



# Niyati Bafna

<http://niyatibafna.github.io/>  
[nbafna1@jh.edu](mailto:nbafna1@jh.edu)

## EDUCATION AND EMPLOYMENT

**M.Sc. Computational Linguistics, Language and Science Technology, EMLCT - 2022**  
Charles University, Saarland University  
GPA: 1.0 (1-5, lower being better)

**B.A. English, Computer Science - 2020**  
Ashoka University, India, 2020  
GPA: 3.96/4, Summa Cum Laude

**Research Engineer, Inria, ALMAnaCH, Paris, Oct 2022- June 2023**  
Supervisors: Benoît Sagot, Rachel Bawden  
Working on incorporating phoneme to morpheme level information in Transformer models for closely related languages on the Hindi Belt

## FIND ME ON...

- [LinkedIn](#)
- [Github](#)
- [Google Scholar](#)

## POSITIONS HELD

- Peer reviewing for ICON '21
- Head of Women in Computing Society, Ashoka Uni, 2019
- Founding member of linguistics society, Sandhi, Ashoka, 2017
- Vice-capt. of Ashoka Basketball team

## HOBBIES

- [Blogging](#)
- Learning German (A2) and French (B2)
- Tennis, chess, afternoon naps...

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, ALMAnaCH 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.

## CURRENTLY...

**PhD Computer Science, Johns Hopkins University, Center for Language and Speech Processing (starting Aug 2023)**  
Advisor: Professor David Yarowsky

## PROJECTS AND INTERSHIPS

### Research Project

*ALMAnaCH, Inria | October 2022 - January 2023*

Introduced a strategy for bilingual lexicon induction between data-imbalanced language pairs, working with languages from the Indic dialect continuum. Paper under review.

### Research Project

*ALMAnaCH, Inria | October 2022 - January 2023*

Explored strategies for language modelling for 5 truly low-resource Indic languages. [Presented paper at TALN-23.](#)

### Master's Thesis

*Charles Uni., Saarland Uni | February 2022 - August 2022*

Developed strategies for cognate induction for 26 languages of the Hindi Belt; collected monolingual data for 16 zero-resource dialects. [Presented paper at CoNLL 2022.](#)

### Research Intern

*Charles University, UFAL | May 2021 - Aug 2021*

Worked on embeddings transfer from high-resource to (artificially) low-resource morphologically rich Indian languages (Hindi - Marathi, Nepali). [Presented paper at SIGMORPHON \(NAACL '22\).](#)

### Research Project Member

*Charles University, UFAL | March 2021 - Aug 2021*

Worked on constrained decoding for technical term retention in En-Hi MT. [Short paper accepted at ICON-'21.](#)

### Applied Scientist Intern

*Amazon India, Bangalore | May 2020 - Aug 2020*

Worked on developing and productionizing 'language-agnostic' multilingual NER models for English, French, German.

## COURSE PROJECTS

- **Building Information Retrieval pipeline from scratch**
- **Marathi morphanalyser within transducer framework**
- **Topic modelling on Enron dataset**
- **Building recommendation system for short stories**
- **Tracking Religious Hate Speech on Indian Twitter**
- **Finetuning DPR for Dense Retrieval (ODQA)**
- **Building rule based lexicon-builder for Hindi-English**

## OTHER THINGS ABOUT ME

- Received a merit scholarship for well-performing students from Charles University for 2020-21
- Attended Canada/US Mathcamp, 2016
- Cleared Regional Math Olympiad twice, 2015, 2016
- Hon. Mention at the Panini Linguistics Olympiad 2015
- I've translated a book from Hindi to English!

## PROJECTS AND INTERNSHIPS (CTD.)

### Research Intern

*Indian Institute of Technology, Delhi | Jan 2020 - May 2020*

Worked on building computational models to understand clause-final verb prediction, Hindi. Presented paper at CMCL, NAACL '21.

### Research Intern

*Indian Institute of Information Technology, Hyderabad | Jun - Aug 2019*

Worked on handling verb phrase ellipsis in English-Hindi MT. Presented paper at ICON 2019.

### Research Group Member

*Charles University, UFAL | October 2021: Jan 2022*

Worked with research team on harmonizing 17 morphological resources. Paper accepted at LREC-'22.

## PUBLICATIONS

- **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, Caech Republic. Faculty of Mathematics and Physics.