

Olga SOZINOVA

PERSONAL DATA

ADDRESS: Giebeleichstrasse 9, 8152 Glattbrugg, Switzerland
PHONE: +41 78 684 89 19
WEBSITE: <http://sozinova.com>
EMAIL: olga@sozinova.com

EDUCATION

Main education

2018–PRESENT PhD in COMPUTATIONAL LINGUISTICS,
URPP Language and Space,
University of Zurich
2016–2018 M.A. in LINGUISTICS,
University of Bern
2012–2016 B.A. in FUNDAMENTAL AND APPLIED LINGUISTICS,
HSE University, Moscow

Exchange studies

2015–2016 Exchange semester in COMPUTATIONAL LINGUISTICS
University of Tübingen
2014 Exchange semester in LINGUISTICS
University of Bern

Summer schools (selected)

2019 9TH LISBON MACHINE LEARNING SCHOOL, LxMLS
Instituto Superior Técnico, Lisbon
2019 REVISITING RESEARCH TRAINING IN LINGUISTICS: THEORY, LOGIC, METHOD
Petnica Science Centre, Valjevo

SCHOLARSHIPS

2016–2018 Master Grant of the University of Bern
2015–2016 Oxford Russia Fund Scholarship
2014–2016 Increased State Academic Scholarship,
for noteworthy academic and research achievements

PUBLICATIONS

Bentz, C., X. Gutierrez-Vasques, T. Samardzic and O. Sozinova (2021). Complexity Trade-Offs and Equi-Complexity in Natural Languages: A Meta-Analysis. *Linguistics Vanguard*, De Gruyter. (to appear)

Ruzsics, T., O. Sozinova, X. Gutierrez-Vasques and T. Samardzic (2021). [Interpretability for Morphological Inflection: from Character-level Predictions to Subword-level Rules](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 3189–3201.

Gutierrez-Vasques, X., C. Bentz, O. Sozinova and T. Samardzic (2021). [From characters to words: the turning point of BPE merges](#). In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 3454–3468.

Sozinova, O. (2016). [Complex Networks-Based Approach to Russian Rhyme History Description: Linguostatistics and Database](#). *Digital Humanities 2016, Conference Abstracts*, 891–893.

Arkhangel'skii, T. and **O. Sozinova** (2015). [A multimedia corpus of the Yiddish language](#). *Automatic Documentation and Mathematical Linguistics*, 49(2), 47–53.

CONFERENCE PRESENTATIONS (SELECTED)

Sozinova, O. and T. Samardzic (2021). Subword Geometry: Picturing Word Shapes. *Workshop of the ACL Special Interest Group on Typology, co-located with the North American Chapter of the Association for Computational Linguistics*, Online. (upcoming)

Bentz, C., **O. Sozinova** and T. Samardzic (2019). Collecting a corpus for 100 typologically diverse languages (100LC). *Workshop on language documentation: multilingual settings and technological advances*. Uppsala, Sweden.

Sozinova, O., T. Samardzic and C. Bentz (2019). Measuring inflectional and derivational complexity. *Interactive Workshop on Measuring Language Complexity*. Freiburg, Germany.

WORK EXPERIENCE

Teacher

- 2020 Tutor at the course *Processing Non-standard Language*, autumn semester
University of Zurich
- Preparing and checking programming assignments: Python, Moses SMT, NMT, baseline and advanced ML solutions
 - Installing, training and testing a [g2p-seq2seq transformer model](#) for normalizing Swiss German data into Standard German
- 2020 Tutor at the course *Techniques of Semantic Analysis (MA)*, spring semester
University of Zurich
- Preparing and checking programming assignments: Python, cosine similarity for vector space, lineal algebra, matrix transformations, introduction to ML and NN
- 2019 Tutor at the course *Programming for Linguists (MA)*, autumn semester
University of Zurich
- Preparing and checking programming assignments: Python, R, explaining basic notions (variable, operator, loop, etc.), explaining basic text processing and data analysis
 - Giving a lecture on *Working with web pages in Python*

Software Developer

- 2020 Assistant at the [sBayes](#) project
University of Zurich
- Restructuring the Python code, designing a well-organized pipeline
 - Designing JSON configuration files
 - Writing documentation parts

Web Developer

- 2018–2019 Full-stack web developer of the *Zurich Tangram Corpus*
University of Zurich
- Backend: Python Flask application based on *Tsakorpus platform*, data preparation (converting transcription files, mapping images to search fields)
 - Frontend: UI development (input fields, video and audio player adjustment)
- 2016 Frontend developer of the *Beserman Dictionary*
HSE University, Moscow
- JavaScript, jQuery (keyboard pop-up, storing and showing search requests, displaying search results)
 - HTML and CSS design (UI, search fields)
- 2014 Frontend developer of the *Yiddish Multimedia Corpus*
HSE University, Moscow
- JavaScript video player, synchronizing video with transcription
 - Preparing Python Django localization files (Russian and English)
 - Full web design

Computational Linguist

- 2018 Linguist Consultant at **Lionbridge Technologies Inc.**
Tampere, Finland (remote)
- Preparing phonemic transcriptions in Russian
 - Recordings quality control
- 2015–2016 Junior Linguist at **Dialogflow** (former Speaktoit, API.AI)
Palo Alto, CA, USA (office in Moscow)
- Developing conversational AI in German and English
 - Technical support for the German speaking users
 - Writing blog articles on the use of the platform

OUTREACH & ACADEMIC SERVICE

- 2021 Co-organizer of a booth at *Scientifica – Zurich Science Days* (upcoming)
ETH Zurich
- Writing a proposal on the topic *Natural vs. Artificial Languages*
 - Planning and coordinating a web app development: writing a collaborative story with a GPT-2 model in four languages
- 2021 Reviewing papers for:
- the Workshop of the ACL Special Interest Group on Typology (SIGTYP)
 - the Annual Meeting of the Association for Computational Linguistics (ACL)
- 2020 Reviewing papers for:
- the International Conference on Computational Linguistics (COLING)
 - the Conference on Empirical Methods in Natural Language Processing (EMNLP)
 - the Annual Meeting of the Association for Computational Linguistics (ACL)
- 2019 Reviewing papers for:
- the Workshop on Balto-Slavic Natural Language Processing (BSNLP)
 - the Fest of Empirical Syntax (SyntaxFest)

LANGUAGES

RUSSIAN:	Native
ENGLISH, GERMAN:	Fluent
FRENCH:	Intermediate
CHINESE, SERBIAN, BULGARIAN:	Elementary

TECHNICAL SKILLS

Programming languages

Python, R, Shell (Unix/Linux), Java (basic);
text processing, data mining (web crawling), data analysis, statistics

Software development

Git, architecture design:
Python classes, JSON files (configuration), general pipeline

Web development

Python Django and Flask, HTML5, CSS, JavaScript,
jQuery/AJAX (basic), SSL certificates (basic)

Databases

MySQL, Neo4j, MongoDB (basic), Elasticsearch (basic), REST API (basic)

Data visualization

R ggplot (static and animated), D3.js and Sigma.js (dynamic)

Markup

Latex, Markdown

Digital graphics

Photoshop, Illustrator, Krita (basic), Inkscape (basic)