entities, Relations, and Events via Modeling Inter-Instance and Inter-Label Dependencies
Minh Van Nguyen, Bonan Min, Franck Dernoncourt and Thien Huu Nguyen ............. 4363

Linguistic Frameworks Go Toe-to-Toe at Neuro-Symbolic Language Modeling
Jakob Prange, Nathan Schneider and Lingpeng Kong ........................................... 4375

Imagination-Augmented Natural Language Understanding
Yujie Lu, Wanrong Zhu, Xin Eric Wang, Miguel Eckstein and William Yang Wang ...... 4392

What company do words keep? Revisiting the distributional semantics of J.R. Firth & Zellig Harris
Mikael Brunila and Jack LaViolette ........................................................................ 4403

Compositional Task-Oriented Parsing as Abstractive Question Answering
Wenting Zhao, Konstantine Arkoudas, Weiqi Sun and Claire Cardie .......................... 4418

Learning Cross-Lingual IR from an English Retriever
Yulong Li, Martin Franz, Md Arafat Sultan, Bhavani Iyer, Young-Suk Lee and Avirup Sil . . 4428

Testing the Ability of Language Models to Interpret Figurative Language
Emmy Liu, Chenxuan Cui, Kenneth Zheng and Graham Neubig ................................ 4437

Multi-Vector Models with Textual Guidance for Fine-Grained Scientific Document Similarity
Sheshera Mysore, Arman Cohan and Tom Hope .................................................... 4453

CHAI: A CHatbot AI for Task-Oriented Dialogue with Offline Reinforcement Learning
Siddharth Verma, Justin Fu, Sherry Yang and Sergey Levine .................................... 4471

Connecting the Dots between Audio and Text without Parallel Data through Visual Knowledge Transfer
Yanpeng Zhao, Jack Hessel, Youngjae Yu, Ximing Lu, Rowan Zellers and Yejin Choi .... 4492

SURF: Semantic-level Unsupervised Reward Function for Machine Translation
Atijit Anuchitanukul and Julia Ive ........................................................................... 4508

Disentangling Categorization in Multi-agent Emergent Communication
Washington Garcia, Hamilton Scott Clouse and Kevin R. B. Butler ......................... 4523

Show, Don’t Tell: Demonstrations Outperform Descriptions for Schema-Guided Task-Oriented Dialogue
Raghav Gupta, Harrison Lee, Jeffrey Zhao, Yuan Cao, Abhinav Rastogi and Yonghui Wu . 4541

Does Pre-training Induce Systematic Inference? How Masked Language Models Acquire Commonsense Knowledge
Ian Porada, Alessandro Sordoni and Jackie CK Cheung ........................................... 4550

Using Paraphrases to Study Properties of Contextual Embeddings
Laura Burdick, Jonathan K Kummerfeld and Rada Mihalcea .................................... 4558

Measure and Improve Robustness in NLP Models: A Survey
Xuezhi Wang, Haohan Wang and Diyi Yang ........................................................... 4569

Learning to Generate Examples for Semantic Processing Tasks
Danilo Croce, Simone Filice, Giuseppe Castellucci and Roberto Basili ......................... 4587

Ivii
Symbolic Knowledge Distillation: from General Language Models to Commonsense Models
Peter West, Chandra Bhagavatula, Jack Hessel, Jena D. Hwang, Liwei Jiang, Ronan Le Bras, Ximing Lu, Sean Welleck and Yejin Choi ................................................................. 4602

GenIE: Generative Information Extraction
Martin Josifoski, Nicola De Cao, Maxime Peyrard, Fabio Petroni and Robert West ........ 4626

Entity Linking via Explicit Mention-Mention Coreference Modeling
Dhruv Agarwal, Rico Angell, Nicholas Monath and Andrew McCallum ...................... 4644

Massive-scale Decoding for Text Generation using Lattices
Jiacheng Xu, Siddhartha Jonnalagadda and Greg Durrett ...................................... 4659

Disentangling Indirect Answers to Yes-No Questions in Real Conversations
Krishna Chaitanya Sanagavarapu, Jathin Pranav Singaraju, Anusha Kakeleti, Anirudh Kaza, Aaron Abraham Mathews, Helen Li, Nathan Raul Brito and Eduardo Blanco ......................... 4677

Quantifying Adaptability in Pre-trained Language Models with 500 Tasks
Belinda Z. Li, Jane A. Yu, Madian Khabسا, Luke Zettlemoyer, Alon Y. Halevy and Jacob Andreas 4696

Counterfactually Augmented Data and Unintended Bias: The Case of Sexism and Hate Speech Detection
Indira Sen, Mattia Samory, Claudia Wagner and Isabelle Augenstein ...................... 4716

A Study of the Attention Abnormality in Trojaned BERTs
Weimin Lyu, Songzhu Zheng, Tengfei Ma and Chao Chen ...................................... 4727

EPiDA: An Easy Plug-in Data Augmentation Framework for High Performance Text Classification
Minyi Zhao, Lu Zhang, Yi Xu, Jiandong Ding, Jihong Guan and Shuigeng Zhou ........... 4742

Partial-input baselines show that NLI models can ignore context, but they don’t.
Neha Srikanth and Rachel Rudinger ................................................................. 4753

Lifelong Pretraining: Continually Adapting Language Models to Emerging Corpora
Xisen Jin, Dejiao Zhang, Henghui Zhu, Wei Xiao, Shang-Wen Li, Xiaokai Wei, Andrew Arnold and Xiang Ren ................................................................. 4764

Learning as Conversation: Dialogue Systems Reinforced for Information Acquisition
Pengshan Cai, Hui Wan, Fei Liu, Mo Yu, Hong yu and Sachindra Joshi ..................... 4781

Dynamic Programming in Rank Space: Scaling Structured Inference with Low-Rank HMMs and PC-FGs
Songlin Yang, Wei Liu and Kewei Tu ................................................................. 4797

Terne Sasha Thorn Jakobsen and Anna Rogers .................................................... 4810

Reducing Disambiguation Biases in NMT by Leveraging Explicit Word Sense Information
Niccolò Campolungo, Tommaso Pasini, Denis Emelin and Roberto Navigli ........... 4824

Mining Clues from Incomplete Utterance: A Query-enhanced Network for Incomplete Utterance Rewriting
Shuzheng Si, Shuang Zeng and Baobao Chang .................................................... 4839
A Corpus for Understanding and Generating Moral Stories
Jian Guan, Ziqi Liu and Minlie Huang .......................................................... 5069

Modeling Multi-Granularity Hierarchical Features for Relation Extraction
Xinnian Liang, Shuangzhi Wu, Mu Li and Zhoujun Li ........................................ 5088

Cross-modal Contrastive Learning for Speech Translation
Rong Ye, Mingxuan Wang and Lei Li ................................................................. 5099

Meet Your Favorite Character: Open-domain Chatbot Mimicking Fictional Characters with only a Few Utterances
Seungju Han, Beomsu Kim, Jin Yong Yoo, Seokjun Seo, Sangbum Kim, Enkhbayar Erdenee and Buru Chang .............................................................. 5114

DynamicTOC: Persona-based Table of Contents for Consumption of Long Documents
Himanshu Maheshwari, Nethraa Sivakumar, Shelly Jain, Tanvi Karandikar, Vinay Aggarwal, Navita Goyal and Sumit Shekhar .................................................. 5133

KALA: Knowledge-Augmented Language Model Adaptation
Minki Kang, Jinheon Baek and Sung Ju Hwang .................................................. 5144

On the Effect of Pretraining Corpora on In-context Learning by a Large-scale Language Model
Seongjin Shin, Sang-Woo Lee, Hwijeen Ahn, Sungdong Kim, HyounSeok Kim, Boseop Kim, Kyunghyun Cho, Gichang Lee, Woomyoung Park, Jung-Woo Ha and Nako Sung ............................. 5168

Sketching as a Tool for Understanding and Accelerating Self-attention for Long Sequences
Yifan Chen, Qi Zeng, Dilek Hakkani-Tur, Di Jin, Heng Ji and Yun Yang ................ 5187

Partner Personas Generation for Dialogue Response Generation
Hongyuan Lu, Wai Lam, Hong Cheng and Helen M. Meng ............................... 5200

Semantically Informed Slang Interpretation
Zhewei Sun, Richard Zemel and Yang Xu ........................................................ 5213

Dual-Channel Evidence Fusion for Fact Verification over Texts and Tables
Nan Hu, Zirui Wu, Yuxuan Lai, Xiao Liu and Yansong Feng .............................. 5232

TreeMix: Compositional Constituency-based Data Augmentation for Natural Language Understanding
Le Zhang, Zichao Yang and Diyi Yang ............................................................ 5243

Syn2Vec: Synset Colexification Graphs for Lexical Semantic Similarity
John Harvill, Roxana Girju and Mark A. Hasegawa-Johnson ............................ 5259

On the Origin of Hallucinations in Conversational Models: Is it the Datasets or the Models?
Nouha Dziri, Sivan Milton, Mo Yu, Osmar Zaiane and Siva Reddy ..................... 5271

Is My Favorite New Movie My Favorite Movie? Probing the Understanding of Recursive Noun Phrases
Qing Lyu, Zheng Hua, Daoxin Li, Li Zhang, Marianna Apidianaki and Chris Callison-Burch 5286

Original or Translated? A Causal Analysis of the Impact of Translationese on Machine Translation Performance
Jingwei Ni, Zhijing Jin, Markus Freitag, Mrinmaya Sachan and Bernhard Schölkopf ................ 5303

Visual Commonsense in Pretrained Unimodal and Multimodal Models
Chenyu Zhang, Benjamin Van Durme, Zhuowen Li and Elias Stengel-Eskin ................ 5321
QuALITY: Question Answering with Long Input Texts, Yes!
Richard Yuanzhe Pang, Alicia Parrish, Nitish Joshi, Nikita Nangia, Jason Phang, Angelica Chen, Vishakh Padmakumar, Johnny L. Ma, Jana Thompson, He He and Samuel R. Bowman ............... 5336

ExSum: From Local Explanations to Model Understanding
Yilun Zhou, Marco Tulio Ribeiro and Julie Shah .................................................. 5359

Maximum Bayes Smatch Ensemble Distillation for AMR Parsing
Young-Suk Lee, Ramón Fernandez Astudillo, Hoang Thanh Lam, Tahira Naseem, Radu Florian and Salim Roukos ................................................................. 5379

When Does Syntax Mediate Neural Language Model Performance? Evidence from Dropout Probes
Mycal Tucker, Tiwalayo Eisape, Peng Qian, Roger P. Levy and Julie Shah .................. 5393

Modeling Task Interactions in Document-Level Joint Entity and Relation Extraction
Liyan Xu and Jinho D. Choi .......................................................... 5409

Few-Shot Semantic Parsing with Language Models Trained on Code
Richard Shin and Benjamin Van Durme .................................................. 5417

CORWA: A Citation-Oriented Related Work Annotation Dataset
Xiangci Li, Biswadip Mandal and Jessica Ouyang .................................................. 5426

Overcoming Catastrophic Forgetting During Domain Adaptation of Seq2seq Language Generation
Dingcheng Li, Zheng Chen, Eunah Cho, Jie Hao, Xiaohu Liu, Fan Xing, Chenlei Guo and Yang Liu .................................................................................................................. 5441

Extreme Zero-Shot Learning for Extreme Text Classification
Yuanhao Xiong, Wei-Cheng Chang, Cho-Jui Hsieh, Hsiang-Fu Yu and Inderjit S Dhillon . 5455

ConflibERT: A Pre-trained Language Model for Political Conflict and Violence
Yibo Hu, MohammadSaleh Hosseini, Erick Skorupa Parolin, Javier Osorio, Latifur Khan, Patrick Brandt and Vito D’Orazio .......................................................... 5469

Automatic Multi-Label Prompting: Simple and Interpretable Few-Shot Classification
Han Wang, Canwen Xu and Julian McAuley .................................................. 5483

Few-shot Subgoal Planning with Language Models
Lajanugen Logeswaran, Yao Fu, Moontae Lee and Honglak Lee .................................. 5493

IDPG: An Instance-Dependent Prompt Generation Method
Zhuofeng Wu, Sinong Wang, Jiatao Gu, Rui Hou, Yuxiao Dong, V.G.Vinod Vydiswaran and Hao Ma .......................................................... 5507

Embedding Hallucination for Few-shot Language Fine-tuning
Yiren Jian, Chongyang Gao and Soroush Vosoughi .................................................. 5522

Cryptocurrency Bubble Detection: A New Stock Market Dataset, Financial Task & Hyperbolic Models
Ramit Sawhney, Shivam Agarwal, Vivek Mittal, Paolo Rosso, Vikram Nanda and Sudheer Chava 5531

Nearest Neighbor Knowledge Distillation for Neural Machine Translation
Zhixian Yang, Renliang Sun and Xiaojun Wan .................................................. 5546

DEMix Layers: Disentangling Domains for Modular Language Modeling
Suchin Gururangan, Mike Lewis, Ari Holtzman, Noah Smith and Luke Zettlemoyer ....... 5557
Contrastive Learning for Prompt-based Few-shot Language Learners
Yiren Jian, Chongyang Gao and Soroush Vosoughi ........................................... 5577

Cross-Lingual Event Detection via Optimized Adversarial Training
Luis Fernando Guzman-Nateras, Minh Van Nguyen and Thien Huu Nguyen ........... 5588

Identifying Implicitly Abusive Remarks about Identity Groups using a Linguistically Informed Approach
Michael Wiegand, Elisabeth Eder and Josef Ruppenhofer ...................................... 5600

Label Definitions Improve Semantic Role Labeling
Li Zhang, Ishan Jindal and Yunyao Li ................................................................. 5613

Shedding New Light on the Language of the Dark Web
Youngjin Jin, Eugene Jang, Yongjae Lee, Seungwon Shin and Jin-Woo Chung .......... 5621

Conceptualizing Treatment Leakage in Text-based Causal Inference
Adel Daoud, Connor Thomas Jerzak and Richard Johansson ................................. 5638

Consistency Training with Virtual Adversarial Discrete Perturbation
Jungsoo Park, Gyuwan Kim and Jaewoo Kang ..................................................... 5646

CONFIT: Toward Faithful Dialogue Summarization with Linguistically-Informed Contrastive Fine-tuning
Xiangru Tang, Arjun Nair, Borui Wang, Bingyao Wang, Jai Amit Desai, Aaron Wade, Haoran Li, Asli Celikyilmaz, Yashar Mehdad and Dragomir Radev ...................................................... 5657

CoMPM: Context Modeling with Speaker’s Pre-trained Memory Tracking for Emotion Recognition in Conversation
Joosung Lee and Woojin Lee .................................................................................... 5669

Investigating Crowdsourcing Protocols for Evaluating the Factual Consistency of Summaries
Xiangru Tang, Alexander Fabbri, Haoran Li, Ziming Mao, Griffin Thomas Adams, Borui Wang, Asli Celikyilmaz, Yashar Mehdad and Dragomir Radev ...................................................... 5680

DialSummEval: Revisiting Summarization Evaluation for Dialogues
Mingqi Gao and Xiaojun Wan .................................................................................... 5693

Hyperbolic Relevance Matching for Neural Keyphrase Extraction
Mingyang Song, Yi Feng and Liping Jing .................................................................. 5710

Template-free Prompt Tuning for Few-shot NER
Ruotian Ma, Xin Zhou, Tao Gui, Yiding Tan, Linyang Li, Qi Zhang and Xuanjing Huang . 5721

Few-Shot Document-Level Relation Extraction
Nicholas Popovic and Michael Färber ..................................................................... 5733

LaMemo: Language Modeling with Look-Ahead Memory
Haozhe Ji, Rongsheng Zhang, Zhenyu Yang, Zhipeng Hu and Minlie Huang .......... 5747

Exploiting Inductive Bias in Transformers for Unsupervised Disentanglement of Syntax and Semantics with VAEs
Ghazi Felhi, Joseph Le Roux and Djamé Seddah .................................................... 5763

Neighbors Are Not Strangers: Improving Non-Autoregressive Translation under Low-Frequency Lexical Constraints
Chun Zeng, Jiangjie Chen, Tianyi Zhuang, Rui Xu, Hao Yang, Qin Ying, Shimin Tao and Yanghua Xiao ................................................................................................................. 5777
What do Toothbrushes do in the Kitchen? How Transformers Think our World is Structured
Alexander Henlein and Alexander Mehler ................................................................. 5791

Less is More: Learning to Refine Dialogue History for Personalized Dialogue Generation
Hanzxun Zhong, Zhicheng Dou, Yutao Zhu, Hongjin Qian and Ji-Rong Wen ............... 5808

A Holistic Framework for Analyzing the COVID-19 Vaccine Debate
Maria Leonor Pacheco, Tunazzina Islam, Monal Mahajan, Andrey Shor, Ming Yin, Lyle Ungar and Dan Goldwasser ......................................................... 5821

Learning to Win Lottery Tickets in BERT Transfer via Task-agnostic Mask Training
Yuanxin Liu, Fandong Meng, Zheng Lin, Peng Fu, Yanan Cao, Weiping Wang and Jie Zhou 5840

You Don’t Know My Favorite Color: Preventing Dialogue Representations from Revealing Speakers’ Private Personas
Haoran Li, Yangqiu Song and Lixin Fan ................................................................. 5858

Explaining Dialogue Evaluation Metrics using Adversarial Behavioral Analysis
Baber Khalid and Sungjin Lee ...................................................................................... 5871

Annotators with Attitudes: How Annotator Beliefs And Identities Bias Toxic Language Detection
Maarten Sap, Swabha Swayamdipta, Laura Vianna, Xuhui Zhou, Yejin Choi and Noah Smith 5884

Non-Autoregressive Chinese ASR Error Correction with Phonological Training
Zheng Fang, Ruiqing Zhang, Zhongjun He, Hua Wu and Yanan Cao .......................... 5897

Hate Speech and Counter Speech Detection: Conversational Context Does Matter
Xinchen Yu, Eduardo Blanco and Lingzi Hong .......................................................... 5918

DACSA: A large-scale Dataset for Automatic summarization of Catalan and Spanish newspaper Articles
Encarnación Segarra Soriano, Vicent Ahuir, Lluís-F. Hurtado and José Ángel González . 5931

Time Waits for No One! Analysis and Challenges of Temporal Misalignment
Kelvin Luu, Daniel Khashabi, Suchin Gururangan, Karishma Mandyam and Noah Smith . 5944

MCSE: Multimodal Contrastive Learning of Sentence Embeddings
Miaoran Zhang, Marius Mosbach, David Ifeoluwa Adelani, Michael A. Hedderich and Dietrich Klakow ................................................................. 5959

HiUERE: Hierarchical Exemplar Contrastive Learning for Unsupervised Relation Extraction
Shuliang Liu, Xuming Hu, Chenwei Zhang, Shu’ang Li, Lijie Wen and Philip S. Yu ........ 5970

Wanrong Zhu, Yuankai Qi, Pradyumna Narayana, Kazoo Sone, Sugato Basu, Xin Eric Wang, Qi Wu, Miguel Eckstein and William Yang Wang .............................................. 5981

Aligning to Social Norms and Values in Interactive Narratives
Prithviraj Ammanabrolu, Liwei Jiang, Maarten Sap, Hannaneh Hajishirzi and Yejin Choi . 5994

MOVER: Mask, Over-generate and Rank for Hyperbole Generation
Yunxiang Zhang and Xiaojun Wan .............................................................................. 6018

Embarrassingly Simple Performance Prediction for Abductive Natural Language Inference
Emils Kadik, Vaibhav Srivastav and Roman Klinger ................................................. 6031

Re-Examining System-Level Correlations of Automatic Summarization Evaluation Metrics
Daniel Deutsch, Rotem Dror and Dan Roth .............................................................. 6048