

Niyati Bafna

✉ nbafna1@jhu.edu

in [LinkedIn](#) | [Github](#) | [Google Scholar](#)

OVERVIEW

I'm a PhD Computer Science student at CSLP, JHU, advised by Professor David Yarowsky. In the past, I worked as a research engineer at INRIA, ALMAAnaCH in Paris for a year, and even before that, I was an Erasmus scholar of the M.Sc. Erasmus Mundus Language and Communication Technologies programme, at Saarland University and Charles University. I'm especially interested in low-resource NLP, linguistic interventions in large language modelling, and code-switching/mixing.

EDUCATION

Johns Hopkins University <i>PhD Computer Science, Center for Language and Speech Processing</i>	Baltimore, Maryland 2023 - Present
Charles University, Saarland University <i>M.Sc. Computational Linguistics, Language and Science Technology, EMLCT</i>	Prague, Saarbrücken 2020 - 2022
Ashoka University <i>B.A. English, Computer Science</i>	Sonepat, Haryana 2017 - 2020

EXPERIENCE

ALMAAnaCH, Inria <i>Research Engineer</i>	Oct. 2022 - Jan. 2023
<ul style="list-style-type: none">• Worked on incorporating phoneme to morpheme level information in Transformer models for closely related languages on the Hindi Belt.• Studied robust approaches to bilingual lexicon induction between data-imbalanced language pairs for languages from the Indic dialect continuum.• Explored strategies for unsupervised machine translation and language modelling for 5 truly low-resource Indic languages.	
Charles University, UFAL ; Saarland University <i>Master's Thesis</i>	Feb. 2022 - Aug. 2022
<ul style="list-style-type: none">• Developed strategies for cognate induction for 26 languages of the Hindi Belt; collected monolingual data for 16 zero-resource dialects.	
Charles University, UFAL <i>Research Assistant</i>	Oct. 2021 - Jan. 2022
<ul style="list-style-type: none">• Worked on embeddings transfer from high-resource to (artificially) low-resource morphologically rich Indian languages (Hindi - Marathi, Nepali).• Participated in effort to harmonize 17 morphological databases• Worked on constrained decoding for technical term retention in En-Hi MT	
Amazon India, Bangalore <i>Applied Scientist Intern</i>	May 2020 - Aug. 2020
<ul style="list-style-type: none">• Worked on developing and productionizing 'language-agnostic' multilingual NER models for English, French, German.	
Indian Institute of Technology, Delhi <i>Research Intern</i>	Jan. 2020 - May. 2020
<ul style="list-style-type: none">• Worked on building computational models to understand clause-final verb prediction, Hindi.	
Indian Institute of Information Technology, Hyderabad <i>Research Intern</i>	Jun 2019 - Aug. 2019
<ul style="list-style-type: none">• Worked on handling verb phrase ellipsis in English-Hindi MT.	

AWARDS AND OLYMPIADS

- Erasmus Mundus LCT scholarship for a 2-year funded Masters
- Merit scholarship for well-performing students, Charles University
- Department English Prize, Ashoka University, 2020
- Summa Cum Laude (Undergraduate)
- Selection to Indian National Mathematics Olympiad, in 2015, 2016
- Hon. Mention at the Panini Linguistics Olympiad, 2015

SKILLS

Programming Languages/Frameworks: Python, C++, Java, PyTorch, HuggingFace Transformers
Natural Languages: Hindi, English, French, Marathi

COMMUNITY

- CLSP Application Support Program, JHU** 2023
Volunteer
- Contributing feedback for PhD applications to prospective students from underrepresented backgrounds.
- AUWS Mentorship Programme** 2022
Mentor
- Providing guidance re. Masters/PhD/job applications, internships, and coursework to 2nd-3rd year (female) undergraduates.
- Women in Computing Society, Ashoka University** 2019 - 2020
Head of Events
- Organized workshops, tutorials, talks to promote inclusivity in tech.
- Sandhi Linguistics Society, Ashoka University** 2018 - 2020
Founding Member
- Organized reading groups, talks, and events around linguistics and language learning.

PUBLICATIONS

- **Niyati Bafna**, Philipp Koehn, David Yarowsky. **Pointer-Generator Networks for Low-Resource Machine Translation: Don't Copy That!**. 2024. Accepted at Workshop on Insights from Negative Results in NLP, NAACL.
- **Niyati Bafna**, Cristina España-Bonet, Josef van Genabith, Benoît Sagot, and Rachel Bawden. 2023. **When your Cousin has the Right Connections: Unsupervised Bilingual Lexicon Induction for Related Data-Imbalanced Languages**. Accepted at LREC-COLING 2024.
- **Niyati Bafna**, Cristina España-Bonet, Josef van Genabith, Benoît Sagot, and Rachel Bawden. 2023. **Cross-Lingual Strategies for Low-Resource Language Modeling: A Study on Five Indic Dialects** In Proceedings of the 18th Conference on Traitement Automatique des Language Naturelles. Paris, France. TALN.
- **Niyati Bafna**, Josef van Genabith, Cristina España-Bonet, and Zdeněk Žabokrtský. 2022. **Combining Noisy Semantic Signals with Orthographic Cues: Cognate Induction for the Indic Dialect Continuum**. In Proceedings of the 26th Conference on Computational Natural Language Learning, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.
- **Niyati Bafna** and Zdeněk Žabokrtský. 2022. **Subword-based Cross-lingual Transfer of Embeddings from Hindi to Marathi and Nepali**. In Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, pages 61–71, Seattle, Washington. Association for Computational Linguistics.

- Zdeněk Žabokrtský, **Niyati Bafna**, Jan Bodnár, Lukáš Kyjánek, Emil Svoboda, Magda Ševčíková, and Jonáš Vidra. 2022. **Towards Universal Segmentations: UniSegments 1.0**. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 1137–1149, Marseille, France. European Language Resources Association.
- **Niyati Bafna**, Martin Vastlik, and Ondrej Bojar. 2021. **Constrained Decoding for Technical Term Retention in English-Hindi MT**. In Proceedings of the 18th International Conference on Natural Language Processing, pages 1–6, National Institute of Technology Silchar, India. NLP Association of India.
- *Kartik Sharma, ***Niyati Bafna**, and Samar Husain. 2021. **Clause Final Verb Prediction in Hindi: Evidence for Noisy Channel Model of Communication**. In Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics, pages 160–170, Online. Association for Computational Linguistics.
- **Niyati Bafna** and Dipti Sharma. 2019. **Towards Handling Verb Phrase Ellipsis in English-Hindi Machine Translation**. In Proceedings of the 16th International Conference on Natural Language Processing, pages 150–159, International Institute of Information Technology, Hyderabad, India. NLP Association of India.

INVITED TALK

- **Empirical Models for an Indic Dialect Continuum**. Linguistic Mondays, Institute of Formal and Applied Linguistics, Charles University, Czech Republic. Faculty of Mathematics and Physics.

OTHER THINGS ABOUT ME

- I have a FIDE chess rating of 1609.
- I played basketball for my city as a teenager. I’ve also been rejected at Maharashtra nationals trials *four times*.
- I’ve translated a **book** from Hindi to English!