

Yulia Tsvetkov

CONTACT INFORMATION

5516 Woodmont St
Carnegie Mellon University
Pittsburgh, PA 15217
USA

phone: +1-412-204-7407
office: Gates-Hillman Complex, 6405
email: ytsvetko@cs.cmu.edu
web: <http://www.cs.cmu.edu/~ytsvetko>

RESEARCH INTERESTS

My research is motivated by a unified goal: to extend the capabilities of human language technology beyond individual cultures and across language boundaries, thereby enabling NLP for disadvantaged groups, the users that need it most. Research projects in my group currently focus on language generation (e.g., controllable text generation, continuous-output generation, GANs for text, structured summarization, evaluation of summarization), multilinguality (e.g., open-vocabulary machine translation, polyglot models, adaptive multilingual dialogue agents), automated negotiation, and NLP for social good (e.g., identification of microaggressions and dehumanization in online interactions, identification of misinformation and agenda-setting in news, preventing scientific misconduct).

PROFESSIONAL EXPERIENCE

Carnegie Mellon University

Language Technologies Institute, School of Computer Science

Assistant Professor

2017–now

Stanford University

Department of Computer Science

Postdoctoral research associate, advised by Dan Jurafsky.

2016–2017

- Research topics: computational social science and NLP for social good, including modeling gender bias and debasing, political framing and agenda setting.

EDUCATION

Carnegie Mellon University

Language Technologies Institute, School of Computer Science

Ph.D., advised by Chris Dyer.

2011–2016

- Research topics: Modeling lexical borrowing using computational phonology and morphology; representation learning and evaluation; machine translation in low-resource settings; integrating linguistic information in statistical systems; cross-lingual semantics; automatic taxonomy construction; nonnative writing; speech translation; disfluencies in speech recognition; pronunciation and language modeling for speech.
- Thesis: “Linguistic Knowledge in Data-Driven Natural Language Processing” PDF, SLIDES.

University of Haifa, Israel

M.Sc. in Computer Science *Cum Laude*, advised by Shuly Wintner.

2007–2010

- Research topics: semi-supervised extraction of multiword expression from parallel and monolingual corpora.
- 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

Title: Multi-media Analytics Leading to Intent and Semantic Evidence (MALISE)
Co-PIs: SRI International
Source of Support: DARPA SemaFor program
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
Source of Support: National Science Foundation
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: PIT-UN Challenge
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, Nathan Schneider, Shuly Wintner
Source of Support: National Science Foundation
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
Period of Performance: 9/1/2016 to 8/31/2019

Title: Language Generation into Embedding Spaces
Source of Support: Google Faculty Research Award
April 2020

Title: Language Generation Without Softmax Layer: Open-Vocabulary, Accurate, Efficient and Compact
Source of Support: Amazon Machine Learning Research Award
September 2019

Okawa Foundation Young Researcher Award
April 2019

Adobe Researcher Gift
May 2019

JOURNAL
ARTICLES

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

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* 5

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

Anjalie Field, PhD at CMU **2017–**

- Natural language processing, computational social science, ethics.

Sachin Kumar, PhD at CMU **2017–**

- Machine learning, language generation.

Chan Young Park, PhD at CMU **2018–**

- Language generation, ethics.

Vidhisha Balachandran, PhD, co-advised with Jaime Carbonell at CMU **2019–**

- Language generation, structured prediction.

Artidoro Pagnoni, MLT at CMU **2019–**

- Natural language processing, summarization.

Xiaochuang Han, MLT at CMU **2019–**

- Natural language processing, interpretability of deep learning models.

Rishabh Joshi, MLT, co-advised with Alan W Black at CMU **2019–**

- Goal-oriented dialogue modeling, automated negotiation.

Tanmey Parekh, MLT, co-advised with Alan W Black at CMU **2019–**

- Multilinguality and code-switching

Emily Ahn, MLT, co-advised with Alan W Black at CMU **2017–2019**

- Multilinguality and code-switching
- PhD student at UW Linguistics

Yiheng Zhou, MLT, co-advised with Alan W Black at CMU **2017–2019**

- Goal-oriented dialogue modeling
- Research Engineer at Amazon Alexa

Gayatri Bhat, MLT at CMU **2017–2020**

- Distributional semantics, multilingual and low-resource NLP.
- Research Engineer at Bloomberg AI

TEACHING

Computational Ethics for NLP **Spring 2020**

Course Instructor (with Alan W Black)
Carnegie Mellon University

Natural Language Processing (undergraduate IITP course) **Spring 2020**

Course Instructor (with David Mortensen)
Carnegie Mellon University

Algorithms for NLP (graduate course) **Fall 2019**

Course Instructor (with Robert Frederking)
Carnegie Mellon University

Natural Language Processing (undergraduate IITP course) **Fall 2019**

Course Instructor (with David Mortensen)
Carnegie Mellon University

Computational Ethics Lab **Fall 2019**

Course Instructor
Carnegie Mellon University

Computational Ethics for NLP **Spring 2019**

Course Instructor (with Alan W Black)
Carnegie Mellon University

Algorithms for NLP **Fall 2018**

Course Instructor (with Robert Frederking)
Carnegie Mellon University

NLP Lab **Fall 2018**

Course Instructor
Carnegie Mellon University

Computational Ethics for NLP **Spring 2018**

Course Instructor (with Alan W Black)
Carnegie Mellon University

Speech recognition and understanding **Fall 2014**

Teaching Assistant

Carnegie Mellon University

Machine translation

Teaching Assistant

Carnegie Mellon University

Spring 2014

Object Oriented Programming

Lectures on programming in Java

University of Haifa

Spring 2010

Introduction to Computers - C language

Teaching Assistant

Technion

2006

INNOVATIONS
IN
TEACHING

Computational Ethics for Natural Language Processing

Developed a graduate-level course at CMU focusing on real-world applications of language technologies and the potential ethical implications associated with them http://demo.c1ab.cs.cmu.edu/ethical_nlp.

TEACHING
BEYOND
UNIVERSITY

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

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

NLP4ALL

Invited talk at UW. March, 2020.

Towards Human-Centered Language Technologies

Invited 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

Area chair

- 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 2018, 2019, 2020
- ISF reviewer 2020 (Israeli Science Foundation)
- 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
- EMNLP 2015, 2016, 2017, 2018
- NAACL HLT 2015, 2016, 2019
- ICML 2019
- ICLR 2019
- COLING 2014, 2016, 2018
- EACL 2014
- 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

Workshop Organizer

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

SOFTWARE
AND
DATA

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).
Citizenships: US, Israeli