Lisa Bylinina

bylinina@gmail.com / bylinina.github.io +31(0)6 183 50 454

PhD in linguistics, researcher and language technology practitioner with 10+ years of experience in theoretical and data-driven linguistic analysis using deep learning, formal languages, and logic. Results published in top linguistic venues (ACL main session, Journal of Semantics and others). Interested in applying my expertise to computational tasks such as natural language processing, understanding and generation.

Highlights

- Design and implementation of solutions to the whole landscape of NLP tasks using cutting-edge ML/DL techniques.
- Studies in language processing comparing human speakers and language models; studies on multilinguality and language variation, with the focus on under-described and low-resource languages.
- Numeral Typology Database: Created a dataset with grammatical properties of number words across languages.
- Industry application of research results on proper name variation in Yandex search ranking algorithm.
- $\approx$ 40 peer-reviewed publications, $\approx$ 60 conference talks, 200+ students taught.

Technical and Data Skills

Python, Pytorch, Keras, Transformers, Pandas, Numpy, all major ML/DL approaches and tools.

Experience

July 20 – Present: Bookarang.com, Netherlands. NLP and ML engineer.

- Developed and implemented several solutions for text analysis (style and topic detection, target age group detection) using cutting-edge NLP techniques and models.
- Full cycle from method research and design to implementation.

November 16 – 20: Leiden University, Netherlands. Postdoctoral Researcher at Leiden University Centre for Linguistics. Principal investigator in a research project on number and quantity in natural language.

- Developed and pursued a research program combining logic, number cognition and language variation.
- Full-cycle budget management; designed the study, hired and managed a team of 16 interns, research assistants in a large-scale data collection project.
- On-site data collection. Design and analysis of a series of linguistic experiments.
- Collected data from 37 languages and developed cutting-edge linguistic theories, resulting in:
  - 13 peer-reviewed papers and 22 talks in international conferences.
  - Numeral Typology Database: Dataset representing grammatical properties of number words across languages.

- Authored study materials and applied them for teaching and mentoring.
- Editorial board member of *Semantics and Pragmatics* journal. Organizer of an international workshop.
- Won a grant in Digital Humanities to advance storage, handling and analysis of cross-linguistic data.

November 13 – November 16: Meertens Institute, the Royal Netherlands Academy of Arts and Sciences, Amsterdam, Netherlands. Postdoctoral Researcher in a research program 'Knowledge and Culture' (Utrecht / Leiden / Amsterdam).

- Developed and tested scientific theories of language structure and meaning, utilizing varied research methodologies including original elicitation work. Collected and analyzed data on quantity expressions in natural language.
- Hired and managed a small team of interns / student research assistants for data collection. Taught and supervised.
- Dissertation published as a book (ISBN: 978-94-6093-130-7)
- 6 papers, 12 conference / workshop presentations.
- Won a highly competitive early career Dutch 'talent scheme' grant.

November 09 – September 13: Utrecht University, Netherlands. PhD Researcher at Utrecht Institute for Linguistics.

- Synthesized a large amount of data from several languages, including English, Russian, Hungarian and Japanese.
- Wrote a thesis, using formal logics to model vagueness and subjectivity in natural language.
- 10 peer-reviewed publications; 20 conference / workshop presentations worldwide.

March – December 09: Yandex LLC, Moscow, Russia. Linguist in Group of Linguistic Projects, IR / NLP.

- Using statistical data-driven methods, enriched lexical representations to enhance IR and other language-related tasks.
- Research on Yandex ranking algorithm: context-dependent proper name forms in search queries.

January 07 – January 08: EPAM Systems, Moscow, Russia. Computational linguist in an NLP project for Clarabridge.

- Designed the linguistic component for a sentiment analysis system for Russian and English.

Education

2014 PhD in Linguistics, **Utrecht University**, the Netherlands. *Detailed under 'Experience'*.

2005 MA (cum laude) in Theoretical and Computational Linguistics, **Moscow State University**, Russia.

- Computational / math courses taken: logic, probability theory, statistics, calculus, algebra, geometry, formal languages, information theory, artificial intelligence.
- 5 publications, conference presentations; won several fellowships, produced formal descriptions of fragments of under-described languages (8 field-trips).

Languages: Russian (native), English (near-native), Dutch (intermediate), French (intermediate).

Latest academic CV: [pdf]