*SEM 2022

The 11th Joint Conference on Lexical and Computational Semantics

Proceedings of the Conference

July 14-15, 2022

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-955917-98-8

Message from the General Chair and the Program Chairs

Welcome to ***SEM 2022**, the 11th Joint Conference on Lexical and Computational Semantics! We are pleased to present this volume containing the accepted long and short papers. *SEM 2022 was held as a hybrid conference following NAACL 2022, on July 14th-15th, 2022, due to the precautions for the COVID-19 pandemic.

Since its first edition in 2012, *SEM has become a major venue to present recent advances in all areas of lexical and computational semantics, including semantic representations, semantic processing, multilingual semantics, and others. *SEM is sponsored by SIGLEX, the ACL Special Interest Group on the Lexicon.

*SEM 2022 had a hybrid format with respect to ARR. We accepted both direct submissions through the START system and also those already reviewed through ARR. In total, we received 52 submissions in 9 areas:

- Theoretical and formal semantics
- Semantics in NLP applications
- Semantic composition and sentence-level semantics
- Resources and evaluation
- Psycholinguistics, cognitive linguistics and semantic processing
- Multilinguality
- Lexical semantics and word representations
- Commonsense reasoning and natural language understanding

We compiled an exciting program across all these areas. This year saw a particularly strong batch of submissions; finally, **30** papers were accepted – **18** long papers and **12** short papers.

The submitted papers were carefully evaluated by a program committee led by 11 area chairs, who coordinated a panel of 100 reviewers (who were assigned papers to review in the START system). Almost all submissions were reviewed by three reviewers, who were encouraged to discuss any divergence in evaluations. The papers in each area were subsequently assessed by the area chairs, who added meta-reviews to explain their accept/reject suggestions. The final selection was made by the program co-chairs after an independent check of all the reviews, meta-reviews, and discussions with the area chairs. The reviewers' recommendations were also used to shortlist a set of papers nominated for the Best Paper Award.

We are also very excited to have two excellent keynote speakers: **Allyson Ettinger** (University of Chicago) discussing controlled examinations of meaning sensitivity in pre-trained NLP models, and **Jacob Andreas** (Massachusetts Institute of Technology) discussing the extent to which language modeling induces representations of meaning.

We are deeply thankful to all area chairs and reviewers for their invaluable help in the selection of the program, for their readiness in engaging in thoughtful discussions about individual papers, and for providing valuable feedback to the authors. We are grateful to our Publicity chair, Jose Camacho-Collados (Cardiff University), who set up and regularly updated *SEM's website and publicized it through social media. We thank the Publication Chair, Alessandro Raganato (University of Milano-Bicocca), for his help with the compilation of the proceedings, and the NAACL 2022 workshop organizers for all the valuable help and support with organisational aspects of the conference. Finally, we thank all our authors and presenters for making *SEM 2022 such an exciting event. We hope you will find the content of these proceedings as well as the program of *SEM 2022 enjoyable, interesting and inspirational!

Ellie Pavlick and **Mohammad Taher Pilehvar**, Program Co-Chairs
Vivi Nastase, General Chair

Organizing Committee

General Chair

Vivi Nastase, University of Zurich

Program Chairs

Ellie Pavlick, Brown University
Mohammad Taher Pilehvar, Tehran Institute for Advanced Studies

Publicity Chair

Jose Camacho-Collados, Cardiff University

Publication Chair

Alessandro Raganato, University of Milano-Bicocca

Program Committee

Area Chairs

Marianna Apidianaki, University of Pennsylvania
Vered Shwartz, University of British Columbia
Allyson Ettinger, University of Chicago
Nafise Sadat Moosavi, TU Darmstadt
Malihe Alikhani, University of Pittsburgh
Anders Søgaard, University of Copenhagen
Najoung Kim, Johns Hopkins University
Daniel Khashabi, Allen Institute for Artificial Intelligence
Gene Kim, University of Rochester
Keisuke Sakaguchi, Allen Institute for AI
Nazneen Rajani, Salesforce Research

Program Committee

Lasha Abzianidze, Utrecht University
Rodrigo Agerri, HiTZ Center - Ixa, University of the Basque Country UPV/EHU
Md. Shad Akhtar, Indraprastha Institute of Information Technology, Delhi
Dimitris Alikaniotis, Grammarly Inc.
Forrest Sheng Bao, Iowa State Univerity
Mohamad Hardyman Barawi, University of Malaysia, Sarawak
Pierpaolo Basile, Department of Computer Science, University of Bari Aldo Moro
Tilman Beck, UKP Lab, Technical University of Darmstadt
Farah Benamara, University of toulouse
Gábor Berend, University Of Szeged
Jean-philippe Bernardy, University of Gothenburg
Eduardo Blanco, Arizona State University
Michael Bloodgood, The College of New Jersey
Marianna Bolognesi, University of Bologna
Johan Bos, University of Groningen
Paul Buitelaar, National University of Ireland Galway
Elena Cabrio, Université Côte d'Azur, Inria, CNRS, I3S
Aoife Cahill, Dataminr
Franklin Chang, Kobe City University of Foreign Studies
Aditi Chaudhary, Carnegie Mellon University
Pinzhen Chen, University of Edinburgh
Emmanuele Chersoni, Hong Kong Polytechnic University
Patricia Chiril, University of Chicago
Christos Christodoulopoulos, Amazon Research
Philipp Cimiano, Univ. Bielefeld
Robin Cooper, University of Gothenburg
Bonaventura Coppola, University of Trento
Walter Daelemans, University of Antwerp, CLiPS
Joachim Daiber, Apple
Luna De Bruyne, LT3, Language and Translation Technology Team, Ghent University
Gaël Dias, Normandie Univ, UNICAEN, ENSICAEN, CNRS, GREYC
Liviu P. Dinu, University of Bucharest

Lucia Donatelli, Saarland University
Jakub Dotlacil, Utrecht University
Luis Espinosa Anke, Cardiff University
Dan Garrette, Google Research
Debela Gemechu, Centre for Argument Technology, School of Science & Engineering,University of Dundee
Voula Giouli, ATHENA Research & Innovation Centre, Institute for Language & Speech Processing
Max Glockner, UKP, Computer Science Dpt., TU Darmstadt
Binod Gyawali, Educational Testing Service
Ivan Habernal, Technische Universität Darmstadt
Udo Hahn, Friedrich-Schiller-Universität Jena
Mareike Hartmann, University of Copenhagen
Alessandro Raganato, University of Milano-Bicocca
Dag Haug, University of Oslo
Yoshihiko Hayashi, Waseda University
Daniel Hershcovich, University of Copenhagen
Xinyu Hua, Northeastern University
Nancy Ide, Vassar College
Joseph Marvin Imperial, National University, Manila, Philippines
Sk Mainul Islam, IIT Kharagpur
Kokil Jaidka, National University of Singapore
Jenna Kanerva, University of Turku
Mladen Karan, Queen Mary University
Omid Kashefi, ETS
Roman Klinger, University of Stuttgart
Thomas Kober, NULL
Elena Kochkina, Queen Mary University
Valia Kordoni, Humboldt-Universität zu Berlin
Dan Lassiter, University of Edinburgh
Yitong Li, Huawei Technology Co. ltd
Zongxi Li, Hong Kong Metropolitan University
Nikola Ljubešić, Jožef Stefan Institute
Lorenzo Malandri, University of Milan - Bicocca
Alda Mari, http://www.institutnicod.org/
Eugenio Martínez-cámara, University of Granada
Jonathan May, USC Information Sciences Institute
Sahisnu Mazumder, Intel Labs
Nick Mckenna, University of Edinburgh, School of Informatics
Julian Michael, University of Washington
Koji Mineshima, Keio University
Amita Misra, IBM
Ali Modarressi, Iran University of Science and Technology
Andrew Moore, Lancaster University
Richard Moot, CNRS
Véronique Moriceau, IRIT, Université Toulouse 3
Gaku Morio, Research & Development Group, Hitachi, Ltd.
Larry Moss, Indiana University, Bloomington
Philippe Muller, IRIT, University of Toulouse
Nona Naderi, University of Applied Sciences HES-SO Genève, Swiss Institute of Bioinformatics (SIB)

Vivi Nastase, University of Stuttgart
Timothy Niven, Doublethink Lab
Debora Nozza, Bocconi University
Tim O'gorman, Thorn
Emerson Paraiso, Pontificia Universidade Catolica do Parana - PUCPR
Patrick Paroubek, University Paris-Saclay - CNRS - LISN
Maxime Peyrard, EPFL
Sandro Pezzelle, University of Amsterdam
Jonas Pfeiffer, TU Darmstadt
Yuval Pinter, Ben-Gurion University of the Negev
Marco Polignano, University of Bari
Sara Rajaee, Iran University of Science and Technology
Carlos Ramisch, Aix Marseille University, CNRS, LIS
Christian Retoré, University of Montpellier
Elijah Rippeth, University of Maryland
Alla Rozovskaya, Queens College, City University of New York
Irene Russo, ILC CNR
Farig Sadeque, Educational Testing Service
Mehrnoosh Sadrzadeh, University College London
Marina Santini, RISE, Research Institutes of Sweden. Division: Digital Systems
Ryohei Sasano, Nagoya University
David Schlangen, University of Potsdam
Sabine Schulte Im Walde, University of Stuttgart
Weiyan Shi, Columbia University
Melanie Siegel, Hochschule Darmstadt - University of Applied Sciences
Egon Stemle, Eurac Research
Kevin Stowe, Educational Testing Services (ETS)
Sara Stymne, Uppsala University
Yoshi Suhara, Grammarly
Alexandros Tantos, Aristotle University of Thessaloniki
Andon Tchechmedjiev, IMT Mines Alès
Gaurav Singh Tomar, Google Research
Samia Touileb, University of Bergen
Shyam Upadhyay, Google
L. Alfonso Ureña-lópez, University of Jaen
Sowmya Vajjala, National Research Council
Tim Van De Cruys, University of Leuven
Rossella Varvara, University of Fribourg
Eva Maria Vecchi, Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung
Noortje Venhuizen, Saarland University
Serena Villata, Université Côte d'Azur, CNRS, Inria, I3S
Ivan Vulić, University of Cambridge
Ekaterina Vylomova, University of Melbourne
Wenbo Wang, GoDaddy Inc.
Jin Wang, Yunnan University
Zhongqing Wang, Soochow University
Jin Wang, Megagon Lab
Bonnie Webber, University of Edinburgh
Michael Wiegand, Alpen-Adria-Universitaet Klagenfurt
Adina Williams, Facebook, Inc.
Genta Winata, Bloomberg

Tak-lam Wong, Department of Computing Studies and Information Systems, Douglas College
Shijie Wu, Johns Hopkins University
Rong Xiang, The Hong Kong Polytechnic University
Ruifeng Xu, Harbin Institute of Technology, Shenzhen
Bei Yu, Syracuse University
Alessandra Zarcone, Hochschule für angewandte Wissenschaften
Chrysoula Zerva, Instituto de Telecomunicações, Instituto Superior Técnico, University of Lisbon
Lei Zhang, LinkedIn
Shuai Zhang, Amazon

Keynote Talk: "Understanding" and prediction: Controlled examinations of meaning sensitivity in pre-trained models

Allyson Ettinger
University of Chicago

Abstract: In recent years, NLP has made what appears to be incredible progress, with performance even surpassing human performance on some benchmarks. How should we interpret these advances? Have these models achieved language "understanding"? Operating on the premise that "understanding" will necessarily involve the capacity to extract and deploy meaning information, in this talk I will discuss a series of projects leveraging targeted tests to examine NLP models' ability to capture meaning in a systematic fashion. I will first discuss work probing model representations for compositional meaning, with a particular focus on disentangling compositional information from encoding of lexical properties. I'll then explore models' ability to extract and use meaning information when executing the basic pre-training task of word prediction in context. In all cases, these investigations apply tests that prioritize control of unwanted cues, so as to target the desired model capabilities with greater precision. The results of these studies suggest that although models show a good deal of sensitivity to word-level information, and to certain semantic and syntactic distinctions, when subjected to controlled tests they show little sign of representing higher-level compositional meaning, or of being able to retain and deploy such information robustly during word prediction. Instead, models show signs of heuristic predictive strategies that are unsurprising given their training, but that differ critically from systematic understanding of meaning. I will discuss potential implications of these findings with respect to the goals of achieving "understanding" with currently dominant pre-training paradigms.

Bio: Allyson Ettinger is an Assistant Professor in the Department of Linguistics at the University of Chicago. Her interdisciplinary work combines methods and insights from cognitive science, linguistics, and computer science to examine meaning extraction and predictive processes executed during language processing in artificial intelligence systems and in humans. She received her PhD in Linguistics from the University of Maryland, and spent a year as research faculty at the Toyota Technological Institute at Chicago (TTIC) before beginning her appointment at the University of Chicago. She holds an additional courtesy appointment at TTIC.

Keynote Talk: Models of meaning?

Jacob Andreas
Massachusetts Institute of Technology

Abstract: The extent to which language modeling induces representations of meaning—and the broader question of whether it is even in principle possible to learn about meaning from text alone—have remained a subject of ongoing debate across the language sciences. I'll present some evidence that transformer language models build (rudimentary) structured representations of the meaning of input sentences; that these representations support LMs' ability to reason about the entities and events described in a discourse; and that they can be modified with predictable effects on downstream language generation. Despite all this, even the largest LMs are prone to glaring semantic errors: they refer to entities that have not yet been mentioned, present contradictory facts, or describe impossible events. By understanding how (and where) LMs build models of meaning, we identify the causes of these errors, and in some cases correct them with extremely small amounts of targeted supervision.

Bio: Jacob Andreas is the X Consortium Assistant Professor at MIT. His research aims to build intelligent systems that can communicate effectively using language and learn from human guidance. Jacob earned his Ph.D. from UC Berkeley, his M.Phil. from Cambridge (where he studied as a Churchill scholar) and his B.S. from Columbia. As a researcher at Microsoft Semantic Machines, he founded the language generation team and helped develop core pieces of the technology that powers conversational interaction in Microsoft Outlook. He has been the recipient of Samsung's AI Researcher of the Year award, MIT's Kolokotrones teaching award, and paper awards at NAACL and ICML.

Table of Contents

Program

Thursday, July 14, 2022

08:30 - 10:00 *Sentence-Level Semantics*

Compositional generalization with a broad-coverage semantic parser
Pia Weißenhorn, Lucia Donatelli and Alexander Koller

DRS Parsing as Sequence Labeling
Minxing Shen and Kilian Evang

Measuring Alignment Bias in Neural Seq2seq Semantic Parsers
Davide Locatelli and Ariadna Quattoni

Semantics-aware Attention Improves Neural Machine Translation
Aviv Slobodkin, Leshem Choshen and Omri Abend

Multilingual Extraction and Categorization of Lexical Collocations with Graph-aware Transformers
Luis Espinosa Anke, Alexander Shvets, Alireza Mohammadshahi, James Henderson and Leo Wanner

Comparison and Combination of Sentence Embeddings Derived from Different Supervision Signals
Hayato Tsukagoshi, Ryohei Sasano and Koichi Takeda

10:00 - 10:30 *Break*

10:30 - 12:00 *Evaluation*

Dyna-bAbI: unlocking bAbI's potential with dynamic synthetic benchmarking
Ronen Tamari, Kyle Richardson, Noam Kahlon, Aviad Sar-shalom, Nelson F. Liu, Reut Tsarfaty and Dafna Shahaf

A Dynamic, Interpreted CheckList for Meaning-oriented NLG Metric Evaluation – through the Lens of Semantic Similarity Rating
Laura Zeidler, Juri Opitz and Anette Frank

A Generative Approach for Mitigating Structural Biases in Natural Language Inference
Dimion Asael, Zachary Ziegler and Yonatan Belinkov

Thursday, July 14, 2022 (continued)

How Does Data Corruption Affect Natural Language Understanding Models? A Study on GLUE datasets
Aarne Talman, Marianna Apidianaki, Stergios Chatzikyriakidis and Jörg Tiedemann

AnaLog: Testing Analytical and Deductive Logic Learnability in Language Models
Samuel Ryb, Mario Giulianelli, Arabella Sinclair and Raquel Fernández

Unsupervised Reinforcement Adaptation for Class-Imbalanced Text Classification
Yuexin Wu and Xiaolei Huang

12:00 - 13:30 *Break*

13:30 - 15:00 *Invited Talk: Models of meaning? - Jacobs Andreas*

15:00 - 15:30 *Break*

15:30 - 17:00 *Lexical Semantics*

Distilling Hypernymy Relations from Language Models: On the Effectiveness of Zero-Shot Taxonomy Induction
Devansh Jain and Luis Espinosa Anke

PropBank Comes of Age—Larger, Smarter, and more Diverse
Sameer Pradhan, Julia Bonn, Skatje Myers, Kathryn Conger, Tim O'gorman, James Gung, Kristin Wright-bettner and Martha Palmer

Pretraining on Interactions for Learning Grounded Affordance Representations
Jack Merullo, Dylan Ebert, Carsten Eickhoff and Ellie Pavlick

Leveraging Three Types of Embeddings from Masked Language Models in Idiom Token Classification
Ryosuke Takahashi, Ryohei Sasano and Koichi Takeda

When Polysemy Matters: Modeling Semantic Categorization with Word Embeddings
Elizabeth Soper and Jean-pierre Koenig

What Drives the Use of Metaphorical Language? Negative Insights from Abstractness, Affect, Discourse Coherence and Contextualized Word Representations
Prisca Piccirilli and Sabine Schulte Im Walde

Assessing the Limits of the Distributional Hypothesis in Semantic Spaces: Trait-based Relational Knowledge and the Impact of Co-occurrences
Mark Anderson and Jose Camacho-collados

Friday, July 15, 2022

08:30 - 10:00 *Discourse and Dialog*

A Simple Unsupervised Approach for Coreference Resolution using Rule-based Weak Supervision
Alessandro Stolfo, Chris Tanner, Vikram Gupta and Mrinmaya Sachan

Online Coreference Resolution for Dialogue Processing: Improving Mention-Linking on Real-Time Conversations
Liyan Xu and Jinho D. Choi

DeepA2: A Modular Framework for Deep Argument Analysis with Pretrained Neural Text2Text Language Models
Gregor Betz and Kyle Richardson

"What makes a question inquisitive?" A Study on Type-Controlled Inquisitive Question Generation
Lingyu Gao, Debanjan Ghosh and Kevin Gimpel

Speech acts and Communicative Intentions for Urgency Detection
Laurenti Enzo, Bourgon Nils, Farah Benamara, Mari Alda, Véronique Moriceau and Courgeon Camille

What do Large Language Models Learn about Scripts?
Abhilasha Sancheti and Rachel Rudinger

10:00 - 10:30 *Break*

10:30 - 12:00 *Events*

Pairwise Representation Learning for Event Coreference
Xiaodong Yu, Wenpeng Yin and Dan Roth

Event Causality Identification via Generation of Important Context Words
Hieu Man, Minh Nguyen and Thien Nguyen

Word-Label Alignment for Event Detection: A New Perspective via Optimal Transport
Amir Pouran Ben Veyseh and Thien Nguyen

Friday, July 15, 2022 (continued)

Capturing the Content of a Document through Complex Event Identification
Zheng Qi, Elior Sulem, Haoyu Wang, Xiaodong Yu and Dan Roth

Improved Induction of Narrative Chains via Cross-Document Relations
Andrew Blair-stanek and Benjamin Van Durme

12:00 - 13:30 *Break*

13:30 - 15:00 *Invited Talk: "Understanding" and prediction: Controlled examinations of meaning sensitivity in pre-trained models - Allyson Ettinger*

15:00 - 15:30 *Break*

15:30 - 17:00 *Panel discussion, Closing*