Smaranda Muresan, *Curriculum Vitae*

Last Updated: May, 2025

Office Address	Barnard College, Columbia University 3009 Broadway, Milstein Center Room 901 New York, NY 10027
Email	smuresan@barnard.edu
Website	http://www.cs.columbia.edu/~smara/

RESEARCH INTERESTS

Human-centric Natural Language Processing for social good and responsible computing. I develop theory-guided and knowledge-aware computational models for understanding and generating language in context (e.g., visual, social, multilingual, multicultural) with applications to computational social science, education, public health. Research topics that I worked on over the years include: argument mining and generation, fact-checking and misinformation detection, figurative language understanding and generation (e.g., sarcasm, metaphor, idioms), language and vision, multilingual language processing for low-resource languages. I am also interested in explainable models and human-AI collaboration frameworks for high-quality datasets creation, helping humans solve tasks, and aligning AI systems with human values.

DEGREES IN HIGHER EDUCATION

2006	PhD Computer Science, Columbia University, New York, NY Title: "Learning Constraint-based Grammars from Representative Examples: Theory and Applications"
2002	M.Sc in Computer Science, Columbia University, New York, NY
1999	B.Sc in Computer Science, Technical University of Cluj Napoca, Romania

PROFESSIONAL EMPLOYMENT

July 2024 —	Associate Professor, Department of Computer Science, Barnard College
July 2024 —	Affiliated Faculty, Department of Computer Science, Columbia University and member of the Data Science Institute
01/2024 – 06/2024	Visiting Associate Professor in Computer Science, Barnard College
2017 – 06/2024	Research Scientist, Data Science Institute, Columbia University
2017 – 06/2024	Adjunct Associate Professor, Department of Computer Science, 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 Scientist, University of Maryland Institute for Advanced Computer Studies, College Park

HONORS and AWARDS

2019 - 2025	*Amazon Scholar*
2018	*Best short paper award* at ACL 2018
2017	*Best paper award* at SIGDIAL 2017
2013	*Distinguished Achievements in Research Award*, LIS Department, Rutgers University
2012	*Best Paper Runner-Up* at PACLIC 2012
2001	*Extraordinary Teaching Assistant Award*, School of Engineering and Applied Sciences, Columbia University, New York, NY

GRANTS

Abbreviations: DARPA = "Defense Advanced Research Projects Agency"; IARPA = "Intelligence Advanced Research Projects Activity"; NIH NIDA = "National Institute of Health, National Institute of Drug Abuse"; NSF = "National Science Foundation"; NSF SBIR= "NSF Small Business Innovation Research"; NSF EAGER = "NSF EArly-concept Grants for Exploratory Research"; NSF SGER = "NSF Small Grants for Exploratory Research"; DSI = "Data Science Institute, Columbia University"

Award size reflects both direct and indirect cost

Active Grants

2024-2025	Co-PI on a Magic Grant from the Brown Institute at Columbia University *"Improper Conduct"* (90K)
2024-2025	Co-PI on DSI at Columbia University Seed Award for " Detecting Emerging Opioid-Related Polysubstance use Patterns, Motivation, and HIV and Overdose Narratives through Machine Learning Based Natural Language"
2022 – 2026	Co-PI NIH NIDA award *"Identifying Suspected Drug Overdose Deaths in Near Real-Time Using Data Collected by Death Investigators"* PI Karli Hochstatter (Friends Research Institute); Co-PI Nabila El-Bassel (Columbia University) Award: $1,165,421 at Columbia; my share $732,605
2022 – 2026	Co-PI IARPA award *"PAUSIT: Privacy protection and Authorship attribution Using Stylebased Interpretable Transfer"* for the Human Interpretable Attribution of Text using Underlying Structure (HIATUS) program

PI Chris Callison-Burch (UPenn); Co-PIs: Marianna Apidianaki (UPenn), Kathleen Mckeown (Columbia University); Owen Rambow (Stony Brook), Niranjan Balasubramanian (Stony Brook)

Award: $2,793,111 at Columbia; my share $993,500

2022 – 2025 Co-PI DARPA award *"CHARM:Cross-Cultural Harmony though Affect and Response Mediation."* for the Computational Cultural Understanding (CCU) program

PI Kathleen McKeown (Columbia University); Co-PIs: Julia Hirschberg (Columbia University), Carl Vondrick (Columbia University), Colin Wayne Leach (Barnard College, Columbia University), Heng Ji (UIUC), Owen Rambow (Stony Brook), Bo Feng (UC Davis)

Award: $5,000,000 total; my share $459,000

2022 – 2025 Co-PI DSI Seed Fund award *"Psychology, Organizational Behavior and Neuroscience Literatures: Harnessing Data Science to Create a Unified Conceptualization of Diversity and Inclusion in Academic Research and the Popular Press"*

PI Valerie Purdie-Greenaway (Columbia University)

Award: $150,000 at Columbia; my share $75,000

Past Grants

2020 – 2023 PI NSF award *"Collaborative:Research:Computational Models for Studying Word Class Distinctions in Polysynthetic Languages"*

PI at UMD, Maria Polinsky

Award: $249,300 at Columbia; my share $249,300

2019 – 2023 Co-I at Columbia NIH/NIDA award *"HEALing: An Innovative County-Level Public Health Response to the Opioid Epidemic in New York State"*

PI Nabila El-Bassel (Columbia University); Co-PIs Louisa Gilbert (Columbia University), Edward Nunes (Columbia University); Co-Is too many to list

Award: $86,000,000 total; my share $536,000

2019 – 2021 PI Facebook Research award *"What it takes to fact-check a claim? Better Benchmarks for End-to-End Fact-Checking"*

Award: $34,000 at Columbia; my share $34,000

2018 – 2020 PI at Columbia NSF EAGER award *"EAGER: Collaborative Research:Automated Instruction Assistant for Argumentative Essays"*.

PI at Penn State, Rebecca Passonneau

Award: $143,000 at Columbia; my share $143,000

2018 – 2021 PI on a Columbia Collaboratory Fellows Fund grant to develop an interdisciplinary course *"Multilingual Technologies and Language Diversity"*

Co-PIs Lydia Liu (Columbia University), Isabelle Zaugg (Columbia University)

Award: $150,000 at Columbia; my share $74,000

2017 – 2021 Co-PI IARPA award *"SCRIPTS:System for CRoss-language Information Processing, Translation, and Summarization"* for the Machine Translation for English Retrieval of Information in Any Language (MATERIAL) program

PI Kathleen McKeown (Columbia University); Co-PIs Doug Oard (UMD), Marine Carpuat (UMD), Dragomir Radev (Yale University), Kenneth Heafield (University of Edinburgh)

Award: $14,000,000 total; my share $1,276,800

2014 – 2019 PI at Columbia NSF award *"RI: Medium: Collaborative Research: Write A Classifier: Learning Fine-Grained Visual Classifiers from Text and Images"*

PI at Rutgers, Ahmed Elgammal

Award: $463,208 at Columbia; my share $463,208

2014 – 2017 Co-PI NSF award *"Reducing Racial and Gender Achievement Gaps in STEM: Use of Natural Language Processing to Understand Why Affirmation Interventions Improve Performance"*

PI Valerie Purdie-Vaughns (Columbia University); Co-PIs Geoffrey Cohen (Stanford University), Jonathan Cook (Penn State University)

Award: $1,007,993 total; my share $72,778

2014 – 2016 Co-PI DSI Seed Fund award on *"Reducing Racial and Gender Achievement Gaps in STEM: Use of Natural Language Processing to Understand Why Affirmation Interventions Improve Performance"*

PI Valerie Purdie-Vaughns (Columbia University)

Award: $199,998 at Columbia; my share $99,000

2012 – 2017 Co-PI DARPA award *"DRATS: Detecting Relations and Anomalies in Text and Speech"* for the Deep Exploration and Filtering of Text (DEFT) program

PI Owen Rambow (Columbia University); Co-PIs: Kathleen McKeown (Columbia University), Julia Hirschberg (Columbia University), Mona Diab (GWU), Mari Ostendorf (University of Washington)

Award: $5,499,999 total; my share $981,000

2011 – 2014 PI at Columbia, NSF award *"Collaborative Research: Teaching Computers to Follow Verbal Instructions"* (2011-2014)

PI at Rutgers, Michael Littman; PI at UMBC, Marie desJardins

Award: $999,925 total; my share $303,000

2010 PI, Rutgers Office of the Vice President for Research to build the *"Laboratory for the Study of Applied Language Technologies and Society (SALTS)"*

Co-PIs Nina Wacholder, Mark Aakhus (Rutgers University)

Award: $112,000

2008 – 2010 PI at Rutgers on subaward from UMD for NSF SGER award, *"Exploiting Alternative Packagings of Source Meaning in Statistical Machine Translation"*

PI Philip Resnik (UMD)

Award: $144,763 total; my share $45,000

PUBLICATIONS (Google Scholar https://bit.do/smaranda-muresan)

h-index=38; h5-index=33

Conferences are the primary publication venue in natural language processing (acceptance rates mentioned when known). Senior authors are typically listed last, when equal supervision/contribution alphabetical order.

Students (PhD, Masters, Undergraduate, High School) and Postdocs directly advised or co-advised in underline (* denotes undergraduate students; ** denotes high school students at the time the work was performed). Smaranda Muresan is listed in **bold**.

Journal Articles

(J_1) Anubhav Jangra, Jamshid Mozafari, Adam Jatowt, **Smaranda Muresan** (2025). Navigating the Landscape of Hint Generation Research: From the Past to the Future. Accepted to Transactions of ACL (TACL).

(J_2) Amith Ananthram, Olivia Winn, **Smaranda Muresan** (2023). FeelingBlue: A Corpus for Understanding the Emotional Connotation of Color in Context. *Transactions of the Association for Computational Linguistics (TACL)*, 11: 176–190. (impact factor = 9.19)

(J_3) Nabila El-Bassel, Karli R Hochstatter, Melissa Slavin, Chenghao Yang Yudong Zhang, **Smaranda Muresan** (2022). Harnessing the Power of Social Media to Understand the Impact of COVID-19 on People Who Use Drugs During Lockdown and Social Distancing. *Journal of Addiction Medicine*. Volume 16, Issue 2, e123-e132. (impact factor = 5.5)

(J_4) Patricia Davies, Rebecca Passonneau, **Smaranda Muresan** and Yanjun Gao (2022). Analytical Techniques for Developing Argumentative Writing in STEM: A Pilot Study. *IEEE Transactions on Education*, vol. 65, no. 3, pp. 373-383 (impact factor = 2.6)

(J_5) Debanjan Ghosh, Alexander Fabbri*, **Smaranda Muresan** (2018). Sarcasm Analysis using Conversation Context. *Computational Linguistics*, 44(4):755–792.(impact factor = 9.3)

(J_6) Elena Musi, Debanjan Ghosh, **Smaranda Muresan** (2018). ChangeMyView Through Concessions: Do Concessions Increase Persuasion? *Discourse and Dialogue*, 9(1), 107-127 (h5-index = 24).

(J_7) **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)*, 67(11), 2725-2737, 2016. (impact factor = 3.5)

(J_8) **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 (impact factor = 3.5)

(J_9) **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 (impact factor = 1.1)

Refereed Conference Proceedings

(C_1) Arkadiy Saakyan, Shreyas Kulkarni, Tuhin Chakrabarty, and **Smaranda Muresan**(2025). Understanding Figurative Meaning through Explainable Visual Entailment. In Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL 2025).

(C_2) Milad Alshomary, Narutatsu Ri, Marianna Apidianaki, Ajay Patel, **Smaranda Muresan**, and Kathleen McKeown. 2025. Latent Space Interpretation for Stylistic Analysis and Explainable Authorship Attribution. In Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025).

(C_3) Prisha Samadarshi*, Mariam Mustafa*, Anushka Kulkarni*, Raven Rothkopf*, Tuhin Chakrabarty, **Smaranda Muresan** (2024). Connecting the Dots: Evaluating Abstract Reasoning Capabilities of LLMs Using the New York Times Connections Word Game. *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)*.

(C_4) Arkadiy Saakyan and **Smaranda Muresan** (2024). ICLEF: In-Context Learning with Expert Feedback for Explainable Style Transfer. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics, ACL 2024*.

(C_5) Tariq Alhindi, **Smaranda Muresan**, Preslav Nakov. Large Language Models are Few-Shot Training Example Generators: A Case Study in Fallacy Recognition. *Proceedings of Findings of the Association for Computational Linguistics ACL 2024*.

(C_6) Tuhin Chakrabarty, Vishakh Padmakumar, Faeze Brahman, **Smaranda Muresan** (2024). Creativity Support in the Age of Large Language Models: An Empirical Study Involving Emerging Writers. *Proceedings of the 16th ACM Conference on Creativity & Cognition* (June 16-21)

(C_7) Chenghao Yang, Tuhin Chakrabarty , Karli R Hochstatter, Melissa N Slavin, Nabila El-Bassel, **Smaranda Muresan** (2024). Identifying Self-Disclosures of Use, Misuse, and Addiction in Community-based Social Media Posts. *Findings of 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics*.

(C_8) Tuhin Chakrabarty, Philippe Laban, Divyansh Agarwal, **Smaranda Muresan**, Chien-Sheng Wu (2024). Art or Artifice? Large Language Models and the False Promise of Creativity. *Proceedings of the ACM (Association of Computing Machinery) CHI conference on Human Factors in Computing Systems (CHI 2024), May 13-16, 2024* (acceptance rate 26.3%).

(C_9) Robert Vacareanu, Siddharth Varia, Kishaloy Halder, Shuai Wang, Giovanni Paolini, Neha Anna John, Miguel Ballesteros, **Smaranda Muresan** (2024). A Weak Supervision Approach for Few-Shot Aspect Based Sentiment. *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024)* (acceptance rate 18%).

(C_{10}) Oliver Li*, Mallika Subramanian, Arkadiy Saakyan, Sky CH-Wang, **Smaranda Muresan** (2023). NormDial: A Comparable Bilingual Synthetic Dialogue Dataset for Modeling Social Norm Adherence and Violation. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, December 2023, (short paper) (acceptance rate 14%).

(C_{11}) Sky CH-Wang, Arkadiy Saakyan, Oliver Li*, Zhou Yu, **Smaranda Muresan** (2023). Sociocultural Norm Similarities and Differences via Situational Alignment and Explainable Textual Entailment. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, December 2023 (acceptance rate 23.3%).

(C_{12}) Yi R Fung, Tuhin Chakrabarty, Hao Guo, Owen Rambow, **Smaranda Muresan**, Heng Ji (2023). Normsage: Multi-lingual multi-cultural norm discovery from conversations on-the-fly. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*, December 2023 (acceptance rate 23.3%).

(C_{13}) Tuhin Chakrabarty, Kanishk Singh, Arkadiy Saakyan, **Smaranda Muresan** (2023). Learning to Follow Object-Centric Image Editing Instructions Faithfully. *Findings of the Association for Computational Linguistics: EMNLP 2023* (acceptance rate 45.4%).

(C_{14}) Tuhin Chakrabarty, Arkadiy Saakyan, Olivia Winn, Artemis Panagopoulou, Yue Yang, Marianna Apidianaki, **Smaranda Muresan** (2023). I Spy a Metaphor: Large Language Models and Diffusion Models Co-Create Visual Metaphors. *Findings of the Association for Computational Linguistics: ACL 2023*, pages 7370–7388 (acceptance rate 40.6%).

(C_{15}) Thomas Scialom, Tuhin Chakrabarty, **Smaranda Muresan** (2022). Fine-tuned Language Models are Continual Learners. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, pages 6107-6122 (acceptance rate 22.1%).

(C_{16}) Tuhin Chakrabarty, Arakadyi Saakyan, Debanjan Ghosh, **Smaranda Muresan** (2022). FLUTE: Figurative Language Understanding and Textual Explanations. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, pages 7139-7159 (acceptance rate 22.1%).

(C_{17}) Sky Wang, Evan Li, Oliver Li*, **Smaranda Muresan**, Zhou Yu (2022). Affective Idiosyncratic Responses to Music. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, pages 1220-1250 (acceptance rate 22.1%).

(C_{18}) Tariq Alhindi, Tuhin Chakrabarty, Elena Musi, **Smaranda Muresan** (2022). Multi-task Instruction-based Prompting for Fallacy Recognition. *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing (EMNLP 2022)*, pages 8172-8187 (acceptance rate 22.1%).

(C_{19}) Tuhin Chakrabartin, Justin Lewis, **Smaranda Muresan** (2022). CONSISTENT: Open-Ended Question Generation From News Articles. *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6954–6968 (acceptance rate 36.1%).

(C_{20}) Ramy Eskander, Cass Lowry, Sujay Khandagale, Judith Klavans, Maria Polinsky and **Smaranda Muresan** (2022). Unsupervised Stem-based Cross-lingual Part-of-Speech Tagging for Morphologically Rich Low-Resource Languages. *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL 2022)*, pages 4061–4072 (acceptance rate 21.96%).

(C_{21}) Sujay Khandagale, Yoann Léveillé, Samuel Miller, Derek Pham, Ramy Eskander, Cass Lowry, Richard Compton, Judith Klavans, Maria Polinsky, and **Smaranda Muresan** (2022). Towards Unsupervised Morphological Analysis of Polysynthetic Languages. *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing* (Volume 2: Short Papers), pages 334–340 (acceptance rate 26.5%).

(C_{22}) Tuhin Chakrabarty, Arkadiy Saakyan* and **Smaranda Muresan** (2021). *Don't Go Far Off*: An Empirical Study on Neural Poetry Translation. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)*, pages 7253–7265 (acceptance rate 25.6%).

(C_{23}) Tuhin Chakrabarty, Aadit Trivedi** and **Smaranda Muresan** (2021). Implicit Premise Generation with Discourse-aware Commonsense Knowledge Models. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP 2021)* (short paper), pages 6247–6252 (acceptance rate 17.9%).

(C_{24}) Arkadiy Saakyan*, Tuhin Chakrabarty and **Smaranda Muresan** (2021). COVID-Fact: Fact Extraction and Verification of Real-World Claims about COVID-19 Pandemic. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL 2021)*, pages 2116–2129 (acceptance rate 21.2%).

(C_{25}) Kevin Stowe, Tuhin Chakrabarty, Nanyun Peng, **Smaranda Muresan** and Iryna Gurevych (2021). Metaphor Generation with Conceptual Mappings. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL 2021)*, pages 6724–6736 (acceptance rate 21.2%).

(C_{26}) Chenghao Yang, Yudong Zhang and **Smaranda Muresan** (2021). Weakly-Supervised Methods for Suicide Risk Assessment: Role of Related Domains. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics (ACL 2021)* (short paper), pages 1049–1057 (acceptance rate 21.2%).

(C_{27}) Ramy Eskander, Cass Lowry, Sujay Khandagale, Francesca Callejas*, Judith Klavans, Maria Polinsky and **Smaranda Muresan** (2021). Minimally-Supervised Morphological Segmentation using Adaptor Grammars with Linguistic Priors. *Findings of the Association for Computational Linguistics: ACL 2021* (short paper), pages 3969–3974 (acceptance rate 37.5%).

(C_{28}) Elsbeth Turcan, Shuai Wang, Rishita Anubhai, Kasturi Bhattacharjee, Yaser Al-Onaizan and **Smaranda Muresan** (2021). Multi-Task Learning and Adapted Knowledge Models for Emotion-Cause Extraction. *Findings of the Association for Computational Linguistics: ACL 2021*, pages 3975–3989 (acceptance rate 37.5%).

(C_{29}) Tuhin Chakrabarty, Debanjan Ghosh, Adam Poliak and **Smaranda Muresan** (2021). Figurative Language in Recognizing Textual Entailment. *Findings of the Association for Computational Linguistics: ACL 2021* (short paper), pages 3354–3361 (acceptance rate 37.5%).

(C_{30}) Tuhin Chakrabarty, Christopher Hidey and **Smaranda Muresan** (2021). ENTRUST: Argument Reframing with Language Models and Entailment. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2021)*, pages 4958–4971 (acceptance rate 28%).

(C_{31}) Tuhin Chakrabarty, Xurui Zhang, **Smaranda Muresan** and Nanyun Peng (2021). MERMAID: Metaphor Generation with Symbolism and Discriminative Decoding. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2021)*, pages 4250–4261 (acceptance rate 28%).

(C_{32}) Elsbeth Turcan, **Smaranda Muresan** and Kathleen McKeown (2021). Emotion-Infused Models for Explainable Psychological Stress Detection. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2021)*, pages 2895–2909 (acceptance rate 28%).

(C_{33}) Debanjan Ghosh, Ritvik Shrivastavk, **Smaranda Muresan** (2021). "Laughing at you or with you": The Role of Sarcasm in Shaping the Disagreement Space. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: EACL 2021*, pages 1998–2010 (acceptance rate 24.7%).

(C_{34}) Tariq Alhindi, Brennan McManus*, and **Smaranda Muresan** (2021) What to Fact-Check: Guiding Check-Worthy Information Detection in News Articles through Argumentative Discourse Structure. *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2021)*, pages 380–391 (acceptance rate 41.5%).

(C_{35}) Tariq Alhindi, **Smaranda Muresan**, Daniel Preotiuc-Pietro (2020). Fact vs. Opinion: the Role of Argumentation Features in News Classification. *Proceedings of the 28th International Conference on Computational Linguistics (COLING 2020)*, pages 6139–6149 (acceptance rate 33.4%).

(C_{36}) Ramy Eskander, **Smaranda Muresan** and Michael Collins (2020). Unsupervised Cross-Lingual Part-of-Speech Tagging for Truly Low-Resource Scenarios. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, pages 4820–4831 (acceptance rate 24.6%).

(C_{37}) Tuhin Chakrabarty, **Smaranda Muresan** and Nanyun Peng (2020). Generating similes ~~effortesly~~ *like a pro*: A Style Transfer Approach for Simile Generation. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)*, pages 6455–6469 (acceptance rate 24.6%).

(C_{38}) Kasturi Bhattacharjee, Miguel Ballesteros, Rishita Anubhai, **Smaranda Muresan**, Jie Ma, Faisal Ladhak and Yaser Al-Onaizan (2020). To BERT or Not to BERT: Comparing Task-specific and Task-agnostic Semi-Supervised Approaches for Sequence Tagging. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP 2020)* (short paper), pages 7927–7934 (acceptance rate 16.6%).

(C_{39}) Tuhin Chakrabarty, Debanjan Ghosh, **Smaranda Muresan** and Nanyun Peng (2020). R^3: Reverse, Retrieve, and Rank for Sarcasm Generation with Commonsense Knowledge. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, pages 7976–7986 (acceptance rate 22.7%).

(C_{40}) Christopher Hidey, Tuhin Chakrabarty, Tariq Alhindi, Siddharth Varia, Kriste Krstovski, Mona Diab and **Smaranda Muresan** (2020). DeSePtion: Dual Sequence Prediction and Adversarial Examples for Improved Fact-Checking. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*, pages 8593–8606 (acceptance rate 22.7%).

(C_{41}) Ramy Eskander, Francesca Callejas*, Elizabeth Nichols*, Judith Klavans and **Smaranda Muresan** (2020). MorphAGram, Evaluation and Framework for Unsupervised Morphological Segmentation. *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, pages 7112–7122 (acceptance rate 60%).

(C_{42}) Debanjan Ghosh, Elena Musi, Kartikeya Upasani and **Smaranda Muresan** (2020). Interpreting Verbal Irony: Linguistic Strategies and the Connection to the Type of Semantic Incongruity. *Proceedings of the Society for Computation in Linguistics 2020*, pages 82–93 (acceptance rate 26%).

(C_{43}) Tuhin Chakrabarty, Christopher Hidey, **Smaranda Muresan**, Kathy McKeown and Alyssa Hwang* (2019). AMPERSAND: Argument Mining for PERSuAsive oNline Discussions. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing (EMNLP 2019)*, pages 2933–2943 (acceptance rate 23.4%).

(C_{44}) Olivia Winn, **Smaranda Muresan** (2018). 'Lighter' Can Still Be Dark: Modeling Comparative Color Descriptions. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018)*, pages 790–795 (acceptance rate 23.8%) (**best short paper award**).

(C_{45}) Elena Musi, Tariq Alhindi, Manfred Stede, Leonard Kriese, **Smaranda Muresan** and Andrea Rocci (2018). A Multi-layer Annotated Corpus of Argumentative Text: From Argument Schemes to Discourse Relations. *Proceedings of Language Resources and Evaluation Conference (LREC 2018)*, pages 1629-1636 (acceptance rate 65%).

(C_{46}) Debanjan Ghosh, **Smaranda Muresan** (2018) "With 1 follower I must be AWESOME :P." Exploring the Role of Irony Markers in Irony Recognition. *Proceedings of the International Conference on Web and Social Media (ICWSM) (short paper)*, pages 588-591 (acceptance rate 25.3%).

(C_{47}) Elena Musi, Mark Aakhus, **Smaranda Muresan**, Andrea Rocci, Manfred Stede (2017). From Theory to practice: the annotation of argument schemes. *Proceedings of the 2nd European Conference on Argumentation (ECA)* (acceptance rate unknown).

(C_{48}) Debanjan Ghosh, Alexander Fabbri*, **Smaranda Muresan** (2017). The Role of Conversation Context for Sarcasm Detection in Online Interactions. *Proceedings of SIGDIAL*, pages 186–196 (acceptance rate 41.6%) (**best paper award**)

(C_{49}) Debanjan Ghosh, Aquila Khanam, Yubo Han*, **Smaranda Muresan** (2016). Coarse-grained Argumentation Features for Scoring Persuasive Essays. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016)* (short paper), pages 549–554, (acceptance rate 22.0%)

(C_{50}) Debanjan Ghosh, Weiwei Guo, **Smaranda Muresan** (2015). Sarcastic or Not: Word-Embeddings to Predict the Literal or Sarcastic Meaning of Words. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pages 1003–1012 (acceptance rate 26.0%).

(C_{51}) 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.*Proceedings of Robotics: Science and Systems (RSS 2015)*, 8 pages (acceptance rate 26.1%).

(C_{52}) 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 (EDM 2015)*, 2074-2081 (acceptance rate 35.5%).

(C_{53}) 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)*, pages 1-12 (acceptance rate unknown).

(C_{54}) Debanjan Ghosh, **Smaranda Muresan** (2012). Relation Classification using Entity Sequence Kernels. *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012)*, pages 391–400 (acceptance rate 25%).

(C_{55}) 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)*, pages 127–136 (acceptance rate 33.0%) (**best paper runner-up**)

(C_{56}) **Smaranda Muresan** (2011). Learning for Deep Language Understanding. *Proceedings of 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 1858-1865 (acceptance rate 17.1%).

(C_{57}) 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 (ACL 2011)* (short paper), pages 581–586 (acceptance rate 25.0%).

(C_{58}) **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*, pages 137–149 (acceptance rate 22.5%).

(C_{59}) **Smaranda Muresan** (2010). A Learnable Constraint-based Grammar Formalism. *Proceedings of 23rd International Conference on Computation Linguistics (COLING) 2010: Posters* pages 885–893 (acceptance rate 41%).

(C_{60}) Christopher Dyer, **Smaranda Muresan** and Philip Resnik (2008). Generalizing Word Lattice Translation. *Proceedings of the Association for Computational Linguistics, ACL 2008*, pages 1012–1020, (acceptance rate 25%).

(C_{61}) **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*, pages 832–839, (acceptance rate 22.3%).

(C_{62}) **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*, pages 1-4 (acceptance rate unknown).

(C_{63}) **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)*, pages 231-234 (short paper) (acceptance rate unknown)

(C_{64}) **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)*, page 179-186 (acceptance rate unknown).

(C_{65}) Judith Klavans and **Smaranda Muresan** (2001). Evaluation of the DEFINDER System for Fully Automatic Glossary Construction. *Proceedings of the American Medical Informatics Association Annual Symposium (AMIA 2001)*, pages 324-328 (acceptance rate unknown)

(C_{66}) **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)* (acceptance rate 31.7%).

(C_{67}) 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)* (short paper) (acceptance rate 30%) (**best short paper runner-up**).

Refereed Workshop Publications

(W_1) Aquia Richburg, Ramy Eskander, **Smaranda Muresan**, Marine Carpuat (2020). An Evaluation of Subword Segmentation Strategies for Neural Machine Translation of Morphologically Rich Languages. *Proceedings of the The Fourth Widening Natural Language Processing Workshop at ACL 2020*.

(W_2) Efsun Sarioglu Kayi, Vishal Anand, **Smaranda Muresan** (2020). MultiSeg: Parallel Data and Subword Information for Learning Bilingual Embeddings in Low Resource Scenarios. *Proceedings of the 1st Joint Workshop of SLTU (Spoken Language Technologies for Underresourced languages) and CCURL (Collaboration and Computing for Under-Resourced Languages) (SLTU-CCURL 2020)*, pages 97–105.

(W_3) Tuhin Chakrabarty, Kilol Gupta, Smaranda Muresan (2019). Pay "Attention" to your Context when Classifying Abusive Language. *Proceedings of 3rd Abusive Language Workshop at ACL 2019*, pages 70–79.

(W_4) Ramy Eskander, Judith Klavans, and **Smaranda Muresan** (2019). Unsupervised Morphological Segmentation for Low-Resource Polysynthetic Languages. *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology at ACL 2019*, pages 189–195.

(W_5) Yanjun Gao, Alex Driban, Brennan Xavier McManus*, Elena Musi, Patricia Davies, **Smaranda Muresan** and Rebecca J. Passonneau (2019). Rubric Reliability and Annotation of Content and Argument in Source-Based Argument Essays. *Proceedings of the 14th Workshop on Innovative Use of NLP for Building Educational Applications at ACL 2019*, pages 507–518.

(W_6) Tariq Alhindi, Savvas Petridis, and **Smaranda Muresan** (2018). Where is your evidence: Improving fact-checking by justification modeling. *Proceedings of The First Workshop on Fact Extraction and Verification (FEVER) at EMNLP 2018*, pages 85–90.

(W_7) Ramy Eskander, Owen Rambow, and Smaranda Muresan (2018). Automatically Tailoring Unsupervised Morphological Segmentation to the Language. *Proceedings of the Fifteenth Workshop on Computational Research in Phonetics, Phonology, and Morphology (SIGMorphon) at EMNLP 2018*, pages 78–83.

(W_8) Christopher Hidey, Elena Musi, Alyssa Hwang*, **Smaranda Muresan**, and Kathy McKeown. 2017. Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum. *Proceedings of the 4th Workshop on Argument Mining at EMNLP 2017*, pages 11–21.

(W_9) 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*.

(W_{10}) 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*, pages 86–91.

(W_{11}) Debanjan Ghosh, **Smaranda Muresan**, Nina Wacholder, Mark Aakhus, Matthew Mitsui (2014). Analyzing Argumentative Discourse Units in Online Interactions. *Proceedings of the First Workshop on Argumentation Mining at ACL 2014*, pages 39–48.

(W_{12}) Gaurav Kharkwal, **Smaranda Muresan** (2014). Surprisal as a Predicator of Essay Quality. *Proceedings of The 9th Workshop on Innovative Use of NLP for Building Educational Applications at ACL 2014*, pages 54–60.

(W_{13}) Nina Wacholder, Debanjan Ghosh, Mark Aakhus, **Smaranda Muresan** (2014). Annotating Multiparty Discourse: Challenges for Agreement Metrics. *Proceedings of the 8th Linguistic Annotation Workshop (LAW VIII 2014) at COLING 2014*, pages 120–128.

(W_{14}) **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*, pages 171–179.

(W_{15}) Choonkyu Lee, **Smaranda Muresan**, Karin Stromswold (2012). Computational Analysis of Referring Expressions in Narrative Discourse. *Proceedings of the Workshop on Computational Linguistics for Literature at NAACL-HLT 2012*, pages 1–7.

(W_{16}) Monica Babeş-Vroman, James MacGlashan, Ruoyuan Gao, Kevin Winner, Marie desJardins, Michael Littman, **Smaranda Muresan** (2012). Learning to Interpret Natural Language Instructions. *Proceedings of the Workshop on Semantic Interpretation in an Actionable Context at NAACL-HTL 2012*, pages 1–6.

(W_{17}) **Smaranda Muresan** (2008). Learning to Map Text to Graph-based Meaning Representations via Grammar Induction. *Proceedings of TextGraphs-3: Graph-based Algorithms for Natural Language Processing Workshop at COLING 2008*, pages 9–16.

(W_{18}) **Smaranda Muresan** (2004). Inducing Constraint-based Grammars using a Domain Ontology. *Proceedings of the Ninth AAAI/SIGART Doctoral Consortium*, pages 995-996.

(W_{19}) **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, 8 pages.

(W_{20}) Evelyne Tzoukermann, **Smaranda Muresan**, and Judith L. Klavans (2001). GIST-IT: Summarizing Email using Linguistic Knowledge and Machine Learning. *Proceedings of the Workshop on Human Language Technology and Knowledge Management at ACL 2001*.

Book Chapters

(B_1) **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, pages 187–207.

(B_2) Philip Resnik, Chris Dyer, **Smaranda Muresan** (2010). Lattice Decoding for Hierarchical Models. In Joe Olive, ed., *Handbook of Natural Language Processing and Machine Translation*.

Edited Volumes (Proceedings and Journal Special Issues)

(E_1) **Smaranda Muresan**, Preslav Nakov, Aline Villavicencio (Eds) (2022). Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022).

(E_2) **Smaranda Muresan**, Preslav Nakov, Aline Villavicencio (Eds) (2022). Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022.

(E_3) Debanjan Ghosh, Beata Beigman Klebanov, **Smaranda Muresan**, Anna Feldman, Soujanya Poria, and Tuhin Chakrabarty (Eds) (2022). Proceedings of the 3rd Workshop on Figurative Language Processing (FLP) . Association for Computational Linguistics, Abu Dhabi, United Arab Emirates (Hybrid).

(E_4) Olivier Pietquin, **Smaranda Muresan**, Vivian Chen, Casey Kennington, David Vandyke, Nina Dethlefs, Koji Inoue, Erik Ekstedt, Stefan Ultes (Eds) (2020). Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2020).

(E_5) Beata Beigman Klebanov, Ekaterina Shutova, Patricia Lichtenstein, **Smaranda Muresan**, Chee Wee, Anna Feldman, and Debanjan Ghosh (Eds) (2020). Proceedings of the Second Workshop on Figurative Language Processing at ACL 2020.

(E_6) Beata Beigman Klebanov, Ekaterina Shutova, Patricia Lichtenstein, **Smaranda Muresan**, and Chee Wee (Eds) (2018). Proceedings of the First Workshop on Figurative Language Processing.

(E_7) Rebecca J Passonneau, Danielle McNamara, **Smaranda Muresan**, Dolores Perin (Eds) (2017). Special Issue on Multidisciplinary Approaches to AI and Education for Reading and Writing. *International Journal of Artificial Intelligence in Education*, 27.

(E_8) Diana Inkpen, **Smaranda Muresan**, Shibamouli Lahiri, Karen Mazidi, Alisa Zhila (Eds) (2015). *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*.

(E_9) Monojit Choudhury, Samer Hassan, Animesh Mukherjee, **Smaranda Muresan** (Eds) (2011). Special Issue on Network-based Models of Cognitive and Social Dynamics of Language. *Computer Speech and Language*, Volume 25 (3).

(E_{10}) Monojit Choudhury, Samer Hassan, Animesh Mukherjee, **Smaranda Muresan** (Eds) (2009). *Proceedings of TextGraphs-4: Graph-based Methods for Natural Language Processing Workshop*.

Other Publications

(O_1) Arkadiy Saakyan, Tuhin Chakrabarty, Debanjan Ghosh, **Smaranda Muresan** (2022). A Report on the FigLang 2022 Shared Task on Understanding Figurative Language. *Proceedings of the Third Workshop on Figurative Language Processing at EMNLP 2022*.

(O_2) Debanjan Ghosh, Avijit Vajpayee, and **Smaranda Muresan** (2020). A Report on the 2020 Sarcasm Detection Shared Task. *Proceedings of the Second Workshop on Figurative Language Processing*.

(O_3) Tariq Alhindi, Jonas Pfeiffer, and **Smaranda Muresan** (2019). Fine-Tuned Neural Models for Propaganda Detection at the Sentence and Fragment levels. *Proceedings of the 2nd workshop on NLP for Internet Freedom (NLP4IF): Censorship, Disinformation and Propaganda @ EMNLP 2019*, (shared task paper) pages 98-102.

(O_4) Tuhin Chakrabarty, **Smaranda Muresan** (2019). ColumbiaNLP at SemEval-2019 Task 8: The Answer is Language Model Fine-tuning. *Proceedings of the 13th International Workshop on Semantic Evaluation, @ NAACL 2019* (shared task paper) , pages 1144–1148.

(O_5) Zhuoran Liu, Shivali Goel, Mukund Yelahanka Raghuprasad, **Smaranda Muresan** (2019). Columbia at SemEval-2019 Task 7: Multi-task Learning for Stance Classification and Rumour Verification. *Proceedings of the 13th International Workshop on Semantic Evaluation @ NAACL 2019* (shared task paper), pages 1110–1114.

(O_6) Tuhin Chakrabarty, Tariq Alhindi and **Smaranda Muresan** (2018). Robust Document Retrieval and Individual Evidence Modeling for Fact Extraction and Verification. *Proceedings of The First Workshop on Fact Extraction and Verification (FEVER) at EMNLP 2018*, (shared task paper), pages 127–131.

(O_7) Elena Musi, Debanjan Ghosh, Mark Aakhus, **Smaranda Muresan**, Nina Wacholder (2017). Building an ontology of (dis)agreement space for argument mining. *Position paper for the SIGDIAL / SEMDIAL 2017 Joint Special Session on Negotiation Dialog.*

(O_8) 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)* (poster).

PATENTS

2025 *Global segmenting and sentiment analysis based on granular opinion detection Inventors*. Sunil Mallya Kasaragod, Abhinav Goyal, Yahor Pushkin, Srikanth Doss Kadarundalagi Raghura, Rishita Rajal Anubhai, Kasturi Bhattacharjee, Smaranda Muresan, Siddharth Chaitanyakumar Varia, Federico Torreti. Application number 18908673.

2003 *System and Method of Generating Dictionary Entries.* Judith Klavans and Smaranda Muresan, Columbia University, New York, NY. Serial No 398,535.

TEACHING

Fall 2024	Barnard College, *BC3997 "Natural Langauge Processing"* (undergraduate)
Spring 2024	Barnard College, *BC3997 "Large Language Models Foundations and Ethics"*(undergraduate)
Spring 2021	Columbia University, *COMS W4995 Multilingual Technologies and Language Diversity* (co-taught with Isabelle Zaugg, cross-listed between Computer Science and the Institute for Comparative Literature and Society) (open to undergraduates & graduates)
Spring 2020	Columbia University, *COMS W4995 Multilingual Technologies and Language Diversity* (co-taught with Isabelle Zaugg, cross-listed between Computer Science and the Institute for Comparative Literature and Society) (open to undergraduates & graduates)
Spring 2018	Columbia University, *COMS E6998: Data Science in Context: Computational Models of Social Meaning*. Department of Computer Science & Data Science Institute (graduate course)
Fall 2017	Columbia University, *ENGI E4800.001 Data Science Capstone and Ethics* Data Science Institute (graduate course)
Spring 2017	Columbia University, *ENGI E4800.001 Data Science Capstone and Ethics* Data Science Institute (graduate course)
Fall 2016	Columbia University, *COMS E6998: Data Science in Context: Computational Models of Social Meaning*, Department of Computer Science & Data Science Institute (graduate course)
Spring 2015	Columbia University, *COMS E6998: Data Science in Context: Computational Models of Social Meaning*, Department of Computer Science & Data Science Institute (graduate course)
Spring 2014	Columbia University, *COMS E6998: Data Science in Context: Computational Models of Social Meaning*, Department of Computer Science & Data Science Institute (graduate course)

Spring 2013	Rutgers University, *Language and Information* (mixed levels: PhD + MLIS)
Fall 2012	Rutgers University, *Retrieving and Evaluating Electronic Information*, ITI Program (undergraduate course)
Fall 2012	Rutgers University, *Database Design and Management*, MLIS program (graduate course)
Fall 2011	Rutgers University, *Language and Information* co-taught with Nina Wacholder (PhD-level course)
Fall 2011	Rutgers University, *Retrieving and Evaluating Electronic Information*, ITI Program (undergraduate course)
Spring 2011	Rutgers University, *Database Design and Management*, MLIS program (graduate course)
Spring 2011	Rutgers University, *Information Technologies for Libraries and Information Agencies*, MLIS program (graduate course)
Fall 2010	Rutgers University, *Information Technologies for Libraries and Information Agencies*, MLIS program (graduate course)
Spring 2010	Rutgers University, *Language and Information* co-taught with Nina Wacholder (PhD-level course)
Spring 2010	Rutgers University, *Retrieving and Evaluating Electronic Information*, ITI Program (undergraduate course).
Fall 2009	Rutgers University, *Information Technologies for Libraries and Information Agencies*, MLIS program (graduate course).
Spring 2009	Rutgers University, *Language and Information* co-taught with Nina Wacholder (PhD-level course)
Fall 2008	Rutgers University, *Information Technologies for Libraries and Information Agencies*, MLIS program (graduate course).

RESEARCH ADVISOR

Postdoctoral Researchers

2024–	**Milad Alshomary**
2018 – 2020	**Efsun Kayi** (Research Scientist at Johns Hopkins University Applied Physics Lab)
2016 – 2018	**Elena Musi** (Senior Lecturer (Associate Professor), University of Liverpool, UK)
2014 – 2016	**Travis Riddle** (Research Scientist at the National Institute of Health)

PhD Students (Chair or Co-Chair; Current and Graduated)

Fall 2024 –	**Amy Pu**, CS PhD student (Columbia University)
Fall 2023 –	**Anubhav Jangra**, CS PhD Student (Columbia University)
2022 –	**Arkadiy Saakyan**, CS PhD student (Columbia University): Explainable Text Generation

2021 – **Sky Wang**, CS PhD student (Columbia University) (co-advised with Zhou Yu): Computational Social Science

2020 – 2024 **Tuhin Chakrabarty**, CS PhD student (Columbia University): *Knowledge-aware Models and Human-AI Collaboration Frameworks for Creativity Support* (will join Stony Brook University as Assistant Professor in 2025, after a Research Scientist at Salesforce research)

2016 – 2023 **Olivia Winn**, CS PhD student (Columbia University): Thesis *"Seeing Red" or "Tickled Pink"?: Investigating the Power of Language and Vision Models through Color, Emotion, and Metaphor*

2017 – 2022 **Tariq Alhindi**, CS PhD student (Columbia University): Thesis *Computational Models of Argument Structure and Quality for Understanding Misinformation* (now Postdoc at MBZUAI)

2018 – 2021 **Ramy Eskander**, CS PhD student (Columbia University): Thesis *Unsupervised Morphological Segmentation and Part-of-Speech Tagging for Low-Resource Scenarios* (first position Senior Research Scientist at Twitter)

2012 – 2018 **Debanjan Ghosh**, PhD Student at SC&I (Rutgers University): Thesis *An Empirical Study of Verbal Irony: Identification, Interpretation, and its Role in Turn-Taking Discourse* (first position postdoc at MIT, then Research Scientist at ETS)

Masters Students

2024 **Shreyas Kulkarni** (MS GRA, CS Columbia University)

2023 **Mallika Subramanian** (MS GRA, CS Columbia University, 1 short paper at EMNLP 2023), **Kanishk Singh** (CS Columbia, 1 paper in Findings of EMNLP 2023)

2022 **Gunkirat Kaur** (MS GRA, CS Columbia University, 1 paper in preparation)

2021 **Derek Pham** (CS Columbia, 1 short paper: AACL-IJCNLP 2022)

2020 **Sujay Khandagale**, MS GRA (CS Columbia, 3 papers: ACL 2021, NAACL 2022, AACL-IJCNLP 2022), **Chenghao Yang** (Cs Columbia, 1 short paper ACL 2021, 1 journal paper, 1 paper under review at EMNLP 2023), **Yudong Zhang** (CS Columbia, 1 short paper ACL 2021, 1 journal paper)

2019 **Vishal Anand** (CS Columbia, 1 workshop paper), **Shivali Goel** (MS GRA, CS Columbia, 1 workshop paper), **Zhuoran Liu** (CS Columbia, 1 workshop paper), **Mukund Yelahanka Raghuprasad** (CS Columbia, 1 workshop paper), **Ritvik Shrivastava** (CS Columbia, 1 paper EACL)

2018 – 2019 **Kilol Gupta** (CS Columbia, 1 workshop paper), **Tuhin Chakrabarty** (CS Columbia, became my PhD Student)

2015 **Sowmya Bhagavatula** (MS GRA, CS Columbia, 1 conference paper EDM 2015), **Aquila Khanam** (CS Columbia, 1 short paper ACL 2016), **Madhavan Kidambi** (CS Columbia, 1 workshop paper)

2012 – 2013 **Gaurav Kharkaval** (CS Rutgers, Master Thesis, 1 workshop paper)

Undergraduate Students

2024 **Grace Li** (Barnard College; winner of the Columbia CS Department "Theodore R. Bashkow

award"), **Sunny Fang** (Barnard College)

2022 – 2023 **Oliver Li** (Columbia, 3 conference papers: 1 paper at EMNLP 2022, 2 at EMNLP 2023 (one first author)), **Avantika Garg** (Columbia), **Jacquie Zhang** (Barnard College)

2020 – 2021 **Arkadiy Saakyan** (Columbia, 2 conference papers as undergraduate (1 first author), now my PhD student).

2019 – 2020 **Francesca Callejas** (Columbia, 2 conference papers), **Brennan McManus** (Columbia, 1 workshop paper, 1 conference paper), **Elizabeth Nichols** (Barnard College, 1 conference paper)

2018 – 2019 **Serina Chang** (REU Columbia, went for PhD at Stanford)

2016 – 2017 **Alexander Fabbri** (Columbia, 1 best paper at SIGDIAL 2017, 1 journal paper, went for PhD at Yale)

2015 – 2016 **Yubo Han** (Columbia, 1 conference paper)

High school Students

2020 – 2021 **Aadit Trivedi** (1 short paper at EMNLP 2021)

PhD THESIS COMMITTEES

2023 **Dustin Wright** (CS, University of Copenhagen)

2022 **Giannis Karamanolakis** (CS, Columbia University), **Rose Sloan** (CS, Columbia University), **Shahab Raji** (CS, Rutgers University)

2021 **Zixiaofan Yang** (CS, Columbia University)

2020 **Chris Hidey** (CS, Columbia University), **Tobias Mayer** (CS, Universite Cote d'Azur)

2019 **Sarah Ita Levitan** (CS, Columbia University), **Noura Farra** (CS, Columbia University)

2018 **Wael Sallum** (CS, Columbia University), **Morgan Ulinski** (CS, Columbia University)

2017 **Daniel Bauer** (CS, Columbia University), **Jessica Ouyang** (CS, Columbia University), **Heba Elfardy** (CS, Columbia University)

2012 **Chang Liu** (School of Communication and Information, Rutgers University)

INVITED TALKS/SEMINARS

5/6-7/2024 *Large Language Models: The Curious Cases of Argumentation and Creativity*. Invited Seminar for the PhD Spring School of the PhD Program in Applied Linguistics: Managing Languages, Arguments and Narratives in the Datafied Society (LAND) in Switzerland.

4/20/2024 *Generative AI: The Curious Case of Commonsense and Creativity*. Invited Talk at Barnard Next.

03/07/2024 *Human-Centeric NLP: From Argumentation to Creativity*. Invited Talk at the Basque Center for Language Technology, Spain.

11/16/2023 *Large Language Models and Scientific Communication: Opportunities and Challenges* at the Visualizing Science Webinar on Responsible AI, Columbia University.

05/25/2023 *Human-centric Natural Language Processing for Social Good and Responsible Computing* at Forum Numerica Seminar of the Université Côte d'Azur's Academy of Excellence "Networks, Information and Digital society".

04/20/2023 *Human-centric Natural Language Processing for Social Good and Responsible Computing*. Invited Talk, Computer Science Seminar, Barnard College.

02/03/2023 Invited talk at the NIST TAC Workshop Panel on "Understanding the Nature of Accurate vs. Inaccurate and Authentic vs. Potentially Inauthentic News/Discourse".

10/24/2022 *The Role of Text Generation in Argumentation.* Invited Plenary Talk at the Dagstuhl Seminar "Towards a Unified Model of Scholarly Argumentation".

10/13/2022 *Role of Text Generation for Responsible Computing* Invited Talk at Rice University's CS Colloquim.

5/27/2022 *The Role of Text Generation in Fighting Hostile Posts.* Keynote Talk at the "Second Workshop on Combating Online Hostile Posts in Regional Languages during Emergency Situation" collocated with ACL 2022.

4/29/2022 *Knowledge-enhanced Text Generation: The Curious Case of Figurative Language and Argumentation.* Invited Talk, Women+@DCS Seminar, University of Sheffield, UK.

10/19/2021 *Knowledge-enhanced Text Generation: The Curious Case of Figurative Language and Argumentation*. Invited talk at the University of Michigan's AI Seminar.

10/07/2021 *Knowledge-enhanced Text Generation: The Curious Case of Figurative Language and Argumentation*. Invited talk, Computer Science Seminar, Barnard College.

8/21/2021 *Knowledge-enhanced Text Generation: The Curious Case of Figurative Language and Argumentation*. Keynote talk at IJCAI 2021 Workshop "Is Neuro-Symbolic SOTA still a myth for Natural Language Inference?"

4/16/2021 *Knowledge-enhanced Text Generation: The Curious Case of Figurative Language and Argumentation*. Invited talk at UIUC NLP seminar.

05/13/2019 *Computational Models for Understanding Language in Social Context*. Invited Talk at Amazon.

10/24/2017 *Identification of Non-literal Language in Social Media: A Case Study on Sarcasm*. Invited Talk at Bloomberg.

9/27/2017 *Argument Mining for Assessing Students' Persuasive Essays.* Invited Talk at the NSF Workshop on "Connecting Language, Interaction and Education in Digital Environments", Penn State University.

9/22/2016 *Computational Models of Understanding Language in Social Context*. Invited Talk at the Social Computing Workshop organized by the Army Research Lab.

4/19/2016 *Argumentation Mining In Online Interactions: Opportunities and Challenges*. Invited Talk at Dagstuhl Seminar on "Natural Language Argumentation: Mining, Processing, and Reasoning over Textual Arguments".

3/16/2016	*Computational Models of Understanding Language in Context*. Invited Talk at the Hearst Corporation Meeting with Hearst's Executive Director of R&D.
2/24/2014	*Identification of Non-literal Language in Social Media: A Case Study on Sarcasm*. Invited talk at Yahoo! Research Lab, NYC
5/22/2014	*Investigating the Content of Values-Affirmation Interventions: A Data-Driven Approach to Discovering Topics in Affirmation Essays*. Invited talk at the APS Symposium *Computational Methods for Linguistic Insights into Personality and Behavior*.
7/30/2012	*Context-dependent Language Understanding*. Invited Talk at Rutgers's Big Data Meeting.
2/9/2012	*Identifying Sarcasm in Twitter*. Columbia University Speech Group Talks.
11/11/2011	*Learning for Deep Language Understanding*. Invited Talk at the CUNY Graduate Center.
07/01/2010	*Learnable Constraint-based Grammars for Deep Language Understanding*. Invited talk at the Educational Testing Services.
05/27/2010	*Learnable Constraint-based Grammars for Deep Language Understanding*. Invited Talk at the IBM T.J. Watson Research Center.
04/13/2010	*Learnable Constraint-based Grammars for Deep Language Understanding*. Invited Talk at the Yahoo Machine Learning Seminar, Rutgers University.
10/29/2009	*Are those fighting words? The expression of opposition in online text* Invited Talk at ISchool, Syracuse University.
08/05/2009	*National Language Processing at Rutgers* Invited Talk at National University of Singapore.
10/02/2008	*A Computational Model for Language Learning and Understanding*. Invited Talk at Rutgers University's Center for Cognitive Science.
02/26/2008	*Learning Constraint-Based Grammars from Representative Data*. Invited Talk at 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.

SERVICE AND LEADERSHIP TO THE DEPARTMENT AND UNIVERSITY

Department/School Service and Leadership

2022	Reviewer for the Inaugural Avanessians Doctoral Fellowship Program for Engineering Thought Leaders and Innovators in Data Science (Columbia University)
2020 –	Reviewer for Data Science Institute's Seed Grant Proposals (Columbia University)
2017 – 2020	Data Science Institute Education Committee (Columbia University)

2017 – 2020	Data Science Institute MS Admission Committee (Columbia University)
2017 – 2020	Faculty Advisor for Data Science Institute's Masters Students (Columbia University)
2008 – 2013	Library and Information Science Department's Research & Speakers Committee (Rutgers University)
Fall 2011	Executive Steering Committee, Information Technology and Informatics (ITI) Program (undergraduate program) (Rutgers University)
Fall 2011	Library and Information Science PhD Area Curriculum Committee (Rutgers University)
Spring 2011	PhD Awards Committee (School of Communication and Information, Rutgers University)
2010 – 2011	Planning Committee for a new MLIS Specialization "Informatics and Design" (Rutgers University)
2009 – 2010	Organizer of Library and Information Science Department's Research Talks (Rutgers University).
2008 – 2010	PhD Admission Committee (School of Communication and Information, Rutgers University)

University Service and Leadership

2014 – 2018	Columbia University Senate: member of the Research Officers Committee (ROC) (Acting Chair for January 1, 2017 – June 30, 2017); member of the Alumni Relations Committee and External Relations Committee (Columbia University).
2010 – 2013	Co-founder and co-director of the university-wide Laboratory for the Study of Applied Language Technologies and Society (SALTS), together with Nina Wacholder (LIS) and Mark Aakhus (Communication Department) (Rutgers University)
2011 – 2013	Co-founder of 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 at Rutgers University (university-wide)

SERVICE AND LEADERSHIP TO THE PROFESSION

Co-Chair	ACL 2025 Demonstration Track
Co-Organizer	The 4th Workshop on Figurative Language Processing @ NAACL 2024.
Co-Organizer	Shared Task on Visual Language Understanding with Textual Explanations @ FigLang Processing at NAACL 2024.
Action Editor	ACL Rolling Review (2023 –)
Action Editor	Transactions of the Association for Computational Linguistics (TACL) (2023 – 2025)
Standing Reviewer	Transactions of the Association for Computational Linguistics (TACL) (2020 – 2023)
Co-Organizer	10th Workshop on Argument Mining (ArgMining 2023) at EMNLP 2023

Program Co-Chair	The 60th Annual Meeting of the Association for Computational Linguistics (ACL 2022).
Co-Organizer	The 3rd Figurative Language Processing Workshop at EMNP 2022
Co-Organizer	Shared Task on "Figurative Language Understanding Through Textual Entailment and Explanations" at FigLang Workshop@EMNLP 2022
NSF Review Panel	2022, 2021, 2019, 2018, 2016, 2014, 2011, 2009
Advisory Board Member	ACL Rolling Review Advisory Board Member (2022 – 2024)
Nominating Committee	Member of the Nominating Committee for the North American Association for Computational Linguistics (NAACL) board (2022 – 2023)
Board Member	North American Association for Computational Linguistics (NAACL) 2020 – 2021.
Area Chair	Sentiment Analysis, Stylistic Analysis, and Argument Mining Track at ACL 2021
Program Co-Chair	The 21st Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2020)
Senior Area Chair	Sentiment Analysis, Stylistic Analysis and Argument Mining Tack at ACL 2020
Co-Organizer	The 2nd Figurative Language Processing Workshop Series at ACL 2020
Co-Organizer	Shared Task on "The Role of Conversation Context for Sarcasm Detection" at FigLang Workshop@ACL 2020
Inaugural Co-Chair	The New York Academy of Sciences' *Annual Symposium on NLP/Dialog/Speech* November 2019 and 2020.
Area Chair	Discourse and Pragmatics Track at ACL 2019
Co-Editor	Special Issue of IJAIED: MARWIDE: Multidisciplinary Approaches to Reading and Writing Integrated with Disciplinary Education (2018)
Co-Organizer	The 1st Figurative Language Processing Workshop Series at NAACL 2018
Best Paper Comm.	Member of the Best Paper Selection Committee at NAACL 2018
Area Chair	Sentiment Analysis and Argument Mining Track at ACL 2018
Area Chair	Sentiment Analysis Track at NAACL 2018
Area Chair	Information Extraction and NLP Applications Track at ACL 2017
Area 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 at NAACL-HLT 2015
Co-Organizer	Multidisciplinary Advances in Reading and Writing for Science Education Workshop (NSF funded workshop), 2015
Faculty Advisor	Student Research Workshop at NAACL-HLT 2015

Area Chair	Second Joint Conference on Lexical and Computational Semantics (*SEM 2013)
Publicity Chair	The 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2012)
Co-Editor	Special Issue on “Network-based Models of Cognitive and Social Dynamics of Language”. Computer Speech and Language. Volume 25 (3), 2011.
Panelist	”What It’s Like to Work at My Institution” Faculty Panel at ”Preparing Future Faculty Workshop” at Columbia University on April 16, 2010.
Expert Evaluator	Killam Research Fellowships, Canada Council (http://canadacouncil.ca/prizes/killam/) (by invitation, 06/2010 – 09/2010).
Co-Organizer	TextGraphs-4: Graph-based Methods for Natural Language Processing. Workshop at ACL-IJCNLP 2009.