Sebastin Santy

PhD student, UNIVERSITY *of* WASHINGTON

⊙ sebast.in   @ ssanty@uw.edu   ⚲ github.com/sebastinsanty   🎓 Google Scholar

📍 CSE 614, Paul G. Allen Center, 185 E Stevens Way NE, Seattle, WA 98195

Education

University of Washington, *Seattle, USA*   Sep 2021 - Present
Ph.D. in Computer Science (ongoing)
Advisor. Katharina Reinecke

Birla Insitute of Technology and Science, *Goa, India*   Aug 2015 - Dec 2018
B.E. in Electronics & Instrumentation Engineering

Experience

Allen Institute for Artificial Intelligence, *Seattle, USA*   Jun 2022 - Sep 2022
Research Intern with Maarten Sap, Ronan Le Bras

Microsoft Research, *Bangalore, India*   Jul 2019 - Jul 2021
AI Center Fellow with Kalika Bali, Monojit Choudhury, Tanuja Ganu

Microsoft Research, *Bangalore, India*   Jan 2019 - Jul 2019
Research Intern with Kalika Bali, Monojit Choudhury

Carnegie Mellon University, *Pittsburgh, USA*   Jul 2018 - Dec 2018
Research Intern with David S. Touretzky

University College London, *London, UK*   May 2018 - Jul 2018
Research Intern with Emine Yilmaz, Rishabh Mehrotra

Julia Language, *Remote*   May 2018 - Aug 2018
Google Summer of Code with Lyndon White

Mozilla, *Remote / Austin, USA*   May 2017 - Aug 2017
Google Summer of Code with Dylan Hardison

Publications

S=In Submission, C=Conference, W=Workshop, P=Poster/Demo, J=Journal

C.6   **Language Translation as a Socio-Technical System: Case-Studies of Mixed-Initiative Interactions.** 📄
Sebastin Santy, Kalika Bali, Monojit Choudhury, Sandipan Dandapat, Tanuja Ganu, Anurag Shukla,
Jahanvi Shah, Vivek Seshadri
ACM SIGCAS Conference on Computing and Sustainable Societies (Virtual)   `COMPASS 2021`

C.5   **Use of Formal Ethical Reviews in NLP Literature: Historical Trends and Current Practices.** 📄
Sebastin Santy, Anku Rani, Monojit Choudhury
Annual Conference of the Association for Computational Linguistics (Virtual)   `ACL Findings 2021`

C.4   **The State and Fate of Linguistic Diversity and Inclusion in the NLP World.** 📄
Sebastin Santy*, Pratik Joshi*, Amar Budhiraja*, Kalika Bali, Monojit Choudhury (* = Equal Contribution)
Annual Conference of the Association for Computational Linguistics (Virtual)   `ACL 2020`

C.3   **Learnings from Technological Interventions in a Low Resource Language: A Case-Study on Gondi.** 📄
Sebastin Santy*, Devansh Mehta*, Ramaravind Mothilal, Brij Mohan Lal Srivastava, Alok Sharma, Anurag Shukla,
Vishnu Prasad, Venkanna U, Amit Sharma and Kalika Bali (* = Equal Contribution)
International Conference on Language Resources and Evaluation (Virtual)   `LREC 2020`

C.2   **Unsung Challenges of Building and Deploying Language Technologies for LRL Communities.** 📄
Pratik Joshi, Christain Barnes, Sebastin Santy, Simran Khanuja, Sanket Shah, Anirudh Srinivasan, Satwik Bhat-
tamishra, Sunayana Sitaram, Monojit Choudhury, Kalika Bali
16th International Conference on Natural Language Processing, Hyderabad, India   `ICON 2019`

C.1   **BITS Darshini: A Modular, Concurrent Protocol Analyzer Workbench.** 📄
Prasad Talasila, Mihir Kakrambe, Anurag Rai, <u>Sebastin Santy</u>, Neena Goveas, Bharat Deshpande
19th ACM International Conference on Distributed Computing and Networking, Varanasi, India   `ICDCN 2018`

P.3   **Deploying Language Technologies for Underserved Communities.** [Invited Poster] 📄
Kalika Bali, Monojit Choudhury, Sunayana Sitaram, <u>Sebastin Santy</u>
UNESCO International Conference on Language Technologies for All, Paris, France   `LT4All 2020`

P.2   **INMT: Interactive Neural Machine Translation Prediction.** 📄
<u>Sebastin Santy</u>, Sandipan Dandapat, Monojit Choudhury, Kalika Bali
Conference on Empirical Methods in Natural Language Processing, Hong Kong [Systems Demo]   `Demo, EMNLP 2019`

P.1   **Towards Task Understanding in Visual Settings.** 📄
<u>Sebastin Santy</u>, Wazeer Zulfikar, Rishabh Mehrotra, Emine Yilmaz.
33rd AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, USA [Student Abstract]   `Abstract, AAAI 2019`

W.2   **BERTologiCoMix: How does Code-Mixing interact with Multilingual BERT?** 📄
<u>Sebastin Santy</u>, Anirudh Srinivasan, Monojit Choudhury
Workshop on Domain Adaptation for NLP, EACL (Virtual)   `AdaptNLP, EACL 2021`

W.1   **CoSSAT: Code-Switched Speech Annotation Tool.** 📄
Sanket Shah, Pratik Joshi, <u>Sebastin Santy</u>, Sunayana Sitaram
Workshop on Aggregating and Analysing Crowdsourced Annotations for NLP, EMNLP, Hong Kong   `AnnoNLP, EMNLP 2019`

J.1   **DataDepsGenerators.jl: Automatic generation of DataDeps.jl registration code.** 📄
Lyndon White, <u>Sebastin Santy</u>
Journal for Open Source Software   `JOSS 2018`

Mentions and Media Coverage

U.S. Federal Trade Commission (FTC) on *"Who is being left behind? Enforcement Priorities for a Tech [...]"* 📰   Aug 2022

The Gradient on *"New Technology, Old Problems: The Missing Voices in Natural Language Processing"* 📰   Apr 2022

Emily Bender [Blog] on *"On Academic Freedom and Ethics Review"* 📰   Jun 2021

Quartz on *"Data scientists are trying to make the internet accessible in every language"* 📰   Oct 2020

Sebastian Ruder [Blog] on *"Why You Should Do NLP Beyond English"* 📰   Aug 2020

Mint Lounge on *"Now a unique machine translation tool from Hindi to Gondi"* 📰   Aug 2020

Underrated ML on *"Language Independence and Material Properties"* 📰   Jun 2020

NLP Newsletter on *"Reviewing, Taking stock, Theme papers, Poisoning and stealing models, multimodal [...]"* 📰   Jun 2020

SIGTYP on *"Recent Developments in Computational Typology & Multilingual NLP"* 📰   Apr 2020

Times of India on *"Chhatisgarh: Now, tribals can listen to news in their own Gondi Language"* 📰   Aug 2019

The Caravan on *"Mind Our Language: The Koitur community is reclaiming their linguistic identity despite [...]"* 📰   Aug 2019

Hindustan Times on *"Gonds in Chhattisgarh get app for news in their language"* 📰   Aug 2019

ETV on *"Now tribals can hear news and stories in their own language [Translated]"* 📰   Aug 2019

Academic Service

PC/Reviewer	*Conferences:* WWW'23, AAAI'23, EMNLP'22, NAACL'22, ACL'22, **CHI'22***, ARR'21, **EMNLP'21***, ACL'21, EACL'21, MLADS'20, ICON'20
	Conference Demos: NAACL'22, ACL'22, EMNLP'21, NAACL'21, EMNLP'20
	Others: UW CSE PhD Applications '22
	(* = Outstanding Reviewer)
Sub-Reviewer	EMNLP'20, ACL'20, LREC'20, CODS-COMAD'20, CoNLL'19, Interspeech'19, ICON'19
Volunteer	ACL'20, AAAI'19, Panini Linguistics Olympiad (PLO) '19/'20
Tech	ACL Rolling Review, ACL'20 Virtual Conference

Invited Talks and Misc

NLP with Friends on *"The State and Fate of Linguistic Diversity and Inclusion in the NLP World"* ◉ ▪️   Nov 2020

PyData New York City on *"Repeatable Data Setup for Repeatable Science using Julia"* ◉ ▪️   Oct 2018

In conversation with Dr. APJ Abdul Kalam (President of India) on *"Technology progress around the world"* ▪️ Jun 2010

Software and Open Source Contributions

Microsoft. Interactive Neural Machine Translation - **git**, ◉

Mozilla. Bugzilla - **git**, Firefox (Gecko Engine) - **git**, TreeHerder - **git**

Others. scikit-learn - **git**, DDGenerators.jl - **git**, Virtual ACL 2020 - **git**, UserContext - ◯

BITS Pilani. BITS-Darshini - **git**, SWD - **git**, ARC - **git**, Quark 2017 - **git**, Waves 2016 - **git**

Teaching and Leadership Roles

Introduction to Human-Computer Interaction, CSE 440 @ UW. *Teaching Assistant*   Fall 2022

Speech & NLP Reading Group, Microsoft Research. *Organizer*   2019 - 2021

Open Source Development (OSD) Labs, BITS Goa. *Founding Member* ◯   2016 - 2018

Introduction to Computer Programming, CS F111 @ BITS. *Teaching Assistant*   Spring 2018

Introduction to Programming, Center for Technical Education (CTE), BITS Goa. *Co-Instructor*   Spring 2017

Waves, Quark, BITSMUN, (Festivals at BITS Goa). *Lead Developer*   2015-2017