

Yulia Tsvetkov

CONTACT INFORMATION

Paul G. Allen Center
185 E Stevens Way NE,
Seattle, WA 98195
USA

phone: +1-412-204-7407
office: CSE 566
email: yuliats@cs.washington.edu
web: <https://homes.cs.washington.edu/~yuliats>

RESEARCH INTERESTS

I work on Natural Language Processing—a subfield of computer science and AI focusing on computational processing of human languages. I am particularly interested in hybrid solutions at the intersection of machine learning and theoretical or social linguistics, i.e., practical solutions that combine sophisticated learning and modeling methods with insights into human languages and the people who speak them. Much of my research is focused on NLP for social good, multilingual NLP, and language generation; it is motivated by a unified goal: to extend the capabilities of human language technology beyond individual populations and across language boundaries, thereby enabling NLP for diverse and disadvantaged users, the users that need it most.

PROFESSIONAL EXPERIENCE

University of Washington

Paul G. Allen School of Computer Science & Engineering

Assistant Professor

2021–now

Carnegie Mellon University

Language Technologies Institute, School of Computer Science

Assistant Professor

2017–2021

Stanford University

Department of Computer Science

Postdoctoral research associate, advised by Dan Jurafsky.

2016–2017

EDUCATION

Carnegie Mellon University

Language Technologies Institute, School of Computer Science

Ph.D., advised by Chris Dyer.

2011–2016

- Thesis: “Linguistic Knowledge in Data-Driven Natural Language Processing”

University of Haifa, Israel

Department of Computer Science

M.Sc. *Cum Laude*, advised by Shuly Wintner.

2007–2010

- Thesis: “Extraction of Multi-word Expressions from Small Parallel Corpora”.
- Awards: Gutwirth award for excellence (awarded top 50 graduate students in Israel); excellence scholarship in honor of Captain Uri Akavia.

Technion - Israel Institute of Technology

B.Sc. in Computer Science and Education.

2000–2005

FUNDING
AND
AWARDS

Sloan Research Fellowship
February 2021

Title: NSF: CAREER: Language Technologies Against the Language of Social Discrimination
Source of Support: National Science Foundation
Total Award Amount: \$550,436
Period of Performance: 9/1/22 to 8/31/27

Title: NSF: Collaborative Research: RI: Small: NL(V)P: Natural Language (Variety) Processing
Co-PIs: Antonios Anastasopoulos (GMU), David Chiang (ND)
Source of Support: National Science Foundation
Total Award Amount (UW portion): \$165,949
Period of Performance: 9/1/21 to 8/31/23

Title: Fostering Civil Interactions and Combating Social Bias in Workplace Communications
Source of Support: Workhuman
Total Award Amount: \$280,000
Period of Performance: 9/1/21 to 8/31/23

Title: NSF: FAI: Quantifying and Mitigating Disparities in Language Technologies
Co-PIs: Graham Neubig (CMU), Geoff Kaufman (CMU), Antonios Anastasopoulos (GMU)
Source of Support: National Science Foundation
Total Award Amount (CMU portion): \$600,000
Period of Performance: 8/1/21 to 7/31/23

Title: NSF: Collaborative Research: RI: Small: NSF-BSF: Multilingual Language Generation via Understanding of Code Switching
Co-PIs: Melinda Fricke (UPitt)
Source of Support: National Science Foundation
Total Award Amount (UW portion): \$345,563
Period of Performance: 10/01/20 to 9/31/23

Title: NNSA: Usable and Explainable Models for Global-Scale Cross-Lingual Proliferation Expertise Identification and Forecasting
Co-PI: Svitlana Volkova (PNNL)
Source of Support: Pacific Northwest National Laboratory (PNNL)
Total Award Amount: \$280,000
Period of Performance: 9/1/2020 to 12/31/2022

Title: Multi-media Analytics Leading to Intent and Semantic Evidence (MALISE)
Co-PIs: SRI International
Source of Support: DARPA SemaFor program
Total Award Amount (UW portion): \$737,885
Period of Performance: 06/01/20 to 05/31/24

Title: NSF: Standard: Promoting Ethics in STEM with Transparency Initiatives that Work: An Interdisciplinary Investigation
Co-PIs: Silvia Saccardo, Gretchen Chapman (CMU)
Source of Support: National Science Foundation
Total Award Amount: \$397,340
Period of Performance: 09/01/19 to 08/31/22

Title: Bridging the "Ethical Gap" in CS Education by Teaching Socially Responsible Language Technologies

Source of Support: Public Interest Technology University Network – New America Foundation
Total Award Amount: \$90,000
Period of Performance: 10/1/2019 to 09/31/2020

Title: NSF-BSF: RI: Small: Collaborative Research: Modeling Crosslinguistic Influences between Language Varieties
Co-PIs: Noah Smith (UW), Nathan Schneider (U Georgetown), Shuly Wintner (U Haifa)
Source of Support: National Science Foundation
Total Award Amount (CMU portion): \$166,000
Period of Performance: 9/1/2018 to 8/31/2020

Title: NSF: RI: Small: Modeling Lexical Borrowing to Bridge the “Linguistic Divide” in Natural Language Processing
Co-PI: Chris Dyer
Source of Support: National Science Foundation
Total Award Amount: \$500,000
Period of Performance: 9/1/2016 to 8/31/2019

Title: Language Generation into Embedding Spaces
Source of Support: Google Faculty Research Award
Total Award Amount: \$80,000
April 2020

Title: Language Generation Without Softmax Layer: Open-Vocabulary, Accurate, Efficient and Compact
Source of Support: Amazon Machine Learning Research Award
Total Award Amount: \$40,000 (plus \$60,000 in AWS credits)
September 2019

Okawa Foundation Young Researcher Award
April 2019

Adobe Researcher Gift
May 2019

JOURNAL
ARTICLES

An Analysis of Emotions and the Prominence of Positivity in #BlackLivesMatter Tweets

Anjalie Field, Chan Young Park, Antonio Theophilo, Jamelle Watson-Daniels, and Yulia Tsvetkov
In review at Proceedings of the National Academy of Sciences (PNAS), 2021.

A Framework for the Computational Linguistic Analysis of Dehumanization

Julia Mendelsohn, Yulia Tsvetkov, and Dan Jurafsky
Frontiers in Artificial Intelligence, 2020.

Native Language Cognate Effects on Second Language Lexical Choice

Ella Rabinovich, Yulia Tsvetkov and Shuly Wintner
Transactions of the Association for Computational Linguistics, 2018.

Cross-Lingual Bridges with Models of Lexical Borrowing

Yulia Tsvetkov and Chris Dyer
Journal of Artificial Intelligence Research, 55:63–93, 2016. PDF.

Identification of Multi-word Expressions by Combining Multiple Linguistic Information Sources

Yulia Tsvetkov and Shuly Wintner
Computational Linguistics, 40(2):449-468, 2014. PDF.

Extraction of Multi-word Expressions from Small Parallel Corpora
Yulia Tsvetkov and Shuly Wintner
Natural Language Engineering 18(4):549-573, 2012. PDF.

CONFERENCE
PUBLICATIONS

**Speaker Information Can Guide Models to Better Inductive Biases:
A Case Study On Predicting Code-Switching.**
Alissa Ostapenko, Shuly Wintner, Melinda Fricke and Yulia Tsvetkov.
In *Proceedings of ACL 2022*

Controlled Analyses of Social Biases in Wikipedia Bios.
Anjalie Field, Chan Young Park, Kevin Z. Lin and Yulia Tsvetkov.
In *Proceedings of TheWebConf 2022*

SimVLM: Simple Visual Language Model Pretraining with Weak Supervision.
Zirui Wang, Jiahui Yu, Adams Wei Yu, Zihang Dai, Yulia Tsvetkov and Yuan Cao.
In *Proceedings of ICLR 2022*

Controlled Text Generation as Continuous Optimization with Multiple Constraints.
Sachin Kumar, Eric Malmi, Aliaksei Severyn and Yulia Tsvetkov.
In *Proceedings of NeurIPS 2021*

SelfExplain: A Self-Explaining Architecture for Neural Text Classifiers.
Dheeraj Rajagopal, Vidhisha Balachandran, Eduard Hovy and Yulia Tsvetkov.
In *Proceedings of EMNLP 2021*

Evaluating the Morphosyntactic Well-formedness of Generated Texts.
Adithya Pratapa, Antonios Anastasopoulos, Shruti Rijhwani, Aditi Chaudhary, David R. Mortensen,
Graham Neubig and Yulia Tsvetkov.
In *Proceedings of EMNLP 2021*

**Influence Tuning: Demoting Spurious Correlations via Instance Attribution and
Instance-Driven Updates.**
Xiaochuang Han and Yulia Tsvetkov.
In *Findings of EMNLP 2021*

Detecting Community Sensitive Norm Violations in Online Conversations.
Chan Young Park, Julia Mendelsohn, Karthik Radhakrishnan, Kinjal Jain, Tushar Kanakagiri,
David Jurgens and Yulia Tsvetkov.
In *Findings of EMNLP 2021*

Efficient Test Time Adapter Ensembling for Low-resource Language Varieties.
Xinyi Wang, Yulia Tsvetkov, Sebastian Ruder and Graham Neubig.
In *Findings of EMNLP 2021*

Simple and Efficient ways to Improve REALM.
Vidhisha Balachandran, Ashish Vaswani, Yulia Tsvetkov and Niki Parmar.
In *Proceedings of MRQA 2021*

Improving the Diversity of Unsupervised Paraphrasing with Embedding Outputs.
Monisha Jegadeesan, Sachin Kumar, John Wieting and Yulia Tsvetkov.
In *Proceedings of MRL 2021*

Improving Span Representation for Domain-adapted Coreference Resolution.

Nupoor Gandhi, Anjalie Field and Yulia Tsvetkov.

In *Proceedings of CRAC 2021*

A Survey of Race, Racism, and Anti-Racism in NLP.

Anjalie Field, Su Lin Blodgett, Zeerak Waseem and Yulia Tsvetkov.

In *Proceedings of ACL 2021*

Machine Translation into Low-resource Language Varieties.

Sachin Kumar, Antonios Anastasopoulos, Shuly Wintner and Yulia Tsvetkov.

In *Proceedings of ACL 2021*

Synthesizing Adversarial Negative Responses for Robust Response Ranking and Evaluation.

Prakhar Gupta, Yulia Tsvetkov, Jeffrey P. Bigham.

In *Proceedings of Findings of ACL 2021*

Understanding Factuality in Abstractive Summarization with FRANK: A Benchmark for Factuality Metrics.

Artidoro Pagnoni, Vidhisha Balachandran, and Yulia Tsvetkov.

In *Proceedings of NAACL 2021*

Controlling Dialogue Generation with Semantic Exemplars.

Prakhar Gupta, Jeffrey P. Bigham, Yulia Tsvetkov, and Amy Pavel.

In *Proceedings of NAACL 2021*

DialoGraph: Incorporating Interpretable Strategy-Graph Networks into Negotiation Dialogues.

Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov.

In *Proceedings of ICLR 2021*

Gradient Vaccine: Investigating and Improving Multi-task Optimization in Massively Multilingual Models. (Spotlight)

Zirui Wang, Yulia Tsvetkov, Orhan Firat, and Yuan Cao.

In *Proceedings of ICLR 2021*

StructSum: Incorporating Latent and Explicit Sentence Dependencies for Single Document Summarization.

Vidhisha Balachandran, Artidoro Pagnoni, Jay Yoon Lee, Dheeraj Rajagopal, Jaime Carbonell, and Yulia Tsvetkov.

In *Proceedings of EACL 2021*

Ranking Transfer Languages with Pragmatically-Motivated Features for Multilingual Sentiment Analysis.

Jimin Sun, Hwijee Ahn, Chan Young Park, Yulia Tsvetkov, and David R. Mortensen.

In *Proceedings of EACL 2021*

Multilingual Contextual Affective Analysis of LGBT People Portrayals in Wikipedia.

Chan Young Park, Xinru Yan, Anjalie Field and Yulia Tsvetkov.

In *Proceedings of ICWSM 2021*

An Exploration of Data Augmentation Techniques for Improving English to Tigrinya Translation.

Lidia Kidane, Sachin Kumar and Yulia Tsvetkov.

In *Proceedings of AfricaNLP 2021*

Unsupervised Discovery of Implicit Gender Bias.

Anjalie Field and Yulia Tsvetkov.

In *Proceedings of EMNLP 2020*

On Negative Interference in Multilingual Models: Findings and A Meta-Learning Treatment.

Zirui Wang, Zachary C. Lipton and Yulia Tsvetkov.

In *Proceedings of EMNLP 2020*

Fortifying Toxic Speech Detectors Against Veiled Toxicity.

Xiaochuang Han and Yulia Tsvetkov.

In *Proceedings of EMNLP 2020*

Automatic Extraction of Rules Governing Morphological Agreement.

Aditi Chaudhary, Antonios Anastasopoulos, Adithya Pratapa, David R. Mortensen, Zaid Sheikh, Yulia Tsvetkov and Graham Neubig.

In *Proceedings of EMNLP 2020*

Understanding Linguistic Accommodation in Code-Switched Human-Machine Dialogues.

Tanmay Parekh, Emily Ahn, Yulia Tsvetkov and Alan W Black.

In *Proceedings of CoNLL 2020*

LTiatCMU at SemEval-2020 Task 11: Incorporating Multi-Level Features for Multi-Granular Propaganda Span Identification.

Sopan Khosla, Rishabh Joshi, Ritam Dutt, Alan W. Black, and Yulia Tsvetkov.

In *Proceedings of SemEval 2020*

A Computational Analysis of Polarization on Indian and Pakistani Social Media. (Nominated for best paper award)

Aman Tyagi, Anjalie Field, Priyank Lathwal, Yulia Tsvetkov, and Kathleen M. Carley.

In *Proceedings of SocInfo 2020*

Demoting Racial Bias in Hate Speech Detection

Mengzhou Xia, Anjalie Field, Yulia Tsvetkov

In *Proceedings of SocialNLP 2020*

A Deep Reinforced Model for Cross-Lingual Summarization with Bilingual Semantic Similarity Rewards

Zi-Yi Dou, Sachin Kumar, Yulia Tsvetkov

In *Proceedings of WNGT 2020*

Balancing Training for Multilingual Neural Machine Translation

Xinyi Wang, Yulia Tsvetkov, and Graham Neubig

In *Proceedings of ACL 2020*

Balancing Training for Multilingual Neural Machine Explaining Black Box Predictions and Unveiling Data Artifacts through Influence Functions

Xiaochuang Han, Byron C. Wallace, and Yulia Tsvetkov

In *Proceedings of ACL 2020*

Stress and Burnout in Open Source: Toward Finding, Understanding, and Mitigating Unhealthy Interactions

Naveen Raman, Minxuan Cao, Yulia Tsvetkov, Christian Kästner, and Bogdan Vasilescu
In *Proceedings of International Conference on Software Engineering – New Ideas Track (ICSE-NIER) 2020*

Augmenting Non-Collaborative Dialog Systems with Explicit Semantic and Strategic Dialog History

Yiheng Zhou, Yulia Tsvetkov, Alan W Black, and Zhou Yu
In *Proceedings of ICLR 2020*

What Code-Switching Strategies are Effective in Dialog Systems?

Emily Ahn, Cecilia Jimenez, Yulia Tsvetkov, and Alan W Black
In *Proceedings of SCiL 2020*

Where New Words Are Born: Distributional Semantic Analysis of Neologisms and Their Semantic Neighborhoods

Maria Ryskina, Ella Rabinovich, Taylor Berg-Kirkpatrick, David R. Mortensen and Yulia Tsvetkov
In *Proceedings of SCiL 2020*

Topics to Avoid: Demoting Latent Confounds in Text Classification

Sachin Kumar, Shuly Wintner, Noah A. Smith, and Yulia Tsvetkov
In *Proceedings of EMNLP 2019*

Finding Microaggressions in the Wild: A Case for Locating Elusive Phenomena in Social Media Posts

Luke M. Breitfeller, Emily Ahn, David Jurgens, and Yulia Tsvetkov.
In *Proceedings of EMNLP 2019*

Learning to Generate Word- and Phrase-Embeddings for Efficient Phrase-Based Neural Machine Translation

Chan Young Park and Yulia Tsvetkov
In *Proceedings of WNGT 2019*

A Margin-based Loss with Synthetic Negative Samples for Continuous-output Machine Translation

Gayatri Bhat, Sachin Kumar and Yulia Tsvetkov
In *Proceedings of WNGT 2019*

A Dynamic Strategy Coach for Effective Negotiation

Yiheng Zhou, He He, Alan W Black and Yulia Tsvetkov
In *Proceedings of SIGdial 2019*

Entity-Centric Contextual Affective Analysis

Anjalie Field and Yulia Tsvetkov
In *Proceedings of ACL 2019*

CMU-01 at the SIGMORPHON 2019 Shared Task on Crosslinguality and Context in Morphology

Aditi Chaudhary, Elizabeth Salesky, Gayatri Bhat, David R. Mortensen, James G. Carbonell, and Yulia Tsvetkov
In *Proceedings of SIGMORPHON 2019* (Interpretability prize)

Quantifying Social Biases in Contextual Word Representations.

Keita Kurita, Nidhi Vyas, Ayush Pareek, Alan W Black, and Yulia Tsvetkov
In *Workshop on Gender Bias for NLP 2019*

Contextual Affective Analysis: A Case Study of People Portrayals in Online #MeToo Stories

Anjalie Field, Gayatri Bhat and Yulia Tsvetkov
In *Proceedings of ICWSM 2019*

Black is to Criminal as Caucasian is to Police: Detecting and Removing Multiclass Bias in Word Embeddings

Thomas Manzini, Yao Chong, Yulia Tsvetkov and Alan W Black
In *Proceedings of NAACL 2019*

Von Mises-Fisher Loss for Training Sequence to Sequence Models with Continuous Outputs

Sachin Kumar and Yulia Tsvetkov
In *Proceedings of ICLR 2019*

Framing and Agenda-setting in Russian News: a Computational Analysis of Intricate Political Strategies

Anjalie Field, Doron Kliger, Shuly Wintner, Jennifer Pan, Dan Jurafsky, and Yulia Tsvetkov
In *Proceedings of EMNLP 2018*.

Machine Translation with Continuous Outputs

Sachin Kumar and Yulia Tsvetkov
In *Proceedings of the ICML Workshop on Theoretical Foundations and Applications of Deep Generative Models 2018*.

Style Transfer Through Back-Translation

Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, and Alan W Black
In *Proceedings of ACL 2018*.

RtGender: A Corpus of Responses to Gender for Studying Gender Bias

Rob Voigt, David Jurgens, Vinodkumar Prabhakaran, Dan Jurafsky, and Yulia Tsvetkov
In *Proceedings of LREC 2018*.

Incorporating Dialectal Variability for Socially Equitable Language Identification

David Jurgens, Yulia Tsvetkov and Dan Jurafsky
In *Proceedings of ACL 2017*. PDF

Writer Profiling Without the Writer's Text

David Jurgens, Yulia Tsvetkov, and Dan Jurafsky
In *Proceedings of SocInfo 2017*. PDF

Learning the Curriculum with Bayesian Optimization for Task-Specific Word Representation Learning

Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Brian MacWhinney, and Chris Dyer
In *Proceedings of ACL 2016*. PDF

Correlation-based Intrinsic Evaluation of Word Vector Representations

Yulia Tsvetkov, Manaal Faruqui and Chris Dyer
In *RepEval 2016*. PDF

Problems With Evaluation of Word Embeddings Using Word Similarity Tasks

Manaal Faruqui, Yulia Tsvetkov, Pushpendre Rastogi, and Chris Dyer
In *RepEval 2016*. PDF

Polyglot Neural Language Models:

Case Study in Cross-Lingual Phonetic Representation Learning

Yulia Tsvetkov, Sunayana Sitaram, Manaal Faruqui, Guillaume Lample, Patrick Littell, David Mortensen, Alan W Black, Lori Levin, and Chris Dyer

In *Proceedings of NAACL 2016*. PDF

Morphological Inflection Generation Using Character Sequence to Sequence Learning

Manaal Faruqui, Yulia Tsvetkov, Graham Neubig, and Chris Dyer.

In *Proceedings of NAACL 2016*. PDF

Evaluation of Word Vector Representations by Subspace Alignment

Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Guillaume Lample, and Chris Dyer

In *Proceedings of EMNLP 2015*. PDF

Not All Contexts Are Created Equal: Better Word Representations with Variable Attention

Wang Ling, Chu-Cheng Lin, Yulia Tsvetkov, Silvio Amir, Ramón Fernández Astudillo, Chris Dyer, Alan W Black, and Isabel Trancoso

In *Proceedings of EMNLP 2015*. PDF

Lexicon Stratification for Translating Out-of-Vocabulary Words

Yulia Tsvetkov and Chris Dyer

In *Proceedings of ACL 2015*. PDF

Sparse Overcomplete Word Vector Representations

Manaal Faruqui, Yulia Tsvetkov, Dani Yogatama, Chris Dyer, and Noah Smith

In *Proceedings of ACL 2015*. PDF

Constraint-Based Models of Lexical Borrowing

Yulia Tsvetkov, Waleed Ammar, and Chris Dyer

In *Proceedings of NAACL 2015*. PDF

A Bottom Up Approach to Category Mapping and Meaning Change

Haim Dubossarsky, Yulia Tsvetkov, Chris Dyer, and Eitan Grossman

In *Proceedings of NetWordS 2015*. PDF

Metaphor Detection with Cross-Lingual Model Transfer

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer

In *Proceedings of ACL 2014*. PDF

Augmenting Translation Models with Simulated Acoustic Confusions for Improved Spoken Language Translation

Yulia Tsvetkov, Florian Metze, and Chris Dyer

In *Proceedings of EACL 2014*. PDF

Augmenting English Adjective Senses with Supersenses

Yulia Tsvetkov, Nathan Schneider, Dirk Hovy, Archana Bhatia, Manaal Faruqui, and Chris Dyer

In *Proceedings of LREC 2014*. PDF

A Unified Annotation Scheme for the Semantic/Pragmatic Components of Definiteness

Archana Bhatia, Mandy Simons, Lori Levin, Yulia Tsvetkov, Chris Dyer, and Jordan Bender

In *Proceedings of LREC 2014*. PDF

Automatic Classification of Communicative Functions of Definiteness

Archana Bhatia, Chu-Cheng Lin, Nathan Schneider, Yulia Tsvetkov, Fatima Talib Al-Raisi, Laleh Roostapour, Jordan Bender, Abhimanu Kumar, Lori Levin, Mandy Simons, and Chris Dyer
Proceedings of COLING 2014. PDF

The CMU Machine Translation Systems at WMT 2014

Austin Matthews, Waleed Ammar, Archana Bhatia, Weston Feely, Greg Hanneman, Eva Schlinger, Swabha Swayamdipta, Yulia Tsvetkov, Alon Lavie, and Chris Dyer
Proceedings of WMT 2014. PDF

Generating English Determiners in Phrase-Based Translation with Synthetic Translation Options

Yulia Tsvetkov, Chris Dyer, Lori Levin, and Archana Bhatia.
In *Proceedings of WMT 2013*. PDF.

The CMU Machine Translation Systems at WMT 2013: Syntax, Synthetic Translation Options, and Pseudo-References

Waleed Ammar, Victor Chahuneau, Michael Denkowski, Greg Hanneman, Wang Ling, Austin Matthews, Kenton Murray, Nicola Segall, Yulia Tsvetkov, Alon Lavie, and Chris Dyer
In *Proceedings of WMT 2013*. PDF.

Identifying the L1 of non-native writers: the CMU-Haifa system

Yulia Tsvetkov, Naama Twitto, Nathan Schneider, Noam Ordan, Manaal Faruqui, Victor Chahuneau, Shuly Wintner, and Chris Dyer
In *Proceedings of the 8th BEA Workshop, 2013*. PDF.

Cross-Lingual Metaphor Detection Using Common Semantic Features

Yulia Tsvetkov, Elena Mukomel, and Anatole Gershman
In *Proceedings of the 1st Meta4NLP Workshop, 2013*. PDF.

Identification and Modeling of Word Fragments in Spontaneous Speech

Yulia Tsvetkov, Zaid Sheikh, and Florian Metze
In *Proceedings of ICASSP, 2013*. PDF.

Identification of Multi-word Expressions by Combining Multiple Linguistic Information Sources

Yulia Tsvetkov and Shuly Wintner
In *Proceedings of EMNLP, 2011*. PDF.

Extraction of Multi-word Expressions from Small Parallel Corpora

Yulia Tsvetkov and Shuly Wintner
In *Proceedings of COLING 2010*. PDF.

Automatic Acquisition of Parallel Corpora from Websites with Dynamic Content

Yulia Tsvetkov and Shuly Wintner
In *Proceedings of LREC 2010*. PDF.

PATENTS

SEQUENCE GENERATION USING NEURAL NETWORKS WITH CONTINUOUS OUTPUTS

Sachin Kumar and Yulia Tsvetkov *U.S. Patent 16/509,930, issued July 12, 2019*.

ADVISING

Lucille Njoo, PhD at UW

- Computational ethics in NLP.

2021–

Orevaoghene Ahia , PhD at UW, co-advised with Noah A. Smith	2021–
<ul style="list-style-type: none"> • Multilingual NLP. 	
Melanie Sclar , PhD at UW, co-advised with Yejin Choi	2021–
<ul style="list-style-type: none"> • Machine learning, misinformation. 	
Xiaochuang Han , MLT at CMU → PhD at UW	2019–
<ul style="list-style-type: none"> • Machine learning, interpretability, computational ethics. 	
Artidoro Pagnoni , MLT at CMU → PhD at UW	2019–
<ul style="list-style-type: none"> • Natural language processing for detecting and mitigating misinformation. 	
Vidhisha Balachandran , PhD, co-advised with Jaime Carbonell at CMU	2019–
<ul style="list-style-type: none"> • Language generation, interpretability, computational ethics. 	
Chan Young Park , PhD at CMU	2018–
<ul style="list-style-type: none"> • Multilinguality, sentiment analysis, computational ethics. 	
Sachin Kumar , PhD at CMU	2017–
<ul style="list-style-type: none"> • Machine learning, language generation, multilinguality. 	
Anjalie Field , PhD at CMU	2017–
<ul style="list-style-type: none"> • Natural language processing, computational social science, computational ethics. 	
Zirui Wang , PhD, advised by Jaime Carbonell; from 2020 co-advised with Emma Strubell at CMU	2017–2021
<ul style="list-style-type: none"> • Multilingual machine learning, machine translation. • Research Engineer at Google Brain 	
Alissa Ostapenko , MLT at CMU	2020–2022
<ul style="list-style-type: none"> • Multilinguality and code-switching 	
Rishabh Joshi , MLT, co-advised with Alan W Black at CMU	2019–2021
<ul style="list-style-type: none"> • Goal-oriented dialogue modeling, information extraction. • Software Engineer at Google Brain 	
Tanmay Parekh , MLT, co-advised with Alan W Black at CMU	2019–2021
<ul style="list-style-type: none"> • Multilinguality and code-switching • PhD student at UCLA 	
Emily Ahn , MLT, co-advised with Alan W Black at CMU	2017–2019
<ul style="list-style-type: none"> • Multilinguality and code-switching • PhD student at UW Linguistics 	
Yiheng Zhou , MLT, co-advised with Alan W Black at CMU	2017–2019
<ul style="list-style-type: none"> • Goal-oriented dialogue modeling • Research Engineer at Amazon Alexa 	
Gayatri Bhat , MLT at CMU	2017–2020
<ul style="list-style-type: none"> • Distributional semantics, multilingual and low-resource NLP. • Research Engineer at Bloomberg AI 	

TEACHING

Natural Language Processing (undergraduate) <i>Course Instructor</i> University of Washington	Spring 2022
Computational Ethics in NLP (graduate) <i>Course Instructor</i> University of Washington	Winter 2022
Computational Ethics Lab <i>Course Instructor</i> Carnegie Mellon University	Spring 2021
Multilingual NLP (graduate) <i>Course Instructor (with Graham Neubig and Alan W Black)</i> Carnegie Mellon University	Fall 2020
Algorithms for NLP (graduate) <i>Course Instructor (with Emma Strubell and Robert Frederking)</i> Carnegie Mellon University	Fall 2020
Computational Ethics for NLP (graduate) <i>Course Instructor (with Alan W Black)</i> Carnegie Mellon University	Spring 2020
Natural Language Processing (undergraduate IITP course) <i>Course Instructor (with David Mortensen)</i> Carnegie Mellon University	Spring 2020
Algorithms for NLP (graduate) <i>Course Instructor (with Robert Frederking)</i> Carnegie Mellon University	Fall 2019
Natural Language Processing (undergraduate IITP course) <i>Course Instructor (with David Mortensen)</i> Carnegie Mellon University	Fall 2019
Computational Ethics Lab (graduate) <i>Course Instructor</i> Carnegie Mellon University	Fall 2019
Computational Ethics for NLP (graduate) <i>Course Instructor (with Alan W Black)</i> Carnegie Mellon University	Spring 2019

Algorithms for NLP (graduate) **Fall 2018**
Course Instructor (with Robert Frederking)
Carnegie Mellon University

NLP Lab **Fall 2018**
Course Instructor
Carnegie Mellon University

Computational Ethics for NLP (graduate) **Spring 2018**
Course Instructor (with Alan W Black)
Carnegie Mellon University

Speech recognition and understanding (graduate) **Fall 2014**
Teaching Assistant
Carnegie Mellon University

Machine translation (graduate) **Spring 2014**
Teaching Assistant
Carnegie Mellon University

Object Oriented Programming **Spring 2010**
Lectures on programming in Java
University of Haifa

Introduction to Computers - C language **2006**
Teaching Assistant
Technion

INNOVATIONS
IN
TEACHING

Computational Ethics in Natural Language Processing
Developed a graduate-level course focusing on real-world applications of language technologies and the potential ethical implications associated with them http://demo.clab.cs.cmu.edu/ethical_nlp, <https://courses.cs.washington.edu/courses/cse599d1/22wi>.

Multilingual Natural Language Processing
Co-developed a graduate-level course focusing on linguistically motivated solutions to core and applied NLP tasks for any language <http://demo.clab.cs.cmu.edu/11737fa20/>, video lectures [URL].

TEACHING
BEYOND
UNIVERSITY

Tutorial on Mitigating Societal Harms of Large Language Models
TheWebConf 2022 <https://sites.google.com/view/societal-impacts-lm-tutorial>

Tutorial on Methods in Multilingual Natural Language Processing
Multilingual NLP summer residency program for girls at CMU, 2020.

Tutorial “Language Technologies for People’s Good”
AI4ALL summer residency program at CMU, 2019.

Tutorial on Socially Responsible Natural Language Processing
WWW 2019. <https://sites.google.com/view/srnlp>

Tutorial on Socially Responsible Natural Language Processing
NAACL 2018

Tutorial on Methods in Low-Resource Natural Language Processing
2017 JSALT Summer School on Human Language Technologies PART1 PART2.

KEYNOTE
TALKS

Controllable Text Generation with Multiple Constraints

Talk at the CtrlGen Workshop at NeurIPS: Controllable Generative Modeling in Language and Vision, December 2021

Multilingual Controllable Text Generation

Talk at the Workshop on Multilingual Representation Learning (MRL Workshop at EMNLP), November 2021

Towards Interpretability Without an Overhead

Talk at the Second Workshop on Simple and Efficient Natural Language Processing (SustainNLP Workshop at EMNLP), November 2021

Towards Language Generation We Can Trust

Talk at the ACL 2021 Workshop on NLP for Positive Impact, August 2021

Proactive NLP: How to prevent social and ethical Problems in NLP systems?

ICML 2021 Workshop on Socially Responsible Machine Learning , July 2021

Proactive NLP: How to prevent social and ethical Problems in NLP systems?

1st ELLIS NLP Workshop, February 2021

Towards Speaker-Centered NLP with Finer-Grained Multilingual Model Proficiency

The Second Workshop on Typology for Polyglot NLP (TyP-NLP), November 2020

Proactively Reducing Social Bias in Conversational AI

Safety in Conversational AI Workshop, October 2020

Responsible Language Technologies

Responsible Machine Learning Summit, UC Santa Barbara, October 2019

Towards Personalized and Adaptive NLP: Modeling Output Spaces in Continuous-Output Language Generation

The 4th Workshop on Representation Learning for NLP (RepL4NLP). July, 2019.

Continuous-Output Language Generation

ICLR workshop Deep Generative Models for Highly Structured Data. May, 2019.

Framing and Agenda-setting in News: a Computational Analysis of Intricate Political Strategies

Credibility on the Web Workshop, Google. October, 2018.

Language Generation with Continuous Outputs

WeCNLP Summit, Facebook. September, 2018.

Towards Flexible but Controllable Language Generation

The Second Workshop on Neural Machine Translation and Generation, ACL. July, 2018.

Towards Flexible but Controllable Language Generation

Plenary Lecture at JSALT. July, 2018.

Linguistic Knowledge in a Data-driven World

SemEval 2016. SLIDES.

INVITED
TALKS

Controlled Analyses of Social Biases in Wikipedia Bios

Invited Talk at Wikimedia Research. May, 2022

Towards Language Generation We Can Trust

Invited Talk at Rochester Institute of Technology. February, 2022

Towards Language Generation We Can Trust

PNNL Mega-AI Seminar Speaker Series. February, 2022

Towards Language Generation We Can Trust

Amazon NLP Talk Series. September, 2021

Proactive NLP: How to prevent social and ethical Problems in NLP systems?

Georgia Tech. April, 2021

Proactive NLP: How to prevent social and ethical Problems in NLP systems?

Google Research, March 2021

Proactive NLP: How to prevent social and ethical Problems in NLP systems?

Hebrew University, March 2021

Ethical Language Technologies

Magee-Womens Research Institute & Foundation, September 2020

NLP for ALL

Talk at UW. March, 2020.

Towards Human-Centered Language Technologies

Talk at UCLA. January, 2020.

Towards Multilingual Personalized Language Generation

Invited talk at Duolingo. May, 2019.

Computational Approaches to Unveiling Biases in Stories and Models

Invited talk at Stanford. May, 2019.

Towards a Computational Analysis of the Language of Veiled Manipulation

Invited talk at University of Pennsylvania. April, 2019. **Towards a Computational Analysis of the Language of Veiled Manipulation**

Invited talk at New York University. April, 2019.

Language Technologies for Social Good (or Evil)

Invited talk at Carnegie Mellon University, Social and Decision Sciences department. January, 2019.

Towards Flexible but Controllable Language Generation

Invited talks at ISI, UCLA, UC Davis. June, 2018.

Linguistic Knowledge in a Data-Driven World

Invited talk at University of Massachusetts. March, 2018.

Leveraging Human Knowledge for Better Statistical Generalization
University of Pittsburgh. February, 2018.

Linguistic Knowledge in Data-Driven NLP

Invited talk at Moscow Institute of Physics and Technology. February, 2018.

On the Synergy of Linguistics and Machine Learning in Natural Language Processing

Invited talk at Georgetown University. October, 2016.

Linguistic Knowledge in a Data-driven World

Invited talk at Stanford University. September, 2016.

On the Synergy of Linguistics and Machine Learning in Natural Language Processing

Job talk at Carnegie Mellon University University. March, 2016. SLIDES.

Cross-Lingual Bridges with Models of Lexical Borrowing

Talk at Army Research Office. November 2015. SLIDES.

Learning an Optimality Theoretic Model of Lexical Borrowing

Student Research Symposium, Carnegie Mellon University, LTI. August 2014.

Best Poster Award POSTER.

Lexical-Semantic Features in Cross-Lingual Applications

Talk at Army Research Office, December 2013.

Identification and Modeling of Word Fragments in Conversational Speech

Student Research Symposium, Carnegie Mellon University, LTI. August 2012.

Identification of Multi-word Expressions by Combining Multiple Linguistic Information sources

Invited talk at Workshop on Machine Translation and Morphologically Rich Languages, University of Haifa. January 2011. SLIDES.

SERVICE

Co-Chair of the ACL Exec Ethics Committee

2021 – present

Program / Area chair

- ICLR 2022
- NAACL 2022 – Tutorials Chair
- EMNLP 2021 – Machine Learning for NLP
- ACL 2021 – Theme Track: NLP for Social Good (NLP4SG)
- EMNLP 2020 – Machine Learning for NLP
- ACL 2020 – Ethics
- CoNLL 2020
- EMNLP-IJCNLP 2019 – Social Media and Computational Social Science
- NAACL-HLT 2019 – Multilingualism, Cross lingual resources
- NAACL-HLT 2018 – Machine Learning for NLP, Semantics
- CoNLL 2018 – Semantics

Program chair/Reviewer

- NSF panelist (CISE/IIS, CISE/SBE GRF/CRII/EAGER/Small/Medium/DEL) 2018, 2019, 2020, 2021, 2022
- ISF reviewer 2020 (Israeli Science Foundation)
- TACL – *Transactions of the Association for Computational Linguistics*
- JAIR – *Journal of Artificial Intelligence Research*
- TASLP – *IEEE/ACM Transactions on Audio Speech and Language Processing*
- LREV – *Language Resources and Evaluation*
- ACL/AAACL/EMNLP/COLING Tutorials reviewer 2020
- ACL 2015, 2016, 2017, 2018, 2019
- AAACL 2020
- EMNLP 2015, 2016, 2017, 2018
- NAACL HLT 2015, 2016, 2019, 2021
- NeurIPS 2021 (Ethics committee)
- ICML 2019
- ICLR 2019
- COLING 2014, 2016, 2018, 2020
- EACL 2014, 2021
- IJCNLP 2017
- LREC 2018
- *SEM 2013, 2017, 2018
- WMT 2014, 2015, 2016, 2018 – Workshop on Statistical Machine Translation
- EthNLP 2017, 2018 – The Workshop on Ethics in NLP
- Fig-Lang18 2018, 2020 – Workshop on Figurative Language Processing
- MWE 2013, 2014, 2015, 2017 – The Workshop on Multiword Expressions
- Meta4NLP 2015, 2016 – The Workshop on Metaphor in NLP
- RELNLP 2018 – Workshop on Relevance of Linguistic Structure in Neural Architectures for NLP
- RepEval 2016, 2017, 2019 – The Workshop on Evaluating Vector Space Representations for NLP
- RepL4NLP 2017 – Workshop on Representation Learning for NLP
- The LAW 2014, 2015 – The Linguistic Annotation Workshop
- TyP-NLP 2019 – Typology for Polyglot NLP
- WiNLP 2017, 2018 – Women and Underrepresented Minorities in NLP Workshop
- NLP4IF 2019 – Workshop on Natural Language Processing for Internet Freedom
- WNGT 2019, 2020 – Workshop on Neural Generation and Translation
- Wiki Workshop 2022 – The Web Conference 2022 forum bringing together researchers exploring all aspects of Wikimedia projects

Workshop Organizer

- MLCL 2016 – Workshop on Multilingual and Cross-lingual Methods in NLP. Co-organized with Dipanjan Das, Chris Dyer, Manaal Faruqui
- SCLeM 2018 – Second Workshop on Subword and Character LEvel Models in NLP. Co-organized with Manaal Faruqui, Hinrich Schütze, Isabel Trancoso, Yadollah Yaghoobzadeh

- SafeConvAI SIGDial 2021: A Special Session on Safety for E2E Conversational AI. Co-organized with Emily Dinan, Y-Lan Boureau, Gavin Abercrombie, Saadia Gabriel, Alborz Geramifard, Verena Rieser, Zeerak Waseem
- mmmPIE 2022 – Workshop on Performance and Interpretability Evaluations of Multimodal, Multipurpose, Massive-Scale Models. Co-organized with Maria F Glenski, Megan Kohagen, and Vidhisha Balachandran

SOFTWARE
AND
DATA

FRANK – A benchmark for factuality metrics

CONGEN — a toolkit for training sequence-to-sequence models with continuous outputs.

THE REDDIT-L2 CORPUS – a large corpus (over 250M sentences) of highly-advanced, fluent, diverse, non-native English, with sentence-level annotations of the native language of each author.

RTGENDER – a corpus of more than 25M comments on social media for studying gender bias and responses to gender online, including posts and responses from from five socially and topically diverse sources tagged for the gender of the addressee.

QVEC – Intrinsic evaluation of word vector representations.

THE CMU CROSS-LINGUAL METAPHOR DETECTOR – a toolkit for identifying instances of figurative language in English and any other language for which a bilingual dictionary is available.

STR2IPA – IPA pronunciation dictionaries for languages with close-to-phonetic writing systems.

ENGLISH ADJECTIVE SUPERSENSSES – A 13-class supersense taxonomy of English adjectives and SUPERSENSE CLASSIFIER.

INDUSTRY
EXPERIENCE

Marvell Semiconductors Israel
Software Engineer

2003–2006

MISCELLANEOUS

Languages: Russian(native), English(fluent), Hebrew(fluent), and Ukrainian(fluent).