# ANASTASIA KUZNETSOVA

**Graduate Research Assistant | PhD student**

@ anakuzne@iu.edu    ☎ (812) 558-80-55    ⚲ GitHub: ana-kuznetsova    in linkedin.com/in/anastasia-kuznetsova-2bb66b116/

## EDUCATION

**PhD student in Computer Science and Computational Linguistics (3.48 GPA)**
**Indiana University**
📅 Aug 2019 – May 2024    📍 Bloomington,IN, USA)

**MA in Computational Linguistics**
**NRU Higher School of Economics**
📅 Sep 2017 – Jun 2019    📍 Moscow, Russia

## EXPERIENCE

**Software Developer**
**MTS, Artificial Intelligence Department**
📅 Jan–Jul 2019    📍 Moscow,Russia

**Mentor**
**Google Summer of Code, Google Code-In, Apertium**
📅 2018, 2019, 2020    📍 Remote

**Student Participant**
**Google Summer of Code, Apertium**
📅 May - Aug, 2018    📍 Remote

## SKILLS

- Languages: Python, R, C++
- Libraries: PyTorch, Torchaudio, Tensorflow, NLTK, Scikit learn, OpenAI gym.

## PUBLICATIONS

- *Kuznetsova, Anastasia, Anurag Kumar, and Francis Tyers.* A bandit approach to curriculum generation for automatic speech recognition (Submitted ICASSP 2021)
- *Zueva, Anna, Anastasia Kuznetsova, and Francis Tyers.* "A finite-state morphological analyser for evenki." Proceedings of The 12th Language Resources and Evaluation Conference. 2020.

## LANGUAGES

- Russian (Native)
- Portuguese (Fluent)
- Spanish (intermediate)
- Lithuanian (intermediate)
- Guarani (basic knowledge)

## CURRENT PROJECTS

**Reinforcement Learning based Speech Enhancement**
Ongoing project explores the possibilities of exploiting RL algorithms in speech enhancement area in low-data setting. Re-implemented and tested existing RL solutions in the domain, working on developing my own approach to the problem.

**A reinforcement-learning approach to curriculum generation for ASR**
Developed bandit-based approach to an Automated Curriculum Learning paradigm in order to mitigate the lack of training data in low-resource setting. The implemented bandit framework shows a 10% WER and 27% CER improvement over the baseline model and the potential of the further investigation of RL algorithms in ASR.

## PAST PROJECTS

**Google Summer of Code: Machine Translation for Guarani–Spanish language pair**
Leveraged FST-based solution for dictionary and transfer rules implementation using HFST framework.

**Speaker Identification system**
Implemented RNN-based speaker ID system exploiting Siamese networks and speaker embeddings with one-shot learning approach (Tensorflow).

**BPE weighting of morphological analyser for Paraguayan Guarani**
Weighting of the morphological analyser based on finite-state technology with Byte Pair encoding algorithm.

**Morphological Disambiguation for Paraguayan Guaraní**
Developed rule-based grammar (Constraint Grammar formalism).

**Named Entity Recognition for Russian Popular Science corpora**
- Implemented rule-based entity extraction using Tomita parser;
- Developed annotation scheme
- Annotated Popular Science Corpus with named entities.