

Aalok Sathe

present address: Cambridge, MA 02139, USA
 email: aalok.sathe@richmond.edu web: aalok-sathe.gitlab.io

Education

May 2021 (Aug '17–)	B.S.	University of Richmond, VA, USA <i>Majors:</i> Computer Science (Honors), Cognitive Science <i>Minors:</i> Linguistics, Mathematics Church-Kent prize: outstanding graduate in CS	GPA: 4/4 class rank 1/775
May 2020 (Jan '20–)		University of Edinburgh, Scotland, UK Visiting student in the School of Informatics Coursework: NLP (NLU+), Linguistics, Theoretical CS	GPA: 3.9/4 (converted) UK scale: 78.33/100

Research Experience

- Research Associate, *Brain & Cognitive Sci.*, **Massachusetts Institute of Technology** Cambridge, MA
 Mentors: Dr Evelina (Ev) Fedorenko Jul 2021 – present
- Research Intern, *NLP Group*, **Microsoft Research** Bengaluru, India
 Mentors: Dr Monojit Choudhury, Dr Somak Aditya May 2020 – Aug 2020
 - ▷ Developed ‘TaxiNLI’, a taxonomic fragmentation of Natural Language Inference (NLI) guided by human reasoning capabilities. Analyzed large pretrained neural models’ capabilities and limitations in NLI through error analysis, factor analysis, and visualization. [\[paper\]](#) [\[data\]](#)
 - ▷ Used TaxiNLI to study NLI in a low-resource cross-lingual setting through few-shot transfer. Investigated which categories are hard to generalize and acquire in English as well as cross-lingually using XNLI.
- Research Assistant, *Dept. of Math & CS*, **University of Richmond** Richmond, VA
 Mentors: Dr Joonsuk (Jon) Park Dec 2018 – May 2021
 - ▷ Formulated a new task and a first-of-its-kind large-scale annotated dataset ‘WikiFactCheck-English’ with 120k+ entries using claims, context, and evidence extracted from Wikipedia citations. Developed a novel feature-based logistic regression classifier for the NLI subtask. [\[paper\]](#) [\[data\]](#)
- Research Assistant, *Dept. of Math & CS*, **University of Richmond** Richmond, VA, USA
 Mentors: Dr Prateek Bhakta, Dr Heather Russell, Dr Sara Krehbiel Apr 2019 – May 2020
 - ▷ Developed an efficient C++ based graphs library wrapped to provide a Python interface and a web frontend to construct, analyze, and visualize graphs of vertex colorings or ‘coloring graphs’ for graph theory research. Contributed new sequences ([A307334](#), [A309315](#), [A309379](#), [A309380](#)) to OEIS. [\[talk\]](#) [\[code\]](#)
- Research Assistant, *Dept. of Math & CS*, **University of Richmond** Richmond, VA
 Mentors: Dr Taylor Arnold Aug 2018 – Jul 2019
 - ▷ Developed a Python-based toolkit for audio analysis in digital media, useful for Digital Humanities and cultural analytics research. Experimented with ways to build models for audio using VGGish and regression on spectrographs. Designed a transfer learning pipeline for laughter detection. [\[talk\]](#) [\[poster\]](#) [\[code\]](#)
- Research Assistant, *Dept. of Psychology*, **University of Richmond** Richmond, VA
 Mentors: Dr Cindy Bukach, Dr Matthew Lowder Sep 2017 – Dec 2020
 - ▷ Designed and programmed an ERP-based psycholinguistics study about visual word perception. Developed a web-based version for crowd-sourcing behavioral data from Amazon Mechanical Turk.
 - ▷ Ran EEG capping experiments on human participants, performed data analysis, and maintained code for multiple projects across the lab. [\[ERP poster\]](#) [\[MTurk poster\]](#)

Peer-Reviewed Publications

#co-first author

4. Vaduguru, S., **Sathe, A.**, Choudhury, M., Sharma, D. M., *Sample-efficient Linguistic Generalizations through Program Synthesis: Experiments with Phonology Problems*, in Proceedings of the 18th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology (SIGMORPHON 2021). [\[paper\]](#) [\[data\]](#)
3. Joshi, P.#, Aditya, S.#, **Sathe, A.#**, Choudhury, M., *TaxiNLI: Taking a Ride up the NLU Hill* in Proceedings of the 24th Conference on Computational Natural Language Learning (CoNLL 2020). [\[paper\]](#) [\[video\]](#) [\[data\]](#)
2. **Sathe, A.**, Ather, S., Le, T. M., Perry, N. M., Park, J., *Automated Fact-Checking of Claims from Wikipedia (WikiFactCheck-English)*, in Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020). [\[paper\]](#) [\[data\]](#)
1. **Sathe, A.**, *A rule-based system for the transcription of Sanskrit from the Devanagari orthography to the International Phonetic Alphabet*, in Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018) Workshop 26: Collaboration and Computing for Under-Resourced Languages (CCURL) Miyazaki, Japan. [\[paper\]](#) [\[poster\]](#) [\[code\]](#)

Talks and Presentations

*abstract reviewed

7. ***Sathe, A.**, *Learning to Fact-check with partial supervision using BERT+attention* in Honors Theses 2021, University of Richmond Scholarship Repository.
6. ***Sathe, A.**, Arnold, T., *Automatic methods of detecting audio events in digital media for cultural analytics* Poster at the Natural Language, Dialogue Systems, and Speech symposium (NDS 2019), New York Academy of Sciences (NYAS), New York City. [\[poster\]](#)
5. ***Sathe, A.**, Arnold, T., *Cultural analytics in moving-image media using automatic audio-event detection*. Talk at the Chicago Colloquium for Digital Humanities and Computer Science (DHCS 2019), Chicago IL. [\[talk\]](#)
4. **Sathe, A.**, Su, W., Bhakta, P., *Constructing and visualizing graphs of vertex colorings ('Coloring Graphs')*. Talk at the annual Shenandoah Undergraduate Mathematics and Statistics conference (SUMS 2019), James Madison University, Harrisonburg VA. [\[talk\]](#) [\[code\]](#)
3. *Laskowitz, S., Deutsch, A., Cole, Z. J., Jackson, W., Cinque, C., **Sathe, A.**, Munezero, P., Kade, P., Couperus, J. W., Reed, C. L., & Bukach, C. M., *Angry Bias for Black Male Faces Associated with Increased Other-Race Effect for Identity during Visual Discrimination*. Poster at the 30th Annual Convention of the Association for Psychological Science (APS 2018), San Francisco, CA
2. **Sathe, A.**, Bukach, C. M., Lowder, M. W., Coperus, J. W., & Reed, C. L. *Understanding visual perception of words using event-related potentials (ERP)*. Poster presentation at the Annual Student Symposium of the School of Arts & Sciences 2018, University of Richmond. [\[ERP poster\]](#) [\[MTurk poster\]](#)
1. **Sathe, A.**, *Determining the Optimal Context in a Markov Language Model for Authorship Prediction*, at the Conference of Association for International Mathematics Education and Research 2017. Bengaluru, India. [\[paper\]](#)

Preprints and Works in Submission

2. [under review] *Symmetries in Graph Coloring Reconfiguration* submitted to Journal of Graph Theory. [\[paper\]](#)
1. [in prep./preprint] *Analyzing the Effects of Reasoning Types on Cross-Lingual Transfer Efficiency in NLI*

Honors (select)

- 2021 **Joseph F. and Mary Church Kent prize in Computer Science** *[awarded to 1 graduate]*
Dept. of Math and Computer Science, University of Richmond.
- 2020 **Outstanding Research in Computer Science** *[awarded to 1 during academic year]*
Dept. of Math and Computer Science, University of Richmond.
- 2020 **Phi Beta Kappa** *[top 1% of junior class]* initiated during junior year
Phi Beta Kappa, Epsilon chapter of Virginia, University of Richmond
- 2018 **First place at testing site, ACM-ICPC** *[#20/200 in region]*
ACM-ICPC 2018 Mid-Atlantic USA Regionals, at the Christopher Newport University testing site.
- 2017-21 **Robins Science Scholar** *[awarded to 5 per class year]* full tuition plus room and board scholarship
Richmond Scholars Program, University of Richmond
- 2015-16 **Bronze Medal; Induction into IOL Hall of Fame** for being a repeat medalist
International Linguistics Olympiad (IOL), Blagoevgrad, Bulgaria (2015), and Mysore, India (2016)

Work Experience

- **Blended instruction assistant** *UR Dept. of Math and Computer Science* *Aug – Dec 2020*
Managed technological setup in a blended classroom format and led virtual discussions on classwork problems for remote students in Math 306: Abstract Algebra I (Dr Della Dumbaugh, Fall 2020).
- **Grading assistant** *UR Dept. of Math and Computer Science* *Aug 2018 – May 2019; Jan – May 2021*
Graded weekly problem sets. CS 221: Data Structures (Dr Jory Denny, Fall 2018); CS 315: Algorithms (Dr Prateek Bhakta, Spring 2019); CS 301: Computer Organization (Dr Prateek Bhakta, Spring 2021).
- **Resident assistant** *UR Residence Life and Undergraduate Student Housing* *Aug 2018 – Dec 2019*
Promoted community development and safety in two first-year dorms with 30+ residents in two academic years.
- **Language tutor: Sanskrit** *UR Global Studio* *Jan – May 2018*
Tutored Sanskrit one-on-one: researched helpful resources, corrected exercises, and acted as a sounding board.

Societies, Activities, and Service

- Sep 2018 – *present* Member of the Jury and Problem Design Committee, Asia-Pacific Linguistics Olympiad
- Jul 2016 – *present* Member of the Jury and Problem Design Committee, Panini Linguistics Olympiad (PLO)
- Aug 2018 – Dec 2020 Educational Chair, ACM Chapter, University of Richmond
- Jan – Dec 2019 Bass Vocalist, UR Schola Cantorum mixed choir, University of Richmond
- Aug 2018 – Dec 2019 President, UR Pick-up Soccer Club, University of Richmond
- Aug 2018 – May 2019 Executive Member, South Asian Student Alliance, University of Richmond

Projects (select) *(year indicates when the project began; projects may be ongoing)*

- 2020 **Perceptions of the Pandemic** — A large-scale survey related to perceptions, attitudes, and wisdom-of-crowds analysis towards COVID-19 in Pune city measured using non-typical metrics relevant to the demographics, such as, availability of water and sanitation, living density at home, etc. [\[report\]](#) [\[webinar recording\]](#)
- 2020 **ipatope** — A web-based interactive and educational IPA chart that allows the user to combine multiple phonological features to sort, filter, and hear sounds. Has modes to allow flashcard-like use. (MIT License). [\[webpage\]](#)
- 2018 **pyMediaAnnotator** — A pyGTK- and vlc-based application to hand-annotate audio and video files for cultural analytics and machine learning research. (GPLv3). [\[package\]](#)
- 2017 **SanskritIPA** — A lossless transcription program written in Python to transcribe Sanskrit text written in Devanagari to IPA preserving phonological and prosodic features. (GPLv3). [\[code\]](#)

Course Projects (select)

- 2020 **Machine translation** Implemented RNN+attention and transformer-based encoder-decoder machine translation models using PyTorch on the EUROPARL data. Analyzed attention patterns across layers, comparing RNNs to Transformer, and tested hypotheses for **en-de** translation. (INFR 11157, NLU+, University of Edinburgh)
- 2020 **Distributed leader election and ring creation** Implemented a distributed system using multithreading in Java. Implemented leader election using Chang-Roberts algorithm and ring creation using distributed Hamiltonian cycle search. (INFR 11022, Distributed Systems, University of Edinburgh)
- 2020 **Register machine simulator** Simulated a register machine (analogous to Turing machine) in Python. Implemented a parser for RM grammar, and macros for high-level functionality. (INFR 10057, Theoretical Computer Science, University of Edinburgh)
- 2019 **Agent-based fire evacuation simulator** Agent-based simulation of a moving danger (fire) and a constrained space (building floor) to evaluate various building layouts and per-agent escape strategies to minimize damage and death. (CS 326, Simulation, University of Richmond) [[report](#)]
- 2019 **CPU pipelining** Simulated pipelining of MIPS assembly instructions in the CPU. (CS 301, Computer Organization, University of Richmond)
- 2018 **Complex Systems: ant colonies** What is a complex system? How do ant colonies function using relatively simple components but nevertheless exhibit complex behavior? How are they related to other complex systems, such as, human society, the brain, and language? [[paper](#)] [[group dynamics talk](#)] [[algorithms talk](#)]