

Yulia Tsvetkov

CONTACT INFORMATION	5516 Woodmont St Pittsburgh, PA 15217 USA	<i>phone:</i> +1-412-204-7407 <i>email:</i> ytsvetko@cs.cmu.edu <i>web:</i> http://www.cs.cmu.edu/~ytsvetko
RESEARCH INTERESTS	Multilinguality and low-resource NLP, NLP for social good, computational phonology and morphology, lexical and distributional semantics, representation learning, text and speech translation, machine learning.	
PROFESSIONAL EXPERIENCE	Carnegie Mellon University Language Technologies Institute, School of Computer Science Assistant Professor	2017–
	Stanford University Department of Computer Science Postdoctoral research associate, advised by Dan Jurafsky. <ul style="list-style-type: none">• Research topics: computational social science and NLP for social good, including modeling gender bias and debasing, political framing and agenda setting.	2016–2017
EDUCATION	Carnegie Mellon University Language Technologies Institute, School of Computer Science Ph.D., advised by Chris Dyer. <ul style="list-style-type: none">• 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.	2011–2016
	University of Haifa, Israel M.Sc. in Computer Science <i>Cum Laude</i> , advised by Shuly Wintner. <ul style="list-style-type: none">• 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.	2007–2010
	Technion - Israel Institute of Technology B.Sc. in Computer Science and Education.	2000–2005
JOURNAL ARTICLES	Cross-Lingual Bridges with Models of Lexical Borrowing Yulia Tsvetkov and Chris Dyer <i>Journal of Artificial Intelligence Research</i> , 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

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, Ramn Fernandez 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.

TEACHING

Computational Ethics for NLP

Spring 2018

Course Instructor (with Alan W Black)

Carnegie Mellon University

SYLLABUS

Speech recognition and understanding

Fall 2014

Teaching Assistant

Carnegie Mellon University

Machine translation

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

ADVISING

Gayatri Bhat, PhD at Carnegie Mellon University

2017–

- Distributional semantics, multilingual and low-resource NLP.

Anjalie Field, PhD at Carnegie Mellon University

2017–

- Computational social science.

Sachin Kumar, PhD at Carnegie Mellon University

2017–

- Machine learning, generation.

Emily Ahn, MLT, co-advised with Alan W Black at Carnegie Mellon University

2017–

- Multilinguality and code-switching.

Sai Bhaskar, MLT, co-advised with Carolyn Rose at Carnegie Mellon University **2017–**
• Computational social science.

Yiheng Zhou, MLT, co-advised with Alan W Black at Carnegie Mellon University **2017–**
• Dialogue modeling.

TALKS

Opportunities and Challenges in Working with Low-Resource Languages
Tutorial at JSALT 2017 Summer School. June 2017. PART1 PART2.

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.

Linguistic Knowledge in a Data-driven World
Keynote talk at SemEval 2016. SLIDES.

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

Journal reviewing: JAIR, ACM TSLP, LREV.
Conference and Workshop reviewing: ACL, NAACL, EMNLP, WWW, EACL, COLING, *SEM, LREC, WMT, MWE, Meta4NLP, LAW, RepEval.
Co-Organizing a Workshop on Multilingual and Cross-lingual Methods in NLP at NAACL 2016 (with Dipanjan Das, Chris Dyer, and Manaal Faruqui).
Co-Organizing a Tutorial on Methods in Low-resource NLP at JSALT 2017 Summer School (with David Mortensen).

SOFTWARE AND DATA

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 supersenses - 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