

ANNA A. RUMSHISKY

Email: arum@cs.uml.edu

Web: <http://www.cs.uml.edu/~arum>

EDUCATION

Massachusetts Institute of Technology, CSAIL, **Postdoctoral training** (advisor: Pete Szolovits) 2011–2012
Brandeis University, Computer Science, **PhD** (advisor: James Pustejovsky) 2009
Northeastern University, Computer Science, **MS** 2001
Brandeis University, Psychology/Linguistics, **BA**
Moscow State University, Linguistics

ACADEMIC APPOINTMENTS AND EMPLOYMENT

Assistant Professor, Dept of Computer Science, Univ. of Massachusetts Lowell, Lowell MA 2012 – Present
Research Affiliate, MIT CSAIL, Cambridge, MA 2012 – Present
Postdoctoral Associate, MIT CSAIL, Cambridge, MA 2011 – 2012
Lecturer / Postdoctoral Fellow, Dept of Computer Science, Brandeis University, Waltham, MA 2009 – 2010

FUNDING

AWARDED

1. **NSF (CISE IIS): EAGER: Exploring Cognitively Plausible Computational Models for Processing Human Language.** Total budget: **\$109,926**. Budget share: 100%. Role: **PI**. 9/2018–8/2020.
2. **NSF (CISE IIS): CAREER: Developing an Underspecified Representation for Temporal Information in Text .** Total budget: **\$499,423**. Budget share: 100%. Role: **PI**. 3/2017–2/2022.
3. **Army Research Office (Core Program: Network Science, Collective Intelligence):** Detecting civil conflict and information biases in polarized environments in social media. Total Budget: **\$377,900**. Budget share: 100%. Role: **PI**. 4/2016 – 3/2019
4. **NIH (NIMH R01): Natural language processing for characterizing psychopathology.** Budget share: **\$367,638**. Role: **Co-PI (Site PI)**. PI Perlis, Massachusetts General Hospital. 2/2016 – 2/2019.
5. **Philips Research: Sponsorship for Clinical Natural Language Processing Workshop at COLING 2016.** Total budget: **\$8,000**. Budget share: 100%. Role: **PI**. 12/2016.
6. **Philips Research: NLP-driven Predictive Models for Patient Deterioration and Therapy Response.** Budget share: **\$167,896**. Role: **Co-PI (Site PI)**. PI Szolovits, MIT CSAIL. 7/2015 – 6/2018.
7. **European Research Council Tools for Cross-Disciplinary Academic Communication / Jews and Slavs in the Middle Ages.** Budget share: **\$15,450**. Role: **Site PI**. PI Kulik, Hebrew University of Jerusalem. 1/2014 – 12/2014.
8. **National Endowment for the Humanities (Digital Humanities Start-Up Grant) Mapping the History of Knowledge: Text-Based Tools and Algorithms for Tracking the Development of Concepts.** Total budget \$50,000, Budget share: **\$25,000**, Role: **Co-PI (Site PI)**. PI: Gronas, Dartmouth College. 9/2010 – 8/2011.

Grants that fund co-supervised students

1. **Dartmouth College (Scholarly Innovation and Advancement Award):** KnowEvo: Tracking the Evolution of Domains and Concepts in Structured Historical Corpora (*Encyclopedia Britannica 1768-2014*), *US Academic Syllabi 2000-2014*, *Library of Congress Catalogs and Acquisition Lists 1898-2014* . Total budget: \$25,400. Budget share: (co-supervision only). Role: **Co-PI**. 3/2017 – 2/2018.
2. **UMass Lowell (seed grant): Computational model for the discovery of new treatment targets in autoimmune diseases.** Total budget: \$10,000 . Budget share: (co-supervision only). Role: **Co-PI**. 7/2016 – 6/2017
3. **Dartmouth College (Neukom CompX) Modeling scholarly trajectories in scientific space.** Total budget: \$19,400. Budget share: (co-supervision only). Role: **Co-PI**. 3/2016 – 2/2017.
4. **UMass Lowell (seed grant): To Run is Fun – or is it just a Pun? Detecting Signs of Enjoyment with Physical Activity and Food Choices through Computational Analysis of Figurative Language in Online Communication.** Total budget: \$10,000. Budget share: 50%. Role: **Co-PI**. 6/2018 – 5/2019.

PENDING

1. **NIH (R01): Social media-driven personalization and topical context to aid information discovery for breast cancer patients.** Total budget: \$451,555. Budget share: 100%. Role: **Co-PI (Site PI)**. PI Farri, Philips Research. 4/2019 – 3/2022.

2. NSF (CISE IIS): Text coherence and language variety for chit-chat dialogue generation. Total budget: \$499,896. Budget share: 100%. Role: PI. 4/2019 – 3/2022.

PUBLICATIONS (* = student/adviser co-author)

Google Scholar citation count in November 2018: 1323

Google Scholar profile: <http://bit.ly/2ztdmet>

PEER-REFEREED CONFERENCES

1. Y.-F. Luo*, W. Sun, **A. Rumshisky**. A Hybrid Normalization Method for Medical Concepts in Clinical Narrative Using Semantic Matching. *Proceedings of AMIA Informatics Summit 2019*. San Francisco, CA. (accepted)
2. O. Kovaleva*, **A. Rumshisky**, A. Romanov*. Similarity-Based Reconstruction Loss for Meaning Representation. Proceedings of EMNLP 2018 (short papers). Brussels, Belgium. Acceptance rate (estimated) 22%
3. Y. Meng*, **A. Rumshisky**. Triad-based Neural Network for Coreference Resolution. *Proceedings of COLING 2018*. Santa Fe, New Mexico. Acceptance rate: 37.4%.
4. A. Rogers*, S. H. Ananthakrishna*, **A. Rumshisky**. What's in Your Embedding, and How It Predicts Task Performance. *Proceedings of COLING 2018*. Santa Fe, New Mexico. Acceptance rate: 37.4%.
5. A. Rogers*, A. Romanov*, **A. Rumshisky**, S. Volkova, M. Gronas, A. Gribov*. RuSentiment: An Enriched Sentiment Analysis Dataset for Social Media in Russian. *Proceedings of COLING 2018*. Santa Fe, New Mexico. Acceptance rate: 37.4%.
6. Y. Meng*, **A. Rumshisky**. Context-Aware Neural Model for Temporal Information Extraction. *Proceedings of ACL 2018*, Jul 16-18, 2018. Melbourne, Australia. Acceptance rate: 24.9%.
7. P. Potash*, R. Bhattacharya*, **A. Rumshisky**. Length, Interchangeability, and External Knowledge: Observations from Predicting Argument Convincingness. *Proceedings of IJCNLP 2017*. Taipei, Taiwan. Acceptance rate: 31%.
8. **A. Rumshisky**, M. Gronas, P. Potash*, M. Dubov, A. Romanov*, S. Kulshreshtha*, A. Gribov*. Combining Network and Language Indicators for Tracking Conflict Intensity. *Proceedings of SocInfo 2017*. Oxford, United Kingdom. Acceptance rate: 30%.
9. P. Potash*, **A. Rumshisky**. Towards Debate Automation: a Recurrent Model for Predicting Debate Winners. *Proceedings of EMNLP 2017*. Denmark, Copenhagen. (oral) Acceptance rate: 26%.
10. Y. Meng*, **A. Rumshisky**, A. Romanov*. Temporal Information Extraction for Question Answering Using Syntactic Dependencies in an LSTM-based Architecture. *Proceedings of EMNLP 2017*. Denmark, Copenhagen. Acceptance rate: 26%.
11. P. Potash*, A. Romanov*, **A. Rumshisky**. Here's My Point: Joint Pointer Architecture for Argument Mining. *Proceedings of EMNLP 2017*. Denmark, Copenhagen. Acceptance rate: 26%.
12. Y.-F. Luo*, **A. Rumshisky**. Interpretable Topic Features for Post-ICU Mortality Prediction. *Proceedings of AMIA 2016*. Chicago, IL. (oral)
13. W. Boag*, R. Campos*, K. Saenko, **A. Rumshisky**. MUTT: Metric Unit TesTing for Language Generation Tasks. *Proceedings of ACL 2016*. Berlin, Germany. Acceptance rate: 28%
14. P. Potash*, A. Romanov*, **A. Rumshisky**. GhostWriter: Using an LSTM for Automatic Rap Lyric Generation. (short paper) *Proceedings of EMNLP 2015*. Lisbon, Portugal. Acceptance rate: 22%
15. M. Ghassemi*, T. Naumann*, F. Doshi-Velez, N. Brimmer*, R. Joshi, **A. Rumshisky**, P. Szolovits. 2014. Unfolding Physiological State: Mortality Modelling in Intensive Care Units. *20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD2014)*. New York City, NY. Acceptance rate: 14.6%
16. M. Verhagen, I. Mani, R. Sauri, J. Littman, R. Knippen, S.B. Jang, **A. Rumshisky**, J. Phillips, J. Pustejovsky. Automating Temporal Annotation with TARSQI. (demo / short paper) *Proceedings of ACL 2005*, Ann Arbor, USA.
17. J. Pustejovsky, P. Hanks, **A. Rumshisky**. Automated Induction of Sense in Context. *Proceedings of COLING 2004*, Geneva, Switzerland. (oral)
18. J. Pustejovsky, **A. Rumshisky**, J. L. Moszkowicz, O. Batiukova. GLML: Annotating Argument Selection and Coercion. *Eighth International Conference on Computational Semantics (IWCS-8) 2009*. Tilburg, Netherlands.
19. J. Pustejovsky, J. Castaño, B. Cochran, M. Kotecki, M. Morrell, **A. Rumshisky**. Extraction and Disambiguation of Acronym-Meaning Pairs in Medline. *Medinfo*, 2001
20. R. Futrelle, **A. Rumshisky**. Discourse Structure of Text-Graphics Documents. *1st International Symposium on Smart Graphics*, March 21st-23rd 2001. Hawthorne, NY, USA.

JOURNAL PAPERS

1. **A. Rumshisky**, M. Ghassemi*, T. Neumann*, P. Szolovits, V. Castro, T. McCoy, R. Perlis. 2016. Predicting early

- psychiatric readmission with natural language processing of narrative discharge summaries. *Translational Psychiatry*. 2016 Oct 18. Impact factor: 5.620.
2. K. Zheng, V. Vydiswaran, Y. Liu, Y. Wang, A. Stubbs, O. Uzuner, A. Gururaj, S. Bayer, J. Aberdeen, **A. Rumshisky**, S. Pakhomov, H. Liu, H. Xu. 2015. Ease of adoption of clinical natural language processing software: An evaluation of five systems. *J Biomed Inform*, 2015 Jul 22. Impact factor: 2.482.
 3. W. Sun*, **A. Rumshisky**, Ö. Uzuner. 2015. Normalization of Relative and Incomplete Temporal Expressions in Clinical Narratives. *J Am Med Inform Assoc*. 2015 Apr 12. Impact factor: 3.932.
 4. R. Chasin*, **A. Rumshisky**, Ö. Uzuner, P. Szolovits. 2014. Word sense disambiguation in the clinical domain: a comparison of knowledge-rich and knowledge-poor unsupervised methods. *J Am Med Inform Assoc*. 2014 Jan 17. Impact factor: 3.932.
 5. W. Sun*, **A. Rumshisky**, Ö. Uzuner. 2013. Annotating temporal information in clinical narratives. *J Biomed Inform*. 2013 Jul 19. Impact factor: 2.482.
 6. W. Sun*, **A. Rumshisky**, Ö. Uzuner. 2013. Temporal reasoning over clinical text: the state of the art. *J Am Med Inform Assoc*. 2013 May 15. Impact factor: 3.932.
 7. W. Sun*, **A. Rumshisky**, Ö. Uzuner. 2013. Evaluating temporal relations in clinical text: 2012 i2b2 Challenge. *J Am Med Inform Assoc*. 2013 Apr 5. Impact factor: 3.932.
 8. S. Bratus, **A. Rumshisky**, A. Khrabrov, R. Magar, P. Thompson. Domain-Specific Entity Extraction from Noisy Unstructured Data Using Ontology-Guided Search. In Subramaniam, Lopresti, Roy and Schulz (eds.) Special Issue on Noisy Text Analytics, *International Journal of Document Analysis and Recognition (IJ DAR)*. Vol 14(2). Springer-Verlag, 2011.
 9. **A. Rumshisky**. 2008. Resolving Polysemy in Verbs: Contextualized Distributional Approach to Argument Semantics. In A. Lenci (ed.) *From Context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science. Special Issue of the Italian Journal of Linguistics. Rivista di Linguistica*, 20(1).

BOOKS/PROCEEDINGS

1. **A. Rumshisky**, K. Roberts, S. Bethard, and T. Naumann, editors. Proceedings of the Clinical Natural Language Processing Workshop (ClinicalNLP). The COLING 2016 Organizing Committee, Osaka, Japan. Dec 2016.
2. **A. Rumshisky**, N. Calzolari, editors. 5th International Conference on Generative Approaches to the Lexicon (GL2009). Pisa, Italy. Sept 2009.

PEER-REFEREED WORKSHOPS / MINOR CONFERENCES

1. P. Potash*, A. Romanov*, **A. Rumshisky**. 2018. Evaluating Creative Language Generation: The Case of Rap Lyric Ghostwriting. *2nd Workshop on Stylistic Variation, NAACL 2018*, Jun 5-6, 2018. New Orleans, USA.
2. Y-F. Luo*, W. Sun, **A. Rumshisky**. A Hybrid Method for Normalization of Medical Concepts in Clinical Narrative. *IEEE International Conference on Health Informatics and Health Information Technology 2018*, Jun 4-7 2018, New York City, NY.
3. Y. Meng*, **A. Rumshisky**, F. Sullivan, K. Keith*. Automatic Labeling of Problem-Solving Dialogues: an Example of Computational Microgenetic Analytics. *Proceedings of LREC 2018*. Miyazaki, Japan.
4. W. Boag, E. Sergeeva, S. Kulshreshtha, P. Szolovits, **A. Rumshisky**, T. Naumann. CliNER 2.0: Accessible and Accurate Clinical Concept Extraction. ML4H: Machine Learning for Health Workshop at NIPS 2017. Long Beach, CA.
5. P. Potash*, A. Romanov*, **A. Rumshisky**, M. Gronas. 2017. Tracking Bias in News Sources Using Social Media: the Russia-Ukraine Maidan Crisis of 2013–2014. *EMNLP 2017 Workshop "NLP Meets Journalism"*. Denmark, Copenhagen.
6. A. Romanov*, **A. Rumshisky**. 2017. Forced to learn: Discovering disentangled representations without exhaustive labels. *ICLR Workshop Track 2017*. Toulon, France.
7. P. Potash*, A. Romanov*, **A. Rumshisky**. 2017. SemEval-2017 Task 6: #HashtagWars: Learning a Sense of Humor. *The 11th International Workshop on Semantic Evaluation (SemEval-2017)*. ACL 2017, Aug 3-4, 2016. Vancouver, Canada.
8. D. Donahue*, A. Romanov*, **A. Rumshisky**. 2017. HumorHawk at SemEval-2017 Task 6: Mixing Meaning and Sound for Humor Recognition. *The 11th International Workshop on Semantic Evaluation (SemEval-2017)*. ACL 2017, Aug 3-4, 2016. Vancouver, Canada. **[1st out of 10 submissions, Task 6A]**
9. P. Potash*, **A. Rumshisky**. 2016. Recommender System Incorporating User Personality Profile through Analysis of Written Reviews. *Proceedings of the EMPIRE workshop at the ACM Conference on Recommender Systems (RecSys 2016)*. Boston, MA.
10. P. Potash*, W. Boag*, A. Romanov*, V. Ramanishka*, **A. Rumshisky**. SimiHawk at SemEval-2016 Task 1: A

- Deep Ensemble System for Semantic Textual Similarity. *The 10th International Workshop on Semantic Evaluation (SemEval-2016)*. NAACL HLT 2016, June 16-17, 2016. San Diego, California. [7th out of 115 submissions]
11. Y.-F. Luo*, **A. Rumshisky**, M. Gronas. Catching the Red Priest: Using Historical Editions of Encyclopedia Britannica to Track the Evolution of Reputations. *The 9th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities. ACL-IJCNLP 2015*.
 12. W. Boag*, P. Potash*, **A. Rumshisky**. 2015. TwitterHawk: A Feature Bucket Approach to Sentiment Analysis. *The 9th International Workshop on Semantic Evaluation (SemEval-2015)*. NAACL HLT 2015, May 31-Jun 5. Dever, Colorado. [1st out of 5 submissions, Task 10C]
 13. **A. Rumshisky**, R. Chasin*, P. Potash*. Bypassing knowledge acquisition bottleneck with Bayesian word sense induction. *NIPS 2013 Workshop on Machine Learning for Clinical Data Analysis and Healthcare*. (poster)
 14. R. Chasin*, **A. Rumshisky**. 2013. Mixing in Some Knowledge: Enriched Context Patterns for Bayesian Word Sense Induction. *Sixth International Workshop On Generative Approaches to the Lexicon (GL2013)*. Pisa, Italy.
 15. **A. Rumshisky**, N. Botchan*, S. Kushkuley*, J. Pustejovsky. Word Sense Inventories by Non-Experts. *LREC 2012*. Istanbul, Turkey.
 16. M. Gronas, **A. Rumshisky**, A. Gabrovski*, S. Kovaka*, H. Chen*. 2012. Tracking the History of Knowledge Using Historical Editions of Encyclopedia Britannica. *LREC 2012 Workshop on Adaptation of Language Resources and Tools for Processing Cultural Heritage Objects*. Istanbul, Turkey.
 17. M. Ghassemi*, T. Naumann*, R. Joshi, **A. Rumshisky**. 2012. Topic Models for Mortality Modeling in Intensive Care Units. *ICML Workshop on Machine Learning for Clinical Data Analysis, ICML 2012*, Edinburgh, Scotland.
 18. **A. Rumshisky**. Crowdsourcing Word Sense Definition. Fifth Linguistic Annotation Workshop (LAW-V). ACL-HLT2011. Portland, Oregon.
 19. J. Pustejovsky, **A. Rumshisky**, A. Plotnick, E. Jezek, O. Batiukova, V. Quochi. 2010. SemEval-2010 Task 7: Argument Selection and Coercion. Fifth International Workshop on Semantic Evaluation (SemEval-2010). Association for Computational Linguistics. Uppsala, Sweden.
 20. J. Pustejovsky, **A. Rumshisky**. SemEval-2010 Task 7: Argument Selection and Coercion. *NAACL HLT 2009 Workshop on Semantic Evaluations: Recent Achievements and Future Directions*. Boulder, Colorado, USA.
 21. **A. Rumshisky**, J. Moscowicz, M. Verhagen. The Holy Grail of Sense Definition: Creating a Sense-Disambiguated Corpus from Scratch. 5th International Workshop on Generative Approaches to the Lexicon (GL 2009). Pisa, Italy.
 22. **A. Rumshisky**, V. Grinberg. Using Semantics of the Arguments for Predicate Sense Induction. 5th International Workshop on Generative Approaches to the Lexicon (GL 2009). Pisa, Italy.
 23. S. Bratus, **A. Rumshisky**, R. Magar, P. Thompson. 2009. Using Domain Knowledge for Ontology-Guided Entity Extraction from Noisy, Unstructured Data. Third Workshop on Analytics for Noisy, Unstructured Text Data (AND 2009). Barcelona, Spain.
 24. **A. Rumshisky** and O. Batiukova. 2008. Polysemy in Verbs: Systematic Relations between Senses and their Effect on Annotation. *COLING Workshop on Human Judgement in Computational Linguistics (HJCL-2008)*. Manchester, England.
 25. **A. Rumshisky**, V. A. Grinberg, J. Pustejovsky. Detecting Selectional Behavior of Complex Types in Text. *Fourth International Workshop on Generative Approaches to the Lexicon*. P. Bouillon, L. Danlos, and K. Kanzaki (eds.) 2007. Paris, France.
 26. C. Havasi, **A. Rumshisky**, J. Pustejovsky. An Evaluation of the Brandeis Semantic Ontology. *Fourth International Workshop on Generative Approaches to the Lexicon*. P. Bouillon, L. Danlos, and K. Kanzaki (eds.) 2007. Paris, France.
 27. B. Wellner, J. Pustejovsky, C. Havasi, **A. Rumshisky**, R. Sauri. Classification of Discourse Coherence Relations: An Exploratory Study using Multiple Knowledge Sources. *7th SIGdial Workshop on Discourse and Dialogue*, Sydney, Australia, 15-16 July 2006.
 28. **A. Rumshisky**, P. Hanks, C. Havasi, J. Pustejovsky. Constructing a Corpus-based Ontology using Model Bias. *FLAIRS 2006*, Melbourne Beach, Florida.
 29. **A. Rumshisky**, J. Pustejovsky. Inducing Sense-Discriminating Context Patterns from Sense-Tagged Corpora. *LREC 2006*, Genoa, Italy.
 30. J. Pustejovsky, C. Havasi, J. Littman, **A. Rumshisky**, M. Verhagen. Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. *LREC 2006*, Genoa, Italy.
 31. J. Pustejovsky, **A. Rumshisky**, J. Castaño. Rerendering Semantic Ontologies: Automatic Extensions to UMLS through Corpus Analytics. *LREC 2002 Workshop on Ontologies and Lexical Knowledge Bases*. Las Palmas, Canary Islands, Spain.
 32. J. Pustejovsky, J. Castaño, R. Sauri, **A. Rumshisky**, J. Zhang, W. Luo. Medstract: Creating Large-scale Information Servers for Biomedical Libraries. *ACL 2002 Workshop on Natural Language Processing in the Biomedical Domain*. Philadelphia, PA.

BOOK CHAPTERS

1. **A. Rumshisky**, A. Stubbs. Machine Learning for Higher-Level Linguistic Tasks. In *Handbook of Linguistic Annotation*, ed. N. Ide and J. Pustejovsky. Berlin: Springer, 2015.
2. J. Pustejovsky, **A. Rumshisky**, O. Batiukova, J. Moszkowicz. Annotation of Compositional Operations with GLML. In H. Bunt, J. Bos and S. Pulman (eds.), *Computing Meaning, Volume 4*. Springer-Verlag, 2014, pp. 217-234.
3. J. Pustejovsky and **A. Rumshisky**. 2010. Mechanisms of Sense Extension in Verbs. In G-M de Schryver (ed.) *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers.
4. J. Pustejovsky and **A. Rumshisky**. 2008. Between Chaos and Structure: Interpreting Lexical Data through a Theoretical Lens. In P. Hanks (ed.) *Special Issue of the International Journal of Lexicography on the legacy of John Sinclair*, Oxford University Press.

CONFERENCE PRESENTATIONS

1. M. Gronas, **A. Rumshisky**, A. Romanov*. Clouds, Rivers, and Mountains of Knowledge: Visualizing the History of Knowledge in Large Historical Datasets. *The 49th Annual Convention of the Association for Slavic, Eastern European, and Eurasian studies (ASEEES 2017)*. Chicago, IL.
2. W. Boag*, K. Wacome*, T. Naumann*, **A. Rumshisky**. CLiNER: A Lightweight Tool for Clinical Named Entity Recognition. *AMIA Joint Summits on Clinical Research Informatics 2015*. San Francisco, CA.
3. R. Chasin*, **A. Rumshisky**, Ö. Uzuner, P. Szolovits. Using UMLS for Word Sense Disambiguation in Clinical Notes. *AMIA 2012*. Chicago, IL.
4. **A. Rumshisky**, K. Buford*, I. Goldstein, Ö. Uzuner. Multi-Scrubber: An Ensemble System for De-identification of Protected Health Information. *SHARPN Summit on Secondary Use of EHR*. 2012. Rochester, MN.
5. **A. Rumshisky**, R. Chasin*. Knowledge-Based vs. Bottom-up Methods for Word Sense Disambiguation in Clinical Notes. *SHARPN Summit on Secondary Use of EHR*. 2012. Rochester, MN.
6. M. Gronas, **A. Rumshisky**, S. Bratus. Digital Pushkin: Computational visualizations of Pushkin's social and political networks. *41st National Convention of the American Association for the Advancement of Slavic Studies (AAASS 2009)*. Boston, MA

INVITED TALKS

1. Modeling Long-Distance Dependencies and Modular Representations for Natural Language Processing. *Computational Linguistics Seminar*, University of Pennsylvania, Oct 2018.
2. Understanding Human Language. *Women in Tech (WiT)*. Tufts University, Sep 2018.
3. Understanding Human Language. *AI4All Educational Program*. Boston University, Jul 2018.
4. Modeling Long-Distance Dependencies and Modular Representations for Natural Language Processing. *Conversational Intelligence Summer School*, Moscow Institute of Physics & Technology, Jul 2018.
5. Modeling Long-Distance Dependencies and Modular Representations for Natural Language Processing. Boston University, Dept of Computer Science, Apr 2018.
6. Learning Computational Representations of Meaning – What We Can and Can Not Do. MIT Brain and Cognitive Sciences, Mar 2018.
7. Can We Do Language? Context-Aware Modeling of Long-Distance Relations and Modularity of Linguistic Representation. MIT CSAIL, Mar 2018.
8. Sequence Modeling for Discourse-Level Tasks. UMass Medical School, Worcester, MA, Nov 2017.
9. Adventures in Sequence Modeling: Dissecting Complex Natural Language Tasks. *Analysis in Motion Seminar Series*. Pacific Northwest National Lab, Apr 2017 .
10. Humour, Argument, Semantics: Learning for Complex Natural Language Tasks. *Image and Video Computing Seminar Series*. Boston University, Boston MA, Oct 2016 .
11. Making Sense of Doctor's Notes: Lexical Disambiguation in the Clinical Domain. *Computational Linguistics Seminar Series*. Brandeis University, Waltham, MA. Apr 2013.
12. Five Lectures in Computational Lexical Semantics. *Summer School at the Center for Chinese Linguistics*, Peking University, Beijing, China. Aug 2011.
13. "What Does This Word Mean?" Word Sense Definition and Polysemy Resolution in Context. Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA. Jan 2011.
14. Robust Semantic Annotation of Compositional Relations in Text. *Copenhagen Dependency Treebank Workshop*, Copenhagen Business School, Copenhagen, Denmark. Aug 2010.
15. "What Does This Word Mean?" Word Sense Definition and Polysemy Resolution in Context. Human Language Technology Center of Excellence at John Hopkins University, Baltimore, MD. Jul 2010.

16. Generative Lexicon Markup Language Annotation: Specification and Interface. *GLML Shared Task Planning Meeting*. Istituto di Linguistica Computazionale del CNR, Pisa, Italy. Sep 2008.
17. Sense Induction: Modeling Selection with Contextualized Similarity. International Computer Science Institute (ICSI), Berkley, CA Dec 2006.
18. Selecting for semantic distinctions in a corpus-driven ontology of semantic types. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, Germany. Feb 2004.

INVITATION-ONLY SEMINARS AND RESEARCH VISITS

1. **Unified Annotation Tooling Workshop** (NSF-sponsored). Sunny Isles, Florida, Mar 2015.
2. **Dagstuhl Seminar** on Computational Models of Meaning in Context, Dagstuhl, Germany, Nov 2013
3. **Copenhagen Dependency Treebank Workshop**, Copenhagen Business School, Copenhagen, Denmark. Aug 2010.
4. **Collaborative visits:**
 1. Institute for Formal and Applied Linguistics, Charles University, Prague, Czech Republic. Aug 2009.
 2. Istituto di Linguistica Computazionale del CNR, Pisa, Italy. Sep 2008.
 3. Department of Informatics, Masaryk University, Brno, Czech Republic. Dec 2005.
 4. Berlin-Brandenburgische Akademie der Wissenschaften, Berlin, Germany. Feb 2004.

SERVICE

PROFESSIONAL ACTIVITIES

- **Panelist/Reviewer (funding agencies):** NSF IIS/PAC CompCog 2018, NIH/NIMH 2018, NSF IIS CRII 2017, NIH/NIMH 2016, NSF IIS 2016; NSF IIS 2017, NSF Smart and Connected Health 2015
- **Organizer/Co-Organizer (conferences/workshops/shared tasks)**
 - Organizer, SemEval-2019 Task 11: Normalization of Medical Concepts in Clinical Narrative (Jun 2018–Present)
 - Organizer, SemEval-2017 Task 6: #HashtagWars: Learning a Sense of Humor (Jun 2016–Apr 2017)
 - Organizer, Clinical Natural Language Processing Workshop at COLING 2016, Dec 2016
 - Organizer, tutorial on *Distributional Compositional Semantics in the Age of Word Embeddings: Tasks, Resources and Methodology* at LREC 2018. Attended by 63 people.
 - Organizer, tutorial on *Deep Semantic Annotation with Shallow Methods* at LREC 2014. Attended by 50 people.
 - Co-organizer, 2012 i2b2 Shared Task and Workshop on Challenges in NLP for Clinical Data
 - Co-organizer, *5th International Conference on Generative Approaches to the Lexicon (GL2009)*
 - Co-organizer, *SemEval-2010 Task 7* (Jun 2009–Mar 2010)
- **Area Chair:** COLING 2018, NAACL-HLT 2019 (area: *Machine Learning for NLP*)
- **Student Volunteer Coordinator and Scholarship Co-Chair:** EMNLP 2018
- **University host** for the North American Computational Linguistics Olympiad (NACLO), 2017, 2018
- **Reviewer/PC (conferences):** ACL 2018, NAACL-HLT 2018, LREC 2018, COLING 2016, SemEval 2016, AMIA 2016, TSD 2016, ACL 2013, ACL 2014, COLING 2014, EMNLP 2009, TSD 2010
- **Reviewer/Referee (journals):** Journal of Language Modeling (JLM), 2017; Journal of Machine Learning, 2017; Language Resources and Evaluation, 2016; Language and Linguistics, 2016; PLOS, 2015; Journal of the American Medical Informatics Association, 2012, 2013, 2017; BMC Bioinformatics, 2012; Journal of Artificial Intelligence Research (JAIR), 2012.
- **Technical Reviewer (publishers):** O'Reilly Media Inc 2012, Routledge 2013

SOCIETIES AND MEMBERSHIPS

- Association for Computational Linguistics (ACL)
- American Medical Informatics Association (AMIA)

MENTORING ACTIVITIES

- **Mentor**, Massachusetts AWIS Mentoring Circle, Assoc. for Women in Science (Fall 2017 – Spring 2018)
- **Invited panelist**, Young Academics Panel for Graduate Women in Science and Engineering (GWISE), Northeastern University (Feb 2016)
- **Organizer**, Mini-Panel for Graduate Students: Industry Job Search and How to Succeed in it. Dept. of Computer Science, UMass Lowell (June 2017)
- **Organizer**, Mini-conference and recruiting visit from Microsoft Research Maluuba. Dept of Computer

Science, UMass Lowell. October 31, 2017.

DEPARTMENTAL SERVICE

- Faculty coordinator for the Data Science Option for Computer Science majors (Aug 2016 – Present)
- Faculty Search Committee (September 2017 – May 2018; January 2017 – April 2017)
- Graduate Committee (September 2013 – Present)
- Department representative for the UMass Lowell Commonwealth Honors Council (Jan 2013–Nov 2015)

INTERNSHIPS AND CONSULTING

2013 – 2014 **QPID Health**, Boston, MA – information extraction for clinical text from electronic health records
2012 – 2013 **Blondinka.ru**, Moscow, Russia – automatic generation of ads for a search engine
2009 **Intermedia.net, Inc**, New York, NY – content strategies for improving customer conversion
2007 **Dartmouth College**, Hanover, NH – entity normalization in General Motors' mechanics' notes
2005 **Google**, Mountain View, CA – automatic quality rankings for the application translation console
2004 – 2005 **Synesis7**, Butte, MT – entity markup algorithms for digitized Boeing manuals

TEACHING AND EVALUATION

CURRICULUM DESIGN

- (UMass Lowell) **Data Science Option for B.S. in Computer Science** 2013 – Present
co-led the effort to create – and currently serving as faculty lead – for a Data Science option for computer science majors, with a more in-depth study of mathematics, and core courses in machine learning, cloud computing, natural language processing, and computer vision.

INSTRUCTOR/LECTURER

- COMP 2010 (UMass Lowell) **Object-Oriented Programming in C++ Honors section** 2013 – 2017
undergraduate course, part of the core sequence for majors
(out of 5) average instructor rating: 3.87, average course rating: 3.67
- COMP 4420/5420 (UMass Lowell) **Introduction to Natural Language Processing** 2012 – 2018
jointly taught graduate / advanced undergraduate course
(out of 5) average instructor rating: 4.16, average course rating: 4.14
- COMP 4600/5300 (UMass Lowell) **Topics: Advanced Natural Language Processing** 2015, 2016
jointly taught graduate / advanced undergraduate course
(out of 5) average instructor rating: 4.31, average course rating: 4.27
- (UMass Lowell) **Computer Science Research** 2014, 2017
individual directed study course for B.S., M.S., and Ph.D. students
- LING 190b (Brandeis University) **Quantitative Methods in Corpus Linguistics** 2009
An NLP course in the Master's program in Computational Linguistics; cross-listed as an elective for graduate and undergraduate computer science students.

TEACHING ASSISTANT/ASSISTANT INSTRUCTOR (Brandeis University)

- Undergraduate-level courses
Discrete Structures (CS 29a, Fall 2001)
Theory of Computation (CS 30a, Spring 2005)
Internet and Society (CS 33b, Spring 2002)
- Graduate/Advanced undergraduate electives:
Cryptology: Cryptography and Cryptanalysis (CS 171a, Fall 2002)
Topics in Computational Linguistics (CS 114, Spring 2006)
Programming Language Theory (CS 190a, Fall 2003)
Topics in Computer Systems: Computational Lexicography (CS 120a, Spring 2003)

ADVISING

CURRENT ADVISEES

- Anna Rogers (Postdoc, summer 2017 – Present)
- Alexey Romanov (Ph.D. in Computer Science expected in December 2019, UMass Lowell)
- Yuanliang Meng (Ph.D. in Computer Science expected in December 2019, UMass Lowell)
- Yen-Fu Luo (Ph.D. in Computer Science expected in December 2018, UMass Lowell)
- Saurabh Kulshreshta (Ph.D. in Computer Science expected in May 2021, UMass Lowell)
- Olga Kovaleva (Ph.D. in Computer Science expected in May 2021, UMass Lowell)
- David Donahue (B.S. in Computer Science expected in May 2019, UMass Lowell), Fall 2016 – Present

GRADUATED PHDS

- Peter Potash (Ph.D. in Computer Science awarded in Aug 2017, UMass Lowell) – now at Microsoft Research Montreal, Canada
- Yen-Fu Luo (Ph.D. in Computer Science awarded in Nov 2018)

PHD DISSERTATION COMMITTEE

- Weiyi Sun (Ph.D. in Informatics awarded in Dec 2014, co-supervised @ SUNY Albany)
- Tristan Naumann (Ph.D. in Computer Science awarded in May 2018, Massachusetts Institute of Technology)
- Julie Blumenthal (Ph.D. in Computer science, expected in Dec 2019, UMass Lowell)

MASTERS THESES

- Rachel Chasin (M. Eng awarded in May 2013, co-supervised @ MIT CSAIL)
 - first author in J. Am. Med. Inf. (impact factor 3.9)
- Nick Botchan (MA in Computational Linguistics, awarded in May 2012, Brandeis University)
 - co-author of a paper in LREC 2012
- Igor Finkelsteyn (MA in Computational Linguistics, awarded in May 2010, Brandeis University)

UNDERGRADUATE HONORS THESES

- Kenneth Sanders (B.S. in Computer Science, awarded Aug 2017, UMass Lowell)
- William Boag (B.S. in Computer Science with Data Science Option, May 2016, UMass Lowell) – now a Ph.D. student at MIT EECS
 - recipient of the NSF Graduate Fellowship in 2015; first author of a paper at ACL 2016; first author of a SemEval 2015 workshop paper for a system that placed 1st in Task 10C; co-author on two other workshop papers.
- Kevin Wacome (B.S. in Computer Science, Aug 2016, UMass Lowell) – now at MITRE Corporation
 - co-author on a poster presentation at AMIA Joint Summits on Clinical Research Informatics 2015

DIRECTED STUDY AND INTERNS

- Matthew Downey, B.S. in Electrical Engineering expected in May 2021
- Kevin Huang, B.S. in Computer Science expected in May 2021
- Rahul Krishnan (high-school intern)
- David Donahue (B.S. in Computer Science expected in May 2019, UMass Lowell) , Fall 2016 – Present;
 - first author of a SemEval 2016 workshop paper for a system that placed 1st in Task 6A
- Robin Bhattacharya (B.S. in Computer Science expected 2018, UMass Lowell), Summer 2017– Present
 - co-authored a paper to at IJCNLP 2017.
- Renan Campos (B.S. in Computer Science expected in May 2018, UMass Lowell), Fall 2015–Spring 2017
 - co-authored a paper at ACL 2016
- Alex Gribov (B.S. In Computer Science awarded in May 2017, UMass Lowell), Summer 2017
 - co-authored a paper at SocInfo 2017
- Shashwath Hosur Ananthakrishna (M.S. in Computer Science awarded in May 2018), Fall 2017
 - co-authored a paper at COLING 2018
- Simthyrearch Dy (B.S. in Computer Science expected in May 2018), Fall 2017
- Connor Cooper (B.S. in Computer Science expected in May 2018, UMass Lowell), Fall 2015 – Fall 2016
- John Cavalieri (B.S. in Computer Science expected May 2015, UMass Lowell), Spring 2014
- Amith Nair (M.S. in Computer Science, May 2014, UMass Lowell) , Spring 2014
- Jon Madden (B.S. in Computer Science, May 2014, UMass Lowell), Spring 2014

VISITING STUDENTS

- He Ren (Visiting PhD student from Peking University PKU, Sep 2015 – Aug 2016)
- Shufan Wang (summer intern, B.S. In Trinity College expected in May 2018, Summer 2017)

UNDERGRADUATE HONORS AND AWARDS

Moscow State University Honor Scholarship
Undergraduate Research Program Grant, Brandeis University
Dean's List, Brandeis University

HONOR SOCIETIES

Upsilon Pi Epsilon Computer Science Honor Society
Psi Chi National Honor Society

PERSONAL

U.S. Citizen

Languages: English/Russian bilingual; reading French; beginner German, Mandarin

MEDIA COVERAGE

- USA TODAY. *Clinton vs. Trump: Everything you need to know about the first debate*. Paul Singer. Published 09/24/2016
- European Hospital. *Disambiguation*. Published 03/04/2013. <http://healthcare-in-europe.com>
- Dark Daily. *New solutions for unstructured data may help with clinical laboratory and anatomic pathology data*. Published 01/07/2013. <http://www.darkdaily.com>
- MIT News. *Mining physicians' notes for medical insights*. Larry Hardesty. Published 10/31/2012.