Jacob Eisenstein

Assistant Professor

School of Interactive Computing
College of Computing
Georgia Institute of Technology
Atlanta, GA 30332-0760, USA

I. Earned degrees

Ph.D.	2008	Massachusetts Institute of Technology	*Computer Science*
M.S.	2002	University of Southern California	*Computer Science*
B.S.	1999	Stanford University	*Philosophy*

II. Professional Employment History

Assistant Professor	School of Interactive Computing Georgia Institute of Technology	*2012-present*
Postdoctoral Scholar	Machine Learning Department Carnegie Mellon University	*2009-2011*
Postdoctoral Fellow	Beckman Institute University of Illinois	*2008-2009*

III. Honors and Awards

A. International or National Awards

1. **Best Student Paper Award**, with Vinodh Krishnan, 2015. North American Association for Computational Linguistics. Awarded for our paper *"You're Mr. Lebowski, I'm The Dude": Inducing address term formality in signed social networks*.
2. **NSF Career Award**, 2015. Awarded for the proposal "Sociolinguistic Structure Induction."
3. **SICSA Distinguished Visiting Fellow**. Scotland, 2013. Awarded by the Scottish Informatics & Computer Science Alliance, to support collaborative research and presentations in Scotland.
4. **Best Presentation Award**. NAACL Workshop on Computational Linguistics in a World of Social Media, 2010. Awarded for the presentation of "Social Links from Latent Topics in Microblogs."

B. Institute or School Awards

1. **James D. Lester III Family Award** for internet research, 2016. Awarded by the College of Computing at the Georgia Institute of Technology.
2. **Class of 1940 Course Survey Teaching Effectiveness Award**. Georgia Tech, Spring 2013. Awarded for outstanding teaching evaluations in CS 7650, Natural Language Understanding.
3. **George M. Sprowls Award**. MIT, 2009. Awarded for the best doctoral theses in Computer Science.

IV. Research and Creative Scholarship

Following Georgia Tech's preferred formatting conventions, asterisks indicate publications written while at Georgia Tech, and boldface is used to indicate student advisees. This includes Georgia Tech undergraduate and graduate students, as well as visiting PhD students Dong Nguyen and Caglar Tirkaz. Students at other universities are not highlighted.

A. Published Books, Parts of Books, and Edited Volumes

A.1. Books

[1] * Jacob Eisenstein. *Natural Language Processing: A Technical Introduction*. MIT Press, in preparation.

A.2. Other Parts of Books: Invited Chapters

[1] * Jacob Eisenstein. Written dialect variation in online social media. In Charles Boberg, John Nerbonne, and Dom Watt, editors, *Handbook of Dialectology*. Wiley, 2017.

[2] Angel Puerta and Jacob Eisenstein. Developing a multiple user interface framework for industry. In A. Seffah and H. Javahery, editors, *Multiple User Interfaces: Engineering and Application Frameworks*. John Wiley and Sons, 2003.

B. Refereed Publications and Submitted Articles

B.1. Published and Accepted Journal Articles

[1] * **Dong Nguyen** and Jacob Eisenstein. A kernel independence test for geographical language variation. *Computational Linguistics*, 43, 2017.

[2] * **Yi Yang** and Jacob Eisenstein. Overcoming language variation in sentiment analysis with social attention. *Transactions of the Association for Computational Linguistics (TACL)*, 5, 2017.

[3] * **Umashanthi Pavalanathan** and Jacob Eisenstein. More emojis, less :) The competition for paralinguistic functions in microblog writing. *First Monday*, 22(11), November 2016.

[4] * Lauren F. Klein, Jacob Eisenstein and **Iris Sun**. Exploratory thematic analysis for digitized archival collections. *Digital Scholarship in the Humanities*, October 2015.

[5] * **Caglar Tirkaz**, Jacob Eisenstein, T. Metin Sezgin and Berrin Yanikoglu. Identifying visual attributes for object recognition from text and taxonomy. *Computer Vision and Image Understanding*, 2015.

[6] * **Umashanthi Pavalanathan** and Jacob Eisenstein. Audience-modulated variation in online social media. *American Speech*, 90(2), May 2015.

[7] * Jacob Eisenstein. Systematic patterning in phonologically-motivated orthographic variation. *Journal of Sociolinguistics*, 19:161–188, 2015.

[8] * **Yangfeng Ji** and Jacob Eisenstein. One vector is not enough: Entity-augmented distributional semantics for discourse relations. *Transactions of the Association for Computational Linguistics (TACL)*, June 2015.

[9] * Jacob Eisenstein, Brendan O'Connor, Noah A. Smith and Eric P. Xing. Diffusion of lexical change in social media. *PLoS ONE*, 9, November 2014.

[10] * David Bamman, Jacob Eisenstein and Tyler Schnoebelen. Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, 18(2):135–160, 2014.

[11] * Lauren Klein and Jacob Eisenstein. Reading thomas jefferson with topicviz: Towards a thematic method for exploring large cultural archives. *Scholarly and Research Communication*, 4(3), 2013.

[12] Tahira Naseem, Benjamin Snyder, Jacob Eisenstein and Regina Barzilay. Multilingual part-of-speech tagging: Two unsupervised approaches. *Journal of Artificial Intelligence Research*, 36, November 2009.

[13] SRK Branavan, Harr Chen, Jacob Eisenstein and Regina Barzilay. Learning document-level semantic properties from free-text annotations. *Journal of Artificial Intelligence Research*, 34(2):569, 2009.

[14] Jacob Eisenstein, Regina Barzilay and Randall Davis. Modeling gesture salience as a hidden variable for coreference resolution and keyframe extraction. *Journal of Artificial Intelligence Research*, 31:353–398, 2008.

[15] Jacob Eisenstein. Book review: Gesture in Human Computer Interaction and Simulation (edited by Sylvie Gibet, Nicolas Courty and Jean-Francois Kamp). *Gesture*, 7(1):119–127, May 2007.

[16] Angel Puerta and Jacob Eisenstein. Towards a general computational framework for model-based interface development systems. *Knowledge-Based Systems*, 12:433–442, 1999.

B.2. Conference Presentation with Proceedings (Refereed)

[1] * Eshwar Chandrasekharan, **Umashanthi Pavalanathan**, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein and Eric Gilbert. You can't stay here: The effectiveness of Reddit's 2015 ban through the lens of hate speech. *Proceedings of the ACM on Human-Computer Interaction*, 1, November 2017.

[2] * **Yuval Pinter**, **Robert Guthrie** and Jacob Eisenstein. Mimicking word embeddings using subword RNNs. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, September 2017.

[3] * **Umashanthi Pavalanathan**, Jim Fitzpatrick, Scott F. Kiesling and Jacob Eisenstein. A multidimensional lexicon for interpersonal stancetaking. In *Proceedings of the Association for Computational Linguistics (ACL)*, Vancouver, 2017.
24% acceptance rate.

[4] * Jacob Eisenstein. Unsupervised learning for lexicon-based classification. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, San Francisco, 2017.
25% acceptance rate.

[5] * **Rahul Goel**, **Sandeep Soni**, **Naman Goyal**, John Paparrizos, Hanna Wallach, Fernando Diaz and Jacob Eisenstein. The social dynamics of language change in online networks. In *The International Conference on Social Informatics (SocInfo)*, November 2016.
29% acceptance rate.

[6] * **Parminder Bhatia**, **Robert Guthrie** and Jacob Eisenstein. Morphological priors for probabilistic neural word embeddings. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, November 2016.
25% acceptance rate.

[7] * **Yi Yang**, Ming-Wei Chang and Jacob Eisenstein. Toward socially-infused information extraction: Embedding authors, mentions, and entities. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, November 2016.
25% acceptance rate.

[8] * **Yangfeng Ji**, Gholamreza Haffari and Jacob Eisenstein. A latent variable recurrent neural network for discourse relation language models. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, San Diego, CA, June 2016.
24% acceptance rate.

[9] * **Yi Yang** and Jacob Eisenstein. Part-of-speech tagging for historical English. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, San Diego, CA, June 2016.
24% acceptance rate.

[10] * **Yangfeng Ji**, **Gongbo Zhang** and Jacob Eisenstein. Closing the gap: Domain adaptation from explicit to implicit discourse relations. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, September 2015.
24% acceptance rate.

[11] * **Parminder Bhatia**, **Yangfeng Ji** and Jacob Eisenstein. Better document-level sentiment analysis from rst discourse parsing. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, September 2015.
24% acceptance rate.

[12] * **Umashanthi Pavalanathan** and Jacob Eisenstein. Confounds and consequences in geotagged twitter data. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, September 2015.
24% acceptance rate.

[13] * **Vinodh Krishnan** and Jacob Eisenstein. "You're Mr. Lebowski, I'm The Dude": Inducing address term formality in signed social networks. In *NAACL*, 2015.
26% acceptance rate; one of two papers selected for the best student paper award.

[14] * **Yi Yang** and Jacob Eisenstein. Unsupervised multi-domain adaptation with feature embeddings. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Denver, CO, May 2015.
26% acceptance rate.

[15] * **Jose Manuel Delgado Valdes**, Jacob Eisenstein and Munmun De Choudhury. Psychological effects of urban crime gleaned from social media. In *Proceedings of the International Conference on Web and Social Media (ICWSM)*, Menlo Park, California, May 2015. AAAI Press.
31% acceptance rate.

[16] * **Yangfeng Ji** and Jacob Eisenstein. Representation learning for text-level discourse parsing. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD, 2014.
26% acceptance rate.

[17] * **Kairit Sirts**, Jacob Eisenstein, Micha Elsner and Sharon Goldwater. Pos induction with distributional and morphological information using a distance-dependent chinese restaurant process. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD, 2014.
26% acceptance rate.

[18] * **Sandeep Soni**, **Tanushree Mitra**, Eric Gilbert and Jacob Eisenstein. Modeling factuality judgments in social media text. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD, 2014.
26% acceptance rate.

[19] * **Yi Yang** and Jacob Eisenstein. Fast easy unsupervised domain adaptation with marginalized structured dropout. In *Proceedings of the Association for Computational Linguistics (ACL)*, Baltimore, MD, 2014.
26% acceptance rate.

[20] * **Yangfeng Ji** and Jacob Eisenstein. Discriminative improvements to distributional sentence similarity. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pages 891–896, October 2013.
27% acceptance rate.

[21] * Aaron K Massey, Jacob Eisenstein, Annie I Antón and Peter P Swire. Automated text mining for requirements analysis of policy documents. In *Proceedings of Requirements Engineering (RE)*, pages 4–13, Rio de Janeiro, 2013.
18% acceptance rate.

[22] * **Rakshit Trivedi** and Jacob Eisenstein. Discourse connectors for latent subjectivity in sentiment analysis. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 808–813, 2013.
30% acceptance rate.

[23] * **Yi Yang** and Jacob Eisenstein. A log-linear model for unsupervised text normalization. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, 2013.
27% acceptance rate.

[24] * Jacob Eisenstein. What to do about bad language on the internet. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 359–369, 2013.
30% acceptance rate.

[25] * Micha Elsner, Sharon Goldwater and Jacob Eisenstein. Bootstrapping a unified model of lexical and phonetic acquisition. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 184–193, Jeju, Korea, 2012.
19% acceptance rate.

[26] * Jacob Eisenstein, Duen H. Chau, Aniket Kittur and Eric Xing. TopicViz: interactive topic exploration in document collections. In *Proceedings of CHI: Works-in-progress*, pages 2177–2182, Austin, TX, 2012.
48% acceptance rate.

[27] * Qirong Ho, Jacob Eisenstein and Eric P. Xing. Document hierarchies from text and links. In *Proceedings of the Conference on World-Wide Web (WWW)*, pages 739–748, Lyon, France, 2012.
12% acceptance rate.

[28] Amr Ahmed, Qirong Ho, Choon H Teo, Jacob Eisenstein, Eric P Xing and Alex J Smola. Online inference for the infinite topic-cluster model: Storylines from streaming text. In *Proceedings of Artificial Intelligence and Statistics (AISTATS)*, pages 101–109, Fort Lauderdale, FL, 2011.
28% acceptance rate.

[29] Amr Ahmed, Qirong Ho, Jacob Eisenstein, Eric Xing, Alexander J Smola and Choon H Teo. Unified analysis of streaming news. In *Proceedings of the Conference on World-Wide Web (WWW)*, pages 267–276, Hyderabad, India, 2011.
12% acceptance rate.

[30] Jacob Eisenstein, Noah A. Smith and Eric P. Xing. Discovering sociolinguistic associations with structured sparsity. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 1365–1374, Portland, OR, 2011.
26% acceptance rate.

[31] Jacob Eisenstein, Amr Ahmed and Eric P. Xing. Sparse additive generative models of text. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 1041–1048, 2011.
25% acceptance rate.

[32] Kevin Gimpel, Nathan Schneider, Brendan O'Connor, Dipanjan Das, Daniel Mills, Jacob Eisenstein, Michael Heilman, Dani Yogatama, Jeffrey Flanigan and Noah A. Smith. Part-of-speech tagging for Twitter: annotation, features, and experiments. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 42–47, Portland, OR, 2011.
26% acceptance rate.

[33] Jacob Eisenstein, Brendan O'Connor, Noah A. Smith and Eric P. Xing. A latent variable model for geographic lexical variation. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, pages 1277–1287, 2010.
25% acceptance rate.

[34] Jacob Eisenstein. Hierarchical text segmentation from multi-scale lexical cohesion. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Boulder, CO, 2009.
29% acceptance rate.

[35] Jacob Eisenstein, James Clarke, Dan Goldwasser and Dan Roth. Reading to learn: constructing features from semantic abstracts. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, Singapore, 2009.
34% acceptance rate.

[36] B. Snyder, T. Naseem, J. Eisenstein and R. Barzilay. Adding more languages improves unsupervised multilingual part-of-speech tagging: A bayesian non-parametric approach. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Boulder, CO, 2009.
29% acceptance rate.

[37] S. R. K. Branavan, Harr Chen, Jacob Eisenstein and Regina Barzilay. Learning document-level semantic properties from free-text annotations. In *Proceedings of the Association for Computational Linguistics (ACL)*, Columbus, OH, June 2008.
25% acceptance rate.

[38] Jacob Eisenstein, Regina Barzilay and Randall Davis. Gestural cohesion for discourse segmentation. In *Proceedings of the Association for Computational Linguistics (ACL)*, Columbus, OH, June 2008.
25% acceptance rate.

[39] Jacob Eisenstein, Regina Barzilay and Randall Davis. Discourse topic and gestural form. In *AAAI*, Chicago, IL, July 2008.
24% acceptance rate.

[40] Jacob Eisenstein and Regina Barzilay. Bayesian unsupervised topic segmentation. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, Honolulu, HI, October 2008.
30% acceptance rate.

[41] Benjamin Snyder, Tahira Naseem, Jacob Eisenstein and Regina Barzilay. Unsupervised multilingual learning for POS tagging. In *Proceedings of Empirical Methods for Natural Language Processing (EMNLP)*, Honolulu, HI, October 2008.
30% acceptance rate.

[42] Jacob Eisenstein, Regina Barzilay and Randall Davis. Turning lectures into comic books with linguistically salient gestures. In *AAAI*, pages 877–882, Vancouver, 2007.
27.5% acceptance rate.

[43] Jacob Eisenstein and Randall Davis. Conditional modality fusion for coreference resolution. In *Proceedings of the Association for Computational Linguistics (ACL)*, pages 352–359, Prague, 2007.
22% acceptance rate.

[44] Jacob Eisenstein and Wendy Mackay. Interacting with communication appliances: An evaluation of two computer vision-based selection techniques. In *CHI*, pages 1111–1114, Montréal, 2006.
23% acceptance rate.

[45] Jacob Eisenstein and Randall Davis. Gesture improves coreference resolution. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 37–40, New York, NY, 2006.
41% short paper acceptance rate.

[46] Jacob Eisenstein and Randall Davis. Gesture features for coreference resolution. In Steve Renals, Samy Bengio, and Jonathan G. Fiscus, editors, *Machine Learning for Multimodal Interaction*, Lecture Notes in Computer Science, Berlin, 2006. Springer.

[47] Jacob Eisenstein and Randall Davis. Visual and linguistic information in gesture classification. In *ICMI*, pages 113–120, State College, PA, 2004.
40% acceptance rate.

[48] Jacob Eisenstein and C. Mario Christoudias. A salience-based approach to gesture-speech alignment. In *Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, pages 25–32, Boston, MA, 2004.
26% acceptance rate.

[49] Jacob Eisenstein, S. Ghandeharizadeh, L. Golubchik, C. Shahabi, D. Yan and R. Zimmermann. Device independence and extensibility in gesture recognition. In *Proceedings of IEEE Virtual Reality*, Los Angeles, 2003.
28% acceptance rate.

[50] Laurent Bouillon, Jean Vanderdonckt and Jacob Eisenstein. Model-based approaches to reengineering web pages. In *Proceedings of Task Model and Diagrams for User Interface design (TAMODIA)*, Bucharest, 2002.
Acceptance rate unknown.

[51] Angel Puerta and Jacob Eisenstein. XIML: A universal language for user interfaces. In *Proceedings of Intelligent User Interfaces (IUI)*, San Francisco, 2002.
44% acceptance rate.

[52] Jacob Eisenstein and Charles Rich. Agents and GUIs from task models. In *Proceedings of Intelligent User Interfaces (IUI)*, pages 47–54, San Francisco, 2002.
44% acceptance rate.

[53] Jacob Eisenstein. Modeling preference for adaptive user interfaces. In *Proceedings of Universal Access in Human-Computer Interaction*, New Orleans, 2001.
Acceptance rate unknown.

[54] Jacob Eisenstein, Jean Vanderdonckt and Angel R. Puerta. Applying model-based techniques to the development of UIs for mobile computers. In *Proceedings of Intelligent User Interfaces (IUI)*, pages 69–76, Santa Fe, NM, 2001.
Acceptance rate unknown.

[55] J. Eisenstein, S. Ghandeharizadeh, L. Huang, C. Shahabi, G. Shanbhag and R. Zimmermann. Analysis of clustering techniques to detect hand signs. In *Proceedings of the International Symposium on Intelligent Multimedia, Video and Speech Processing*, 2001.
Acceptance rate unknown.

[56] Jacob Eisenstein, Shahram Ghandeharizadeh, Cyrus Shahabi, Gautum Shanbhag and Roger Zimmermann. Alternative representations and abstractions for moving sensors databases. In *Proceedings of Conference on Information and Knowledge Management (CIKM)*, pages 318–325, Atlanta, GA, 2001.
25% acceptance rate.

[57] Jacob Eisenstein and Angel R. Puerta. Adaptation in automated user-interface design. In *Proceedings of Intelligent User Interfaces (IUI)*, pages 74–81, New Orleans, 2000.
Acceptance rate unknown.

[58] J. Eisenstein, J. Vanderdoncki and A. Puerta. Adapting to mobile contexts with user-interface modeling. In *Proceedings of IEEE Workshop on Mobile Computing Systems and Applications*, pages 83–92, Monterey, CA, 2000.
Acceptance rate unknown.

[59] Angel R. Puerta and Jacob Eisenstein. Towards a general computational framework for model-based interface development systems. In *Proceedings of Intelligent User Interfaces (IUI)*, pages 171–178, Redondo Beach, CA, 1999.
30% acceptance rate.

[60] Angel R. Puerta and Jacob Eisenstein. Interactively mapping task models to interfaces in MOBI-D. In *Proceedings of Design, Specification and Verification of Interactive Systems*, pages 261–273, Abingdon, UK.
Acceptance rate unknown.

B.3. Other Refereed Material

[1] * **Naman Goyal** and Jacob Eisenstein. A joint model of rhetorical discourse structure and summarization. In *EMNLP Workshop on Structured Prediction for NLP*, 2016.

[2] * **Vinodh Krishnan** and Jacob Eisenstein. Nonparametric bayesian storyline detection from microtexts. In *EMNLP Workshop on Computing News Storylines (CNewsStory)*, 2016.

[3] * **Umashanthi Pavalanathan** and Jacob Eisenstein. Emoticons vs. emojis on twitter: A causal inference approach. In *Proceedings of AAAI Spring Symposium on Observational Studies through Social Media and Other Human-Generated Content (OSSM)*, March 2016.

[4] * **Yangfeng Ji** and Jacob Eisenstein. Entity-augmented distributional semantics for discourse relations. Presented at the workshop track of the International Conference of Learning Representations (ICLR) in San Diego, 2015.

[5] * Jacob Eisenstein, **Iris Sun** and Lauren Klein. Exploratory thematic analysis for historical newspaper archives. Presented at Digital Humanities (DH) in Lausanne, 2014.

[6] * Jacob Eisenstein. Phonological factors in social media writing. In *Proceedings of the Workshop on Language Analysis in Social Media*, pages 11–19, Atlanta, 2013.

[7] Jacob Eisenstein, Tae Yano, William W Cohen, Noah A Smith and Eric P Xing. Structured databases of named entities from bayesian nonparametrics. In *Proceedings of the First Workshop on Unsupervised Learning in NLP (at EMNLP)*, pages 2–12, Edinburgh, UK, 2011.

[8] Kriti Puniyani, Jacob Eisenstein, Shay Cohen and Eric P. Xing. Social links from latent topics in microblogs. In *Proceedings of NAACL Workshop on Social Media*, Los Angeles, 2010.
Winner of the workshop's best presentation award.

[9] Aaron Adler, Jacob Eisenstein, Michael Oltmans, Lisa Guttentag and Randall Davis. Building the design studio of the future. In *AAAI Workshop on Making Pen-Based Interaction Intelligent and Natural*, pages 1–7, Menlo Park, CA, 2004.

C. Other Publications and Creative Projects

C.1. Technical Reports

[1] Jacob Eisenstein, Duen Horng "Polo" Chau, Aniket Kittur and Eric P. Xing. TopicViz: Semantic navigation of document collections. Technical Report 1110.6200, ArXiv, 2011.

[2] Jacob Eisenstein and Eric P. Xing. The CMU 2008 political blog corpus. Technical report, Carnegie Mellon University, 2010.

[3] Jacob Eisenstein and Randall Davis. Gestural cues for sentence segmentation. Technical report, MIT AI Memo, 2005.

[4] Jacob Eisenstein. Evolving robocode tank fighters. Technical Report AIM-2003-023, MIT AI Memo, 2003.

D. Presentations

D.1. Keynote Addresses and Plenary Lectures

This section includes presentations that were invited at venues that also had peer-reviewed presentations. Presentations at venues where every talk was invited are listed in D.2, Invited Conference and Workshop Presentations.

D.1.1 *ACL Workshop on Representation Learning for NLP*, Vancouver, August 2017. "Learning Sociolinguistic Representations."

D.1.2 *Workshop on Data-Driven Approaches to Networks and Language*, Lyon, May 2016. "Computational Models of Language Change in Online Social Media."

D.1.3 *DiSpoL 2015: Identification and Annotation of Discourse Relations in Spoken Language*, Saarbrücken, October 2015. "Automatic Identification of Discourse Relations from Distributed Representations."

D.1.4 *EMNLP Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics*, Lisbon, September 2015. "From Distributed Semantics to Discourse, and Back."

D.1.5 *SocialNLP workshop at the North American Assocation for Computational Linguistics*, Denver, June 2015. "Variation and change in social media language."

D.1.6 *NIPS Workshop on Modern Machine Learning and Natural Language Processing*, Montreal, December 2014. "Representation learning for discourse parsing."

D.1.7 *Methods in Dialectology XV*, Groningen, 2014. "Dialect variation in online social media."

D.1.8 *IEEE Workshop on Interactive Visual Text Analytics*, Atlanta, 2013. "Understanding language variation in social media."

D.1.9 *Symposium about Language and Society (SALSA)*, Austin, 2013. "Large-scale analysis of language variation and change in social media."

D.2. Invited Conference and Workshop Presentations

D.2.1 Bridging disciplines in analysing text as social and cultural data, Turing Institute (London), September 2017

D.2.2 Symposium on Text as Social and Cultural Data, University of Twente (Enschede, Netherlands), March 2017

D.2.3 Linguistic Society of America Workshop on Preparing Your Corpus for Archival Storage, Wahsington D.C., January 2016

D.2.4 American Association for the Advancement of Science, Annual Meeting, February 2015

D.2.5 Stanford Conference on Computational Social Science, April 2014

D.2.6 Conference on Patterns of Macro- and Micro-Diversity in the Languages of Europe and the Middle East. Computational Issues in Studying Language Diversity: Storage, Analysis and Inference. Groningen, Netherlands, July 2013

D.2.7 Workshop on Big Data and Language Technology, Copenhagen, Denmark, May 2013

D.3. Conference and Workshop Presentations

The following presentations were accepted based on a peer-reviewed abstract submission. Presentations associated with published conference papers are listed above in section B, and not here.

D.3.1 "#thighgap to #thyghgapp: Incrementation of orthographic variation on Instagram," presented by **Ian Stewart** at *Diversity and Variation in Language*, Atlanta, February 2017.

D.3.2 "Stance markers, Reddit thread structure, and community," presented by **Umashanthi Pavalanathan** at *New Ways of Analyzing Variation*, Vancouver, November 2016.

D.3.3 "A Nonparametric Test for Spatial Dependence," presented by **Dong Nguyen** at *New Ways of Analyzing Variation*, Toronto, October 2015.

D.3.4 "Rhetorical Patterns in Congressional Floor Speeches," presented at *New Directions in Text as Data*, NYU, October 2015.

D.3.5 "Reading Thomas Jefferson with TopicViz: Towards a Thematic Method for Exploring Large Cultural Archives," presented (with Lauren Klein) at *Research Foundations for Understanding Books and Reading in the Digital Age: E/Merging Reading, Writing, and Research Practices*, Havana, December 2012.

D.3.6 "Mapping the Spread of New Words in Social Media." NIPS Workshop on Social Network and Social Media Analysis, Lake Tahoe, CA, December 2012.

D.3.7 "Statistical Exploration of Geographical Lexical Variation in Social Media" at *Meeting of the Linguistic Society of America*, Pittsburgh, January 2011.

D.4. Invited Seminar Presentations

Since beginning at Georgia Tech in 2012.

D.4.1 University of Colorado, June 2017

D.4.2 Georgetown University Linguistics Series, April 2017

D.4.3 Language Technology Institute (LTI) Seminar, Carnegie Mellon University, December 2016

D.4.4 Xerox Research Center Europe, Grenoble, May 2016

D.4.5 University of Toulouse, May 2016

D.4.6 École Normale Supérieure de Lyon, May 2016

D.4.7 Text-as-data speaker series, New York University, February 2016

D.4.8 Social Science Data Analytics Series, Michigan State University, January 2016

D.4.9 IGERT Distinguished Speaker Series, Columbia University, November 2015

D.4.10 University of Copenhagen, October 2015

D.4.11 Digital Humanities Day at the University of Georgia, April 2015

D.4.12 Stanford Linguistics Colloquium, February 2015

D.4.13 Google New York, July 2014

D.4.14 Facebook Research, April 2014

D.4.15 University of Alabama at Birmingham, December 2013

D.4.16 University of Washington, October 2013

D.4.17 SICSA Distinguished Visiting Fellow Lectures: Universities of Edinburgh, Glasgow, and Saint Andrews, July 2013.

D.4.18 University of Texas, May 2012

D.4.19 University of Maryland CLIP colloquium, May 2012

D.4.20 Emory University Math and Computer Science seminar, May 2012

D.4.21 Columbia University, March 2012

D.5. Other Presentations

D.5.1 MLConf (The Machine Learning Conference), Atlanta, September 2017.

D.5.2 Lisbon Machine Learning Summer School, July 2012.

E. Grants and Contractors

1. **CAREER: Sociolinguistic Structure Induction**
   Sponsor: National Science Foundation
   Amount and role: $499,995, Sole PI
   Period of award: 2015-2020

2. **Social Media Signals for Post-traumatic Stress and Anxiety in Crisis-Inflicted Communities**
   Sponsor: National Institutes of Health
   Amount and role: $1,448,921, Senior/Key personnel (Funding share: 40%)
   Collaborators: Munmun De Choudhury (PI), James Pennebaker (Senior/Key personnel, University of Texas)
   Period of award: 2015-2020

3. **Hybrid distributional-entity semantics for discourse relations**
   Sponsor: Google Faculty Research Award
   Amount and role: $57,191, Sole PI
   Period of award: 2015-2016

4. **Expanding Lexical Resources with Knowledge Elicitation and representation Learning to Improve Automated Detection and Classification of WMD Events in Text**
   Sponsor: Defense Threat Reduction Agency
   Amount and role: $310,000, Co-PI (Funding share: 80%)
   Collaborators: Bryan Lee (PI; Monterey Institute for International Studies) and Erica Briscoe (Co-PI; Georgia Tech Research Institute)
   Period of award: 2014-2017

5. **Tracing Reuse in Political Language**
   Sponsor: Google Computational Journalism Research Awards
   Amount and role: $60,000, Co-PI (Funded share: two semester GRA)
   Collaborators: Irfan Essa (PI) and Eric Gilbert (Co-PI)
   Period of award: 2014-2015

6. **Socio-digital influence networks from language analysis**
   Sponsor: Air Force Office of Scientific Research, Young Investigator Research Program
   Amount and role: $360,000, Sole PI
   Period of award: 2014-2017

7. **EAGER: Exploring Adapting Language Technology Across a Network of Domains**
   Sponsor: National Science Foundation
   Amount and role: $100,000, Sole PI
   Period of award: 2013-2015

8. **TOME: Interactive TOpic Model and MEtadata Visualization**
   Sponsor: National Endowment for the Humanities
   Amount and role: $59,999, Co-Projector Director (Funding share: 50%)
   Collaborator: Lauren Klein (Project Director)
   Period of award: 2013-2015

9. **Latent discourse trees**
   Sponsor: Google Faculty Research Award
   Amount and role: $50,500, Sole PI
   Period of award: 2012-2013

10. **Collaborative Research: Discovering and Exploiting Latent Communities in Social Media**
    Sponsor: National Science Foundation
    Amount and role: $547,805, Co-PI (Funding share: $80,000 subcontract to Georgia Tech after joining faculty)
    Collaborators: Eric P. Xing (PI) and Scott F. Kiesling (Co-PI)
    Period of award: 2011-2015

F. Other Scholarly Accomplishments

Senior team member for **Wide-band Machine Translation** at the Frederick Jelinek Memorial Summer Workshop on language technology, July-August 2015. The team was selected for funding among many other proposals at an interactive peer review meeting in Baltimore in November 2014.

G. Societal And Policy Impacts

My research on computational linguistic analysis of social media has been featured in a number of mass media venues, including both print (New York Times, Time Magazine, the Wall Street Journal, the Washington Post, the Boston Globe) and radio (National Public Radio programs *All Things Considered* and *Here & Now*, the BBC program *Click*).

H. Other Professional Activities

Consultant, Microsoft Research New York City, 2014.

V. Teaching

A. Courses Taught

Semester	Course Number	Course Name	Number of students
Fall, 2017	CS8803-CSS	Computational Social Science	20
Spring, 2017	CS4650	Natural Language	39
Spring, 2017	CS7650	Natural Language	49
Spring, 2016	CS4464	Computational Journalism	30
Spring, 2016	CS6465	Computational Journalism	10
Fall, 2015	CS 4650	Natural Language	28
Fall, 2015	CS 7650	Natural Language	59
Spring, 2015	CS 8803-CSS	Computational Social Science	15
Fall, 2014	CS 4650	Natural Language	24
Fall, 2014	CS 7650	Natural Language	32
Spring, 2014	CS 8803-CSS	Computational Social Science	11
Fall, 2013	CS 4650	Natural Language	5
Fall, 2013	CS 7650	Natural Language	24
Spring, 2013	CS 4650	Natural Language	5
Spring, 2013	CS 7650	Natural Language	16
Spring, 2012	CS 4650	Natural Language	8
Spring, 2012	CS 7650	Natural Language	19

B. Individual Student Guidance

B.1. Ph.D. Students

B.1.a Graduated

Yangfeng Ji *Computer Science*
Fall 2012 – Summer 2016; was previously in ECE Doctoral Program
Defended dissertation in May 2016, now a postdoc at University of Washington
Dissertation title: Semantic Representation Learning for Discourse Processing

Yi Yang *Computer Science*
Fall 2013 – Fall 2016; was previously a student of Calton Pu in the CS Doctoral Program
Defended dissertation in November 2016, now a research scientist at Bloomberg
Dissertation title: Robust Adaptation of Natural Language Processing for Language Variation

B.1.b In Process

Umashanthi Pavalanathan *Computer Science*
Fall 2013 – present
Thesis proposal approved in Spring 2017
Research topic: Sociolinguistic style in social media

Sandeep Soni *Computer Science*
Fall 2015 – present
Passed qualifier in Spring 2017
Research topic: Modeling sociolinguistic influence

Ian Stewart *Human-Centered Computing*
Fall 2015 – present

Passed qualifier in Spring 2017
Research topic: Orthographic variation as a sociolinguistic variable

Yuval Pinter *Computer Science*
Fall 2016 – present
Research topic: Neural probabilistic models of lexical semantics and morphology

Sarah Wiegreffe *Computer Science*
Fall 2017 – present
Research topic: Deep learning for analysis of electronic health records

B.2. M.S. Students

Zhewei Sun MS in Computer Science
Fall 2017
Project: *Discourse-driven machine translation*

Murali Raghu Babu MS in Computer Science
Fall 2017
Project: *Social network homophily in natural language processing*

Taha Merghani MS in Computer Science
Fall 2017
Project: *Social network homophily in natural language processing*

James Mullenbach MS in Computer Science
Spring 2017 – present
Project: *Predicting diagnosis codes from the text of electronic health records*

Karishma Malkan MS in Computer Science
Summer 2016
Project: *Neural Turing machines for discourse parsing*
Now a software engineer at Amazon

Akanksha MS in Computer Science
Spring 2016
Project: *Discourse parsing with Bayesian optimization*
Now a software engineer at Amazon

Rahul Goel MS in Computer Science
Fall 2015 – Spring 2016
Project: *Modeling language evolution in social media*
Now a software engineer at Amazon

Naman Goyal MS in Computer Science
Spring 2015 – Spring 2016
Project: *Modeling language evolution in social media; discourse parsing from distant supervision*
Now a software engineer at Yelp

Parminder Bhatia MS in Computer Science
Spring 2015 – Fall 2016
Project: *Discourse parsing for the analysis of document-level sentiment; subword modeling for word representations*
Research scientist at Amazon, previously research software engineer at Yik Yak

Gongbo Zhang MS in Computer Science
Fall 2014 – January 2015
Project: *Discourse parsing from distant supervision*
Now a PhD student at the University of Delaware

Vinodh Krishnan MS in Computer Science
Spring 2014 – Spring 2015
Project: *Inducing signed social networks from dialogue*
Now a software engineer at Google

Ivan Sysoev MS in Computer Science
Spring 2014
Project: *Semantic dependency parsing*
Now a PhD student in the MIT MediaLab

Rakshit Trivedi MS in Computer Science
Spring 2013
Project: *Discourse connectives in sentiment analysis*
Now a PhD student at Georgia Tech

Sandeep Soni MS in Computer Science
Fall 2012 – Spring 2014
Project: *Uncertainty and bias in attributed quotes*
Subsequently a software engineer at Yahoo!, now a PhD student at Georgia Tech

Arindam Das MS in Computer Science
Fall 2012
Project: *Social media variation in social networks*

Sahil Patwardhan MS in Computer Science
Fall 2012
Project: *Automatic summarization of Twitter feeds*

Neha Gupta MS in Computer Science
Spring 2012
Project: *Analyzing political blogs for polarity in thread structure*

B.3. Undergraduate Students

Jay Devanathan *Bachelors in Computer Science*
Fall 2017 – present
Project: *Discourse-driven machine translation*

Marc Marone *Bachelors in Computer Science*
Fall 2017 – present
Project: *Discourse-driven machine translation*

Xiaochuang Han *Bachelors in Computer Science*
Fall 2017 – present
Project: *Stance and style matching in online social media*

Robert Guthrie *Bachelors in Computer Science*
Spring 2016 – Spring 2017
Project: *Word embeddings as latent variables*

Patrick Violette *Bachelors in Computer Science*
Fall 2015 – Spring 2016
Project: *Tracking evolving stories in social media* .

Yijie Wang *Bachelors in Computer Science*
Spring 2015 – present
Project: *Representation learning for semantic dependency parsing* .

Ana L. Smith *Bachelors in Computer Science*
Fall 2013 – Spring 2015
Project: *Resolving ambiguous toponyms in social media text* .
Winner of the President's Undergraduate Research Award (PURA) in Fall 2014. Now a PhD student at Cornell University.

Naomi Robert *Bachelors in Computer Science*
Fall 2012
Project: *Visualizing geographical language variation*

B.4. Service on thesis or dissertation committees

B.4.a Internal

Tanushree Mitra Computer Science, Georgia Institute of Technology, April 2017.
Thesis Title: Understanding Social Media Credibility
Principal Advisor: Eric Gilbert
Accepted faculty position at Virginia Tech

Seungyeon Kim Computer Science, Georgia Institute of Technology, Summer 2015.
Thesis Title: Novel Document Representations based on Labels and Sequential Information
Principal Advisor: Guy Lebanon

Boyang "Albert" Li Computer Science, Georgia Institute of Technology, Fall 2014.
Thesis Title: Learning Domain-Independent Narrative Intelligence
Principal Advisor: Mark Riedl
Research scientist at Disney Research

Ke Zhou Computer Science, Georgia Institute of Technology, August 2013.
Thesis Title: Extending Low-Rank Matrix Factorizations for Emerging Applications
Principal Advisor: Hongyuan Zha

B.4.b External

Dong Nguyen Computer Science, University of Twente, March 2017.
Thesis Title: Text as Social and Cultural Data: A Computational Perspective on Variation in Text
Principal Advisor: Franciska de Jong, Mariet Theune, and Antal van den Bosch
Postdoctoral research fellow at the Alan Turing Institute

J. Jessy Li Computer and Information Science, University of Pennsylvania., Expected Spring 2017.
Thesis Title: From discourse structure to text specificity: Studies of coherence preferences
Principal Advisor: Ani Nenkova
Accepted faculty position at the University of Texas-Austin

Yulia Tsvetkov School of Computer Science, Carnegie Mellon University., August 2016.
Thesis Title: Cross-lingual transfer of linguistic and metalinguistic knowledge via lexical borrowing
Principal Advisor: Chris Dyer
Faculty at Carnegie Mellon University

Baekkwan Park Political Science, Emory University, Summer 2015.
Thesis Title: The Politics of Naming and Shaming
Principal Advisor: Jeffrey Staton
Senior Research Officer at the Human Rights Center of the University of Essex

David Bamman School of Computer Science, Carnegie Mellon University., Summer 2015.
Thesis Title: People-Centric Natural Language Processing
Principal Advisor: Noah A. Smith
Faculty at the University of California, Berkeley

Qiaoling Liu Computer Science, Emory University, Expected Fall 2014.
Thesis Title: Improving Query Understanding and Search Quality with Community Question Answering Data
Principal Advisor: Eugene Agichtein

Bo Han Computer Science, University of Melbourne, Summer 2014.
Thesis Title: Improving the Utility of Social Media with Natural Language Processing
Principal Advisor: Timothy Baldwin
Postdoctoral researcher at IBM Research

B.5. Mentorship of visiting scholars

Dong Nguyen *PhD Student, University of Twente, Netherlands*
Fall 2015
Project: *Linking language variation in social media and speech*
Now postdoctoral research fellow at the Alan Turing Institute

Caglar Tirkaz *Fulbright Scholar; PhD Student, Sabanci University, Turkey*
Fall 2012-Spring 2013
Project: *Identifying visual attributes from natural language descriptions of visual objects*
Now research scientist at Amazon

C. Other Teaching Activities

C.1. Curriculum Development

CS 8803-CSS Computational Social Science: I created a new course on computational social science, which combines techniques from computational data analysis with traditional social science. The course includes a methodological focus on network analysis and text mining, and briefly treats other methodologies such as time series analysis and causal inference. Students engage with applications of these techniques to political science, sociology, linguistics, economics, and psychology. Foundational readings are paired with recent papers from the research literature in this rapidly-evolving cross-disciplinary field.

C.2. Course improvement

CS 4650/7650 Natural Language I completely redesigned Georgia Tech's course on natural language processing. Substantively, core concepts now include machine learning for language processing, linguistic foundations and their computational representations, and the design of efficient algorithms for predicting and learning linguistic structures. Pedagogically, the course now features rigorous problem sets that help students build a powerful toolset in text mining and linguistic analysis. We begin with basic classifiers (including sentiment analysis), and move to structured prediction (part-of-speech tagging and parsing), concluding with unsupervised techniques for measuring distributional similarity between words. To develop a basic understanding of linguistics, the course includes several short homework assignments that are designed to provide practical reinforcement for the reading and lectures. From 2012-2014, the course

had an independent project in which students often attack cutting-edge research problems; one such project resulted in a short paper at NAACL (Trivedi and Eisenstein, 2013), one of the most prestigious NLP conferences. In 2012, the course was added to the elective list for the MSCS specialization in Machine Learning. In Spring 2013, CS 4650 was added as a pick to the Intelligence thread by the UCC. As the course became larger, the independent final project was replaced with an exam.

CS 4464/6465 Computational Journalism In Spring 2016, I taught Computational Journalism for the first time. The course had previously emphasized multimedia, but there was a rising tide of interest in *data journalism*, in which statistics and machine learning are used to produce news stories and to analyze the news. I therefore re-oriented the course towards these trends, while maintaining the course's unique blend of computation and substantive journalism. Specifically, I introduced units on machine learning classification, sentiment analysis, geographic information systems, and topic models. Students then built on these ideas in their independent projects.

C.3. Professional Development/Continuing Education

No data

C.4. Other Teaching Activities: Ph.D. Qualifier Committees

Shagun Jhaver Social Computing Qualifier, Georgia Institute of Technology, Computer Science Ph.D., Spring 2017.
Principal Advisor: Eric Gilbert.

Eshwar Chandrasekharan Social Computing Qualifier, Georgia Institute of Technology, Computer Science Ph.D., Spring 2017.
Principal Advisor: Eric Gilbert.

Lara Martin Human-Centered Computing Qualifier, Georgia Institute of Technology, Human Centered Computing Ph.D., Spring 2017.
Principal Advisor: Mark Riedl.

Stevie Chancellor Human-Centered Computing Qualifier, Georgia Institute of Technology, Human Centered Computing Ph.D., Spring 2015.
Principal Advisor: Munmun De Choudhury.

Michael Pettinati Intelligent Systems Qualifier, Georgia Institute of Technology Computer Science Ph.D., Spring 2015.
Principal Advisor: Ron Arkin.

Unaiza Ahsan Intelligent Systems Qualifier, Georgia Institute of Technology Computer Science Ph.D., Spring 2015.
Principal Advisor: Irfan Essa.

Daniel Kohlsdorf Machine Learning Qualifier, Georgia Institute of Technology Computer Science Ph.D., Spring 2014.
Principal Advisor: Thad Starner.

Tanushree Mitra Social Computing Qualifier, Georgia Institute of Technology Computer Science Ph.D., Spring 2013.
Principal Advisor: Eric Gilbert.

VI. Service

A. Professional Activities

A.1. Editorial Board Memberships

A.1.1 *Language Variation Studies*, Language Science Press. 2014–present

A.1.2 *Linguistic Issues in Language Technology*, CSLI. 2014–present

A.1.3 *Computational Linguistics*, MIT Press, 2017–present

A.2. Society Offices, Activities, and Memberships

A.2.1 Member, Association for Computational Linguistics (ACL)

A.3. Organization and Chairmanship of Technical Sessions, Workshops and Conferences

A.3.1 **Co-chair and Co-founder**, Atlanta Workshop on Computational Social Science, 2013-2015.

A.3.2 **Faculty advisor**, Student Research Workshop, North American Association for Computational Linguistics (NAACL), 2016.

A.3.3 **Co-Chair**, Workshop on Language Technology for Computational Social Science, Association for Computational Linguistics (ACL), 2014; Workshop on Computational Social Science and Natural Language Processing, Empirical Methods in Natural Language Processing (EMNLP), 2016.

A.3.4 **Student award coordinator**, International Conference on Machine Learning (ICML), 2013.

A.3.5 **Student volunteer coordinator**, North American Association for Computational Linguistics (NAACL), 2013.

A.3.6 **Tutorial Chair**, North American Association for Computational Linguistics (NAACL), 2012; Association for Computational Linguistics (ACL), 2018.

A.3.7 **Area Chair**, Empirical Methods in Natural Language Processing (EMNLP), 2017; North American Association for Computational Linguistics (NAACL), 2016; Association for Computational Linguistics (ACL), 2014; European Association for Computational Linguistics (EACL), 2014.

A.4. Technical Journal or Conference Referee Activities

A.4.1 **Book reviewer**: Morgan Claypool Synthesis Lectures on Human Language Technologies.

A.4.2 **Journal reviewer**: Computational Linguistics (2013-2016); Transactions of the Association for Computational Linguistics (2012-present); Communications of the ACM (2013), Language in Society (2013); IEEE/ACM Transactions on Audio, Speech, and Language Processing (2013); ACM Transactions on Intelligent Systems and Technology (2013); Journal of Artificial Intelligence Research (2016)

A.4.3 **Senior Program Committee**: International Conference on Web and Social Media (ICWSM), 2014.

A.4.4 **Program Committee**: World Wide Web (WWW) conference, 2012, 2014; International Conference on Machine Learning (ICML), 2010, 2012-present; Conference on Natural Language Learning (CoNLL), 2009, 2011, 2013, 2016, 2017; Artificial Intelligence and Statistics (AISTATS), 2013; North American Association for Computational Linguistics (NAACL), 2012, 2013, 2015; Association for Computational Linguistics (ACL), 2009, 2012, 2015-2017; International Conference on Computational Linguistics (COLING), 2010, 2012; Empirical Methods in Natural Language Processing (EMNLP), 2010-2012, 2014-present; European Association for Computational Linguistics,

2012, 2016; Knowledge Discovery and Data Mining (KDD), 2011; SIGDial meeting on Discourse and Dialogue, 2016; Joint Conference on Lexical and Computational Semantics (*SEM), 2016.

A.4.5 **Reviewer**: Neural Information Processing Systems (NIPS), 2011, 2013-present; International Joint Conference on Natural Language Processing (IJCNLP), 2011; SIGGRAPH, 2010; Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities, 2016-present; International Conference on Learning Representations (ICLR), 2015-2016.

A.5. Proposal Panels and Reviews

A.5.1 Reviewer and Panelist, National Science Foundation, Directorate for Computer and Information Science and Engineering, 2012, 2014-2017; Directorate for Social, Behavioral, and Economic Sciences, 2016-2017.

A.5.2 Reviewer, Defense Threat Reduction Agency, Fall 2015.

A.5.3 Reviewer, Air Force Office of Scientific Research: Fall 2014, Spring 2017.

A.5.4 Reviewer, Army Research Office, Summer 2014.

B. Public and Community Service

- **Symposium organizer and moderator**, Annual Meeting of the American Association for the Advancement of Science (AAAS), "The linguistics of status, influence, and innovation: a computational perspective." February 2015.
- Participated in the CS10K Community Event on *Teaching Data*, a discussion on big data for high school teachers, in Fall 2013.

C. Institute Contributions

C.1. Institute Committee Service

C.1.1 Tech Fee committee, 2016-2017

C.2. College Committee Service

C.2.1 Search Committee for Chair of the School of Interactive Computing, 2017 – present.

C.2.2 Representative, College of Computing Technical Support Organization (TSO) advisory board, 2015 – present.

C.2.3 College of Computing Search Committee for Endowed Chair in Machine Learning, 2015 – present.

C.3. School Committee Service

C.3.1 Chair, School of Interactive Computing School Advisory Committee (SAC), 2014 – 2015.

C.3.2 School of Interactive Computing School Advisory Committee (SAC), 2013 – 2014.

C.3.3 School of Interactive Computing Faculty Hiring Committee, 2013 – 2014.

C.3.4 MSCS Area Coordinator (Machine Learning), 2012 – 2014.

C.3.5 School of Interactive Computing Student Recruiting Co-Coordinator, 2012 – 2013.

C.4. Program Development: Research

No data

C.5. Program Development: Academic

C.5.1 Participating Faculty on Georgia Tech's Linguistics Certificate, 2012 – present.

C.6. Other Institute Service Contributions

C.6.1 Organized Digital Humanities Hack Day in Spring 2015, and Computational Social Science hackathon in Spring 2013 and 2014.