

# Manaal Faruqui

---

LAST UPDATED 01/2017

CONTACT INFORMATION Google  
111 8th Avenue *E-mail: mfaruqui@google.com*  
New York 11201, NY USA *WWW: www.manaalfaruqui.com*

RESEARCH INTERESTS Representation learning, distributional semantics, multilingual learning, morphology, natural language processing, deep learning and machine learning.

EXPERIENCE **Google**, New York City, NY, USA **Aug 2016 – present**  
Research Scientist

EDUCATION **Carnegie Mellon University**, Pittsburgh, PA, USA  
Ph.D., Language and Information Technology, 2016

- Research area: Representation learning, distributional semantics, multilingual learning.
- Advisor: Prof. Chris Dyer

**Indian Institute of Technology**, Kharagpur, India  
B.Tech. (Hons.), Computer Science, May, 2012  
M.Tech., Computer Science, May, 2012

HONORS AND AWARDS

- Best Student Paper Award at NAACL 2015
- Selected to attend Heidelberg Laureate Forum 2015
- Best Reviewer Award at NAACL 2015
- Best Presentation Award at LTI Students' Research Symposium 2012
- Microsoft Research India Travel Grant to attend ACL-2011
- DAAD Working Internship in Science & Engineering scholarship 2010
- MITACS Globalink Research Internship Grant 2009

INTERNSHIPS **Google Research**, London, UK **May – Aug, 2015**  
Supervisor: Dr. Ryan McDonald  
Multilingual Morpho-Syntactic Lexicon Generation

**Google Research**, New York, USA **June – Aug, 2014**  
Supervisor: Dr. Shankar Kumar  
Multilingual Relation Extraction

**Yahoo! labs**, Bangalore, India **May – July, 2011**  
Supervisor: Dr. Narayan Bhamidipati  
Personalized Advertisement Search

**University of Stuttgart**, Stuttgart, Germany **May – July, 2010**  
Supervisor: Prof. Sebastian Padó  
German Named Entity Recognition, Textual Entailment

**Simon Fraser University**, Vancouver, Canada **May – July, 2009**  
Supervisor: Prof. Anoop Sarkar  
Model Adaptation in Statistical Machine Translation

JOURNAL  
PUBLICATIONS

Morpho-syntactic Lexicon Generation Using Graph-based Semi-supervised Learning.  
**Manaal Faruqui**, Ryan McDonald, and Radu Soricut.  
Transactions of the ACL (TAACL) 2016.

CONFERENCE  
PUBLICATIONS

Morphological Inflection Generation Using Character Sequence to Sequence Learning.  
**Manaal Faruqui**, Yulia Tsvetkov, Graham Neubig, and Chris Dyer.  
Proceedings of NAACL 2016 (long).

Polyglot Neural Language Models: Case Study in Cross-Lingual Phonetic Representation Learning.  
Tsvetkov, Sitaram, **Manaal Faruqui**, Lample, Littell, Mortensen, Black, Levin and Dyer.  
Proceedings of NAACL 2016 (long).

Cross-lingual Models of Word Embeddings: An Empirical Comparison.  
Shyam Upadhyay, **Manaal Faruqui**, Chris Dyer, and Dan Roth.  
Proceedings of ACL 2016 (long).

Learning Curriculum with Bayesian Optimization for Task-Specific Word Representation Learning.  
Yulia Tsvetkov, **Manaal Faruqui**, Wang Ling, Brian MacWhinney and Chris Dyer.  
Proceedings of ACL 2016 (long).

Sparse Overcomplete Word Vector Representations.  
**Manaal Faruqui**, Yulia Tsvetkov, Dani Yogatama, Chris Dyer, and Noah Smith.  
Proceedings of ACL 2015 (long).

Non-distributional Word Vector Representations.  
**Manaal Faruqui** and Chris Dyer.  
Proceedings of ACL 2015 (short).

Retrofitting Word Vectors to Semantic Lexicons.  
**Manaal Faruqui**, Jesse Dodge, Sujay Jauhar, Chris Dyer, Ed Hovy and Noah Smith.  
Proceedings of NAACL 2015 (long). **Best Student Paper Award.**

Multilingual Open Relation Extraction Using Cross-lingual Projection.  
**Manaal Faruqui** and Shankar Kumar.  
Proceedings of NAACL 2015 (short).

Learning Word Representations with Hierarchical Sparse Coding.  
Dani Yogatama, **Manaal Faruqui**, Chris Dyer and Noah Smith.  
Proceedings of ICML 2015 (long).

Evaluation of Word Vector Representations by Subspace Alignment.  
Yulia Tsvetkov, **Manaal Faruqui**, Wang Ling, Guillaume Lample and Chris Dyer.  
Proceedings of EMNLP 2015 (short).

Improving Vector Space Word Representations Using Multilingual Correlation.  
**Manaal Faruqui** and Chris Dyer.  
Proceedings of EACL 2014 (long).

Augmenting English Adjective Senses with Supersenses.  
Yulia Tsvetkov, Nathan Schneider, Dirk Hovy, Archana Bhatia, **Manaal Faruqui** and Chris Dyer.  
Proceedings of LREC 2014 (long).

An Information Theoretic Approach to Bilingual Word Clustering.  
**Manaal Faruqui** and Chris Dyer.  
Proceedings of ACL 2013 (short).

Towards a model of formal and informal address in English.  
**Manaal Faruqui** and Sebastian Padó.  
Proceedings of EACL 2012 (long).

Handling OOV words in Indian-language–English CLIR.  
Parin Chheda, **Manaal Faruqui** and Pabitra Mitra.  
Proceedings of ECIR 2012 (short).

“I thou thee, thou traitor”: Predicting Formal vs. Informal Address in English Literature.  
**Manaal Faruqui** and Sebastian Padó.  
Proceedings of ACL 2011 (short).

Acquiring Positive Entailment Pairs Across Languages & Domains: A data analysis.  
**Manaal Faruqui** and Sebastian Padó.  
Proceedings of IWCS 2011 (long).

Training and Evaluating a German Named Entity Recognizer with Semantic Generalization.  
**Manaal Faruqui** and Sebastian Padó.  
Proceedings KONVENS 2010 (short).

WORKSHOP  
PUBLICATIONS

Problems With Evaluation of Word Embeddings Using Word Similarity Tasks.  
**Manaal Faruqui**, Yulia Tsvetkov, Pushpendre Rastogi, Chris Dyer.  
Proceedings of the Representation Evaluation Workshop at ACL 2016.

Correlation-based Intrinsic Evaluation of Word Vector Representations.  
Yulia Tsvetkov, **Manaal Faruqui** and Chris Dyer.  
Proceedings of the Representation Evaluation Workshop at ACL 2016.

Community Evaluation and Exchange of Word Vectors at [wordvectors.org](http://wordvectors.org)  
**Manaal Faruqui** and Chris Dyer.  
Proceedings of ACL Demo Session 2014 (short).

A Framework for (Under)specifying Dependency Syntax without Overloading Annotators.  
Nathan Schneider, Brendan O'Connor, Naomi Saphra, David Bamman, **Manaal Faruqui**, Jason  
Baldrige, Noah A. Smith and Chris Dyer.  
Proceedings of the Linguistic Annotation Workshop at ACL 2013.

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.  
Proceedings of the Workshop on NLP for Building Educational Applications at NAACL 2013.

Soundex-based Translation Correction in Urdu–English CLIR.  
**Manaal Faruqui**, Prasenjit Majumder and Sebastian Padó.  
Proceedings of Workshop on Cross Lingual Information Access at IJCNLP 2011.

TEACHING  
EXPERIENCE

Carnegie Mellon University  
*Teaching Assistant*  
Instructors: Prof. Chris Dyer  
Machine Translation 11-731, Graduate level course

**Spring, 2016**

Carnegie Mellon University  
*Teaching Assistant*  
Instructors: Prof. Noah Smith & Prof. Chris Dyer  
Natural Language Processing 11-411, Undergraduate level course

**Spring, 2014**

Carnegie Mellon University  
*Teaching Assistant*  
Instructors: Prof. Noah Smith  
Advanced Natural Language Processing Seminar 11-713, Graduate level course

**Fall, 2014**

SERVICE

Tutorial:

2017: Tutorial on Cross-lingual word representations at EMNLP

Workshop Organizer:

2017: Workshop on Subword and Character-level models in NLP at EMNLP

2016: Workshop on Multilingual and Cross-lingual Methods in NLP at NAACL

Area Chair:

2017: Semantics, ACL

Session Chair:

2016: Word Embeddings, NAACL

2014: Semantics and Discourse, EACL

Journal Reviewing:

2017: Transactions of ACL, Computational Linguistics

2016: Computational Linguistics, TALLIP

Conference reviewing:

2016: NAACL, ACL, ICML, CoNLL, ICLR

2015: EMNLP, NAACL, \*SEM

2014: ICML, ACL, EACL, CoNLL

2013: NAACL, CoNLL

2012: EACL

Departmental service:

Moderator, LTI Open House Student Panel, 2015

Member, LTI Colloquium Committee, 2013-14, 2014-15

Moderator, LTI Open House Student Panel, 2014

Member, LTI Open House Committee, 2014

Member, LTI Students' Research Symposium Committee, 2014

TALKS

Invited Talks:

08/16: Inducing Morpho-syntactic Lexicons & Morphological Inflections. Ohio State University.

07/16: Beyond the Distributional Hypothesis. University of Tokyo.

03/16: Beyond the Distributional Hypothesis. IBM Research, Yorktown Hts.

03/16: Beyond the Distributional Hypothesis. Bloomberg, New York City.

03/16: Beyond the Distributional Hypothesis. Google, New York City.  
03/16: Beyond the Distributional Hypothesis. Microsoft Research, Seattle.  
03/16: Beyond the Distributional Hypothesis. Allen Institute for AI, Seattle.  
08/15: Improving and Better Understanding Word Vectors. University of Stuttgart.  
08/15: Improving and Better Understanding Word Vectors. University College London.  
07/15: Improving and Better Understanding Word Vectors. University of Edinburgh.  
06/15: Improving and Better Understanding Word Vectors. University of Copenhagen.  
06/15: Improving and Better Understanding Word Vectors. Cambridge University.  
01/13: Machine Learning for NLP: An Introduction. VIT Chennai, India.

Conference Talks:

08/16: Generating Morpho-syntactic Lexicons. ACL in Berlin.  
06/16: Morphological Inflection Generation. NAACL in San Diego.  
07/15: Sparse Overcomplete Word Vector Representations. ACL in Beijing.  
06/15: Retrofitting Word Vectors to Semantic Lexicons. NAACL in Denver.  
04/14: Improving Word Vectors Using Multilingual Correlation. EACL in Gothenburg.  
08/13: An Information Theoretic Approach to Bilingual Word clustering. ACL in Sofia.

At CMU:

10/15: Graph-based Models for Lexical Semantics. Machine Learning Lunch.  
03/14: Lexical Semantics. A lecture in the undergraduate NLP course (11-411).  
08/13: Multilinguality to the Rescue. LTI Students' Research Symposium.  
08/12: Towards a model of formal and informal address in English. LTI Students' Research Symposium. Won the **Best Presentation Award**.

LANGUAGES

C++, Python, Native speaker of Hindi, Proficient in English.