

Smaranda Muresan
- CURRICULUM VITAE -

Office Address Data Science Institute, Columbia University
475 Riverside Drive MC 7717, New York, NY 10115

Office Phone +1 212-870-1294

Email smara@columbia.edu

WWW <http://www.cs.columbia.edu/~smara/>

RESEARCH INTERESTS

Natural Language Processing: context-dependent language understanding (language and vision; argumentation mining) with applications to computational social science and education

EDUCATION

1999 B.Sc. in Computer Science and Engineering, *Cum Laude*
Technical University of Cluj-Napoca, Romania

2002 M.Sc. in Computer Science
Columbia University, New York, NY

2006 Ph.D. in Computer Science
Columbia University, New York, NY

PROFESSIONAL EMPLOYMENT POST PhD

Fall 2017 - Research Scientist, Data Science Institute & Computer Science Department, Columbia University

2013-2017 Research Scientist, Center for Computational Learning Systems, Columbia University

2008 - 2013 Assistant Professor, Library and Information Science Department, School of Communication and Information, Rutgers University, New Brunswick, NJ.

2010 - 2013 Graduate Faculty, Computer Science Department, Rutgers University

2010 -2013 Co-director of the Laboratory for the Study of Applied Language Technologies and Society, Rutgers University.

2006 - 2008 Postdoctoral Research Associate, University of Maryland Institute for Advanced Computer Studies, College Park

GRANTS AND AWARDS

2013 *Distinguished Achievements in Research Award*, LIS Department, Rutgers University

2017-2021 Co-PI on the IARPA proposal "SCRIPTS: System for Cross-language Information Processing, Translation, and Summarization" for the MATERIAL program.

2014-2018 PI at Columbia on NSF award "RI: Medium: Collaborative Research: Write A Classifier: Learning Fine-Grained Visual Classifiers from Text and Images" (\$463,208). PI at Rutgers Ahmed Elgammal.

2014-2017	Co-PI on NSF award "Reducing Racial and Gender Achievement Gaps in STEM: Use of Natural Language Processing to Understand Why Affirmation Interventions Improve Performance"(\$1,007,993).PI Valerie Purdie Vaughn (Columbia University), Co-PIs Geoffrey Cohen (Stanford University), Jonathan Cook (Penn State University).
2014-2016	Co-PI on a ROADS award from Columbia University's Institute for Data Sciences and Engineering: "Reducing Racial and Gender Achievement Gaps in STEM: Use of Natural Language Processing to Understand Why Affirmation Interventions Improve Performance". PI Valerie Purdie-Vaughns (\$199,998)
2012-2017	Co-PI on a DARPA award "DRATS: Detecting Relations and Anomalies in Text and Speech" for the DEFT program.(total \$5,499,999). PI Owen Rambow, co-PI Kathy McKeown, Julia Hirschberg, Mona Diab; Mari Ostendorf (PI at University of Washington)
2011-2014	Co-PI, NSF "Collaborative Research: Teaching Computers to Follow Verbal Instructions" (2011-2014) (total award \$999,925), PI Michael Littman, co-PI Marie desJardins
2008-2010	PI at Rutgers on subaward from UMD (PI Philip Resnik) for NSF SGER award, "Exploiting Alternative Packagings of Source Meaning in Statistical Machine Translation" (total grant award \$144,763)
2011	Co-PI, Rutgers Office of the Vice President for Research and Economic Development to build the "Laboratory for the Study of Applied Language Technologies and Society (SALTS)" (Internal Grant: \$112,000)

PROFESSIONAL ACTIVITIES

Co-Editor	Special Issue of IJAIED: MARWIDE: Multidisciplinary Approaches to Reading and Writing Integrated with Disciplinary Education (2017)
Area Co-Chair	Sentiment Analysis Track at NAACL 2018
Area Co-Chair	Information Extraction and NLP Applications Track at ACL 2017
Area Co-Chair	Computational Social Science and Social Media Track at EMNLP 2016
Publicity Chair	The 25th International Joint Conference on Artificial Intelligence (IJCAI-16)
Co-Organizer	The 2nd Workshop on Argumentation Mining (ArgMining 2015) at NAACL-HLT 2015
Co-Organizer	Multidisciplinary Advances in Reading and Writing for Science Education Workshop (NSF funded workshop)
Faculty Advisor	Student Research Workshop at NAACL-HLT 2015
Area Co-Chair	Second Joint Conference on Lexical and Computational Semantics (*SEM 2013)
Publicity Chair	NAACL-HLT 2012 (The 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies)
Co-Editor	Special Issue on "Network-based Models of Cognitive and Social Dynamics of Language". Computer Speech and Language. Volume 25 (3), 2011.
Co-organizer	TextGraphs-4: Graph-based Methods for Natural Language Processing. Workshop at the Association for Computational Linguistics Conference (ACL/IJCNLP 2009)

NSF Panelist	2014, 2011, 2009, 2016
Expert Evaluator	Killam Research Fellowships, Canada Council (http://canadacouncil.ca/prizes/killam/) (by invitation,06/2010-09/2010).
Panelist	"What Its Like to Work at My Institution" Faculty Panel at "Preparing Future Faculty Workshop" at Columbia University on April 16, 2010.

PATENTS

2003	<i>System and Method of Generating Dictionary Entries</i> . Judith Klavans and Smaranda Muresan, Columbia University, New York, NY. Serial No 398,535.
------	--

INVITED TALKS

9/27/2016	<i>Argument Mining for Assessing Students' Persuasive Essays</i> . Workshop on "Connecting Language, Interaction and Education in Digital Environments", Penn State University.
9/22/2016	<i>Computational Models of Understanding Language in Social Context</i> at the Social Computing Workshop organized by the Army Research Lab.
4/19/2016	<i>Argumentation Mining In Online Interactions: Opportunities and Challenges</i> . Invited Talk at Dagstuhl Seminar on "Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments".
3/16/2016	<i>Computational Models of Understanding Language in Context</i> . Invited Talk at Hearst Corporation Meeting with Hearst's Executive Director of R&D and Director of Data Science.
2/24/2014	<i>Identification of Non-literal Language in Social Media: A Case Study on Sarcasm</i> . Yahoo! Research Lab, NYC
5/22/2014	<i>Investigating the Content of Values-Affirmation Interventions: A Data-Driven Approach to Discovering Topics in Affirmation Essays</i> . APS Symposium <i>Computational Methods for Linguistic Insights into Personality and Behavior</i> , May 22, 2015.
7/30/2012	<i>Context-dependent Language Understanding</i> . Rutgers's Big Data Meeting.
2/9/2012	<i>Identifying Sarcasm in Twitter</i> . Columbia University Speech Group Talks (joined work with Roberto Gonzalez Ibanez and Nina Wacholder).
11/11/2011	<i>Learning for Deep Language Understanding</i> . CUNY Graduate Center.
07/01/2010	<i>Learnable Constraint-based Grammars for Deep Language Understanding</i> . Educational Testing Services, Princeton, NJ.
05/27/2010	<i>Learnable Constraint-based Grammars for Deep Language Understanding</i> . IBM T.J. Watson Research Center
04/13/2010	<i>Learnable Constraint-based Grammars for Deep Language Understanding</i> . Yahoo Machine Learning Seminar, Department of Computer Science, Rutgers University.
10/29/2009	<i>Are those fighting words? The expression of opposition in online text</i> Ischool, Syracuse University.(together with Nina Wacholder).
08/05/2009	<i>National Language Processing at Rutgers</i> National University of Singapore.

- 10/02/2008 *A Computational Model for Language Learning and Understanding*. The Rutgers University Center for Cognitive Science, Rutgers University.
- 02/26/2008 *Learning Constraint-Based Grammars from Representative Data*. The Center for Language and Speech Processing, Johns Hopkins University.
- 03/01/2006 *Inducing Constraint-based Grammars using a Domain Ontology*. Computational Linguistics and Information Processing Laboratory (CLIP) at the University of Maryland Institute for Advanced Computer Studies (UMIACS).
- 10/22/2004 *Learning Constraint-based Grammars using a Small Semantic Treebank*. Computational Linguistics Group, Department of Computer Science, University of Toronto.

PUBLICATIONS

Elena Musi, Mark Aakhus, Smaranda Muresan, Andrea Rocci, Manfred Stede (2017). From Theory to practice: the annotation of argument schemes. *Proceedings of the European Conference on Argumentation (ECA)*.

Debanjan Ghosh, Alexander Fabri, Smaranda Muresan (2017). The Role of Conversation Context for Sarcasm Detection in Online Interactions. *Proceedings of SIGDIAL*. (best paper award)

Elena Musi, Debanjan Ghosh, Mark Aakhus, Smaranda Muresan, Nina Wacholder (under review). Building an ontology of (dis)agreement space for argument mining. *Position paper for the SIGDIAL/SEMDIAL 2017 Joint Special Session on Negotiation Dialog*.

Debanjan Ghosh, Aquila Khanam, Smaranda Muresan (2016). Coarse-grained Argumentation Features for Scoring Persuasive Essays. *Proceedings of ACL 2016 (short papers)*

Elena Musi, Debanjan Ghosh, Smaranda Muresan (2016). Towards Feasible Guidelines for the Annotation of Argument Schemes. *Proceedings of the Argumentation Mining Workshop at ACL 2016*.

Olivia Winn, Madhavan Kidambi, Smaranda Muresan (2016). Detecting Visually Relevant Sentences for Fine-Grained Image Classification. *Proceedings of the Vision and Language Workshop at ACL 2016*

Smaranda Muresan, Roberto Gonzalez-Ibanez, Debanjan Ghosh, Nina Wacholder (2016). Identification of Non-literal Language in Social Media: A Case Study on Sarcasm. *Journal of the American Society for Information Science and Technology (JASIST)*.

Debanjan Ghosh, Weiwei Guo, Smaranda Muresan (2015). Sarcastic or Not: Word-Embeddings to Predict the Literal or Sarcastic Meaning of Words. *In Proceedings of EMNLP , 2015*, Lisbon, Portugal.

J. MacGlashan, M. Babes-Vroman, M. desJardins, M. Littman, S. Muresan, S. Squire, S. Tellex, D. Arumugam, and L. Yang (2015). Grounding English commands to reward functions. *In Robotics: Science and Systems , 2015*

Travis Riddle, Sowmya Bhagavatula, Weiwei Guo, Smaranda Muresan, Geoff Cohen, Jonathan Cook and Valerie Purdie-Vaughns (2015). Mining a Written Values Affirmation Intervention to Identify the Unique Linguistic Features of Stigmatized Groups. *Proceedings of Educational Data Mining, 2015*.

Debanjan Ghosh, Smaranda Muresan, Nina Wacholder, Mark Aakhus, Matthew Mitsui (2014). Analyzing Argumentative Discourse Units in Online Interactions. *First Workshop on Argumentation*

Mining at ACL 2014

Gaurav Kharkwal, Smaranda Muresan (2014). Surprisal as a Predictor of Essay Quality. *The 9th Workshop on Innovative Use of NLP for Building Educational Applications*, at ACL 2014

Smaranda Muresan (2013). Ontology-based Semantic Interpretation via Grammar Constraints. In A. Oltramari, L. Qin, P. Vossen, E. Hovy (eds) *New Trends of Research in Ontologies and Lexical Resources*. Springer Series on Theory and Applications of Natural Language Processing.

Nina Wacholder, Debanjan Ghosh, Mark Aakhus, Smaranda Muresan (2014). Annotating Multiparty Discourse: Challenges for Agreement Metrics. *The 8th Linguistic Annotation Workshop (LAW VIII 2014)*, at COLING 2014.

Smaranda Muresan, Judith Klavans (2013). Inducing Terminologies from Text: A Case Study for the Consumer Health Domain. *Journal of the American Society for Information Science and Technology (JASIST)*., 64(4); pp 727-744.

Mark Aakhus, Smaranda Muresan, Nina Wacholder (2013). Integrating Natural Language Processing and Pragmatic Argumentation Theories for Argumentation Support. *Proceedings of the Ontario Society for the Study of Argumentation: VIRTUES of ARGUMENTATION (OSSA 2013)*.

Debanjan Ghosh, Smaranda Muresan (2012). Relation Classification using Entity Sequence Kernels. *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)* (short paper - 8 pages) [25% acceptance rate]

Hao Li, Yu Chen, Heng Ji, Smaranda Muresan, Dequan Zheng (2012). Combining Social Cognitive Theories with Linguistic Features for Multi-genre Sentiment Analysis. *Proceedings of 26th Pacific Asia Conference on Language, Information, and Computation (PACLIC 26)*. [best paper nomination]

Smaranda Muresan (2012). Search Space Properties for Learning a Class of Constraint-based Grammars. *Proceedings of TAG+11: The 11th International Workshop on Tree Adjoining Grammars and Related Formalisms*.

Choonkyu Lee, Smaranda Muresan, Karin Stromswold (2012). Analysis of Referring Expressions in Narrative Discourse. *NAACL-HLT 2012 Workshop on Computational Linguistics for Literature* [30% acceptance rate]

James MacGlashan, Monica Babeş-Vroman, Kevin Winner, Ruoyuan Gao, Marie desJardins, Michael Littman, Smaranda Muresan (2012). Learning to Interpret Natural Language Instruction. *Proceedings of AAAI-2012 Workshop on Grounding Language for Physical Systems*. (extended version of NAACL-HTL Workshop paper).

Monica Babeş-Vroman, James MacGlashan, Ruoyuan Gao, Kevin Winner, Marie desJardins, Michael Littman, Smaranda Muresan (2012). Learning to Interpret Natural Language Instructions. *Proceedings of NAACL-HTL Workshop on Semantic Interpretation in an Actionable Context*.

Smaranda Muresan (2011). "Learning for Deep Language Understanding". *Proceedings of 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)* [17% acceptance rate]

Roberto Gonzalez-Ibanez, Smaranda Muresan and Nina Wacholder (2011). Identifying sarcasm in Twitter: a Closer Look. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers* [25% acceptance rate].

Monojit Choudhury, Samer Hassan, Animesh Mukherjee, Smaranda Muresan. Editorial: Network based Models of Cognitive and Social Dynamics of Human Languages. *Computer Speech and*

Language: Special Issue on Network-based Models of Cognitive and Social Dynamics of Language. Volume 25 (3), July 2011.

Smaranda Muresan (2010). Ontology-based Semantic Interpretation as Grammar Rule Constraints. *Proceedings of The 11th International Conference on Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science.* 6008, Springer. [22.5% acceptance rate].

Smaranda Muresan (2010). A Learnable Constraint-based Grammar Formalism. *Proceedings of 23rd International Conference on Computational Linguistics (COLING) 2010.* [22.2% acceptance rate]

Philip Resnik, Chris Dyer, Smaranda Muresan (2010). Lattice Decoding for Hierarchical Models. In Joe Olive, ed., *Handbook of Natural Language Processing and Machine Translation*, Springer.

Smaranda Muresan (2008). Learning to Map Text to Graph-based Meaning Representations via Grammar Induction. *Proceedings of the COLING 2008 Workshop TextGraphs-3: Graph-based Algorithms for Natural Language Processing.*[31.5% acceptance rate]

Christopher Dyer, Smaranda Muresan and Philip Resnik (2008). Generalizing Word Lattice Translation. *Proceedings of the Association for Computational Linguistics, ACL 2008.* [25% acceptance rate]

Smaranda Muresan and Owen Rambow (2007). Grammar Approximation by Representative Sublanguage: A New Model for Language Learning. *Proceedings of the Association for Computational Linguistics, ACL 2007, Prague, Czech Republic.* [22% acceptance rate]

Smaranda Muresan (2004). Inducing Constraint-based Grammars using a Domain Ontology. *Proceedings of the Ninth AAI/SIGART Doctoral Consortium.*

Smaranda Muresan, Tudor Muresan, and Judith Klavans (2004). Inducing Constraint-based Grammars from a Small Semantic Treebank. *AAAI Spring Symposium on Language Learning: An Interdisciplinary Perspective*, Stanford University.

Smaranda Muresan, Samuel D. Popper, Peter T. Davis, and Judith L. Klavans (2003). Building a Terminological Database from Heterogeneous Definitional Sources. *Proceedings of the National Conference on Digital Government Research*, Boston, Massachusetts.

Smaranda Muresan and Judith L. Klavans (2002). A Method for Automatically Building and Evaluating Dictionary Resources. *Proceedings of the Language Resources and Evaluation Conference (LREC 2002).*

Smaranda Muresan, Tudor Muresan, and Rodica Potolea (2002). Data Flow Coherence Constraints for Pruning the Search Space in ILP Tools. *International Journal of Artificial Intelligence Tools*, volume 11(2), 2002.

Judith Klavans and Smaranda Muresan (2001). Evaluation of DEFINDER: A System to Mine Definitions from Consumer-oriented Medical Text. *Proceedings of The ACM+IEEE Joint Conference on Digital Libraries (JCDL 2001).*

Smaranda Muresan, Tudor Muresan, and Rodica Potolea (2001). Data flow criteria in ILP tools. *Proceedings of IEEE 13th International Conference on Tools with Artificial Intelligence (ICTAI 2001).*

Judith Klavans and Smaranda Muresan (2001). Evaluation of the DEFINDER System for Fully Automatic Glossary Construction. *Proceedings of the American Medical Informatics Association Symposium (AMIA 2001).*

Smaranda Muresan, Evelyne Tzoukermann, and Judith Klavans (2001). Combining Linguistic and Machine Learning Techniques for Email Summarization. *Proceedings of the Conference on Computational Natural Language Learning (CoNLL-2001), held as a Workshop at the 39th Annual meeting of the Association for Computational Linguistics (ACL-2001)*.

Noemie Elhadad, Min-Yen Kan, Simon Lok, Smaranda Muresan (2001). PERSIVAL: personalized summarization over multimedia health-care information. *In Proceedings of The First ACM+IEEE Joint Conference on Digital Libraries, JCDL 2001*. (demo presentation)

Judith Klavans and Smaranda Muresan (2000). DEFINDER: Rule-based Methods for the Extraction of Medical Terminology and their Associated Definitions from On-line Text. *Proceedings of the American Medical Informatics Association Symposium (AMIA 2000)*.

BOOKS

Smaranda Muresan (under contract). Grammar Learning for Deep Language Understanding. **Research monograph** for the *Theory and Applications of Natural Language Processing Series*. Springer-Verlag

RECENT SERVICE TO UNIVERSITY AND OTHER PUBLIC BODIES

- 2014-2018 Columbia Senate: Research Officers Committee (ROC) (Acting Chair for January 1, 2017- June 30, 2017); member of the Alumni Relations Committee and External Relations Committee.
- 2010 - 2013 Co-Director of the Laboratory for the Study of Applied Language Technologies and Society (SALTS), together with Nina Wacholder (LIS) and Mark Aakhus (Communication Department)
- 2011 - 2013 Co-Founder for the New Jersey Natural Language Processing Group (with colleagues from AT&T Labs Research, ETS, Montclair State University, Siemens Research)
- Fall 2011 Organizer of the Yahoo! Data Sciences Seminar
- Fall 2011 Executive Steering Committee, Information Technology and Informatics (ITI) Program

CURRENT TEACHING

- Fall 2017 ENGI E4800.001 Data Science Capstone and Ethics; Data Science Institute, Columbia University
- Spring 2017 ENGI E4800.001 Data Science Capstone and Ethics; Data Science Institute, Columbia University
- Fall 2016 *COMS E6998: Data Science in Context: Computational Models of Social Meaning*. Data Science Institute and Department of Computer Science, Columbia University
- Spring 2015, 2014 *COMS E6998: Computational Models of Social Meaning*. Department of Computer Science, Columbia University
- Spring 2016 3 Projects in Computer Science (2 undergraduates)
- Fall 2015 6 Projects in Computer Science

CURRENT RESEARCH ADVISING

Debanjan Ghosh, PhD student (Rutgers) [PhD thesis advisor]

Olivia Winn, PhD student (Columbia) [PhD thesis advisor]

Elena Musi, Postdoctoral Researcher

Travid Riddle, Postdoctoral Researcher

Vikas Mohandos (supervisor MS GRA)

Sowmya Sree Bhagavatula (supervisor MS GRA)

Madhavan Kidambi (supervisor MS GRA)

Academic Committee

Daniel Bauer, Columbia University 2017 (PhD thesis committee)

Sakhar AlKhereyf, Columbia University, 2017 (Candidacy Committee)

Chris Hidey, Columbia University, 2017 (Candidacy Committee)

Morgan Ulinski, Columbia University, 2017 (Phd thesis proposal committee)

Sarah Ita Levitan, Columbia University, 2017 (PhD thesis committee)

Heba Elfardy Columbia University, 2017 (PhD thesis committee)

Noura Farra, Columbia University, 2015 (Candidacy committee)

Daniel Bauer, Columbia University 2016 (PhD thesis proposal committee)