

SUMMARY A versatile data scientist, software architect, instructor, and researcher with a proven track record of solving complex challenges in data-driven research. Provided technical leadership and methodological expertise to interdisciplinary projects. Enabled new research and empowered scholars by developing and deploying digital resources and tools. Architect of AUTOTYP, the world's largest database of linguistic structures.

Extended academic CV is available at tinyurl.com/mus3ys3u

EXPERIENCE **DEPARTMENT OF COMPARATIVE LANGUAGE SCIENCE • UNIVERSITY OF ZURICH • SWITZERLAND** **since 2011**

Data Scientist & Scientific Programmer (since 2014)

PhD Student (2011-2014)

- Carried out highly complex analyses of heterogeneous data using various statistical and data mining methods, created visualizations and layered maps, interpreted the results, wrote reports
- Built data integration and transformation pipelines powering interdisciplinary studies, promoted clean data architecture, prepared and reviewed supplementary code for publications
- Created tailored software solutions and machine learning models for cutting-edge research, led teams developing digital resources and databases, authored R packages and open-source tools
- Developed theoretical concepts and methods enabling novel research on linguistic diversity, created models of complex phenomena, participated in theory-building, supervised research teams
- Secured ~\$1'500'000 in research funds through co-writing grants, authored and contributed to dozens of published papers and datasets, presented at international conferences
- Led the technical development of AUTOTYP, the world's largest database of linguistic structures, developed the architecture, published the database as an openly accessible FAIR resource
- Designed the Zurich Database of Agreement in Italo-Romance, developed the corpus annotation tool and the database query backend, contributed to the development of the web frontend
- Pioneered the data-as-code approach for building linguistic databases and developed domain-specific data entry languages for research projects (Tymber database, Evolution of Linguistic Units)
- Developed software tools and infrastructure for working with corpora of endangered languages, curated endangered language data, typeset community dictionaries
- Supported interdisciplinary research project teams and individual researchers with research design and method selection, sustainable data modeling, computational infrastructure, and writing code
- Mentored PhD students and PostDocs and trained researchers on the best practices in data science and programming, improving the quality of the research output

Consultant IT Strategy and Cloud Computing (since 2021)

- Supported the department in matters related to IT services and operations, advised on technology and platform choices, developed workflows, and curated technical documentation
- Represented the department in committees such as the UZH Cloud Strategy Taskforce, consulted for the Linguistic Research Infrastructure Center, negotiated with the central IT Services
- Managed the Cloud & Cluster Computing policies and resource usage, interfaced with the Research IT unit, and offered continuous education masterclasses on cloud technologies and usage
- Developed tools for cluster job submission and monitoring, substantially reducing the cloud computing expenses of the department

Mandated Instructor at the Faculty of Arts and Social Sciences (since 2013)

- Taught a wide range of undergraduate courses, with a focus on programming, data science, and database technology, supervised BA and MA theses
- Developed and coordinated the Quantitative Methods study module group for the Linguistics curriculum, significantly improving students' proficiency in data science and programming

Manager IT Resources and Strategy (2014-2020)

- Managed IT resources and strategy, ensuring that the needs of the department were met during its transformation into an internationally acclaimed National Research Center

- Oversaw the annual IT budget of CHF 60,000, prepared funding requests for the acquisition of specialized hardware, supervised the IT support team, and interfaced with the central IT Services
- Identified strategic needs, implemented and maintained new critical services (version control, backups, HPC, project and bibliography management, shared data resources)
- Drove the adoption of version control and cluster computing usage throughout the department, improving the productivity and the quality of the scientific output
- Streamlined the IT services and the onboarding process, negotiated a shared IT support model with the central IT Services, significantly reducing the operating costs and the administrative overhead

DEPARTMENT OF GENERAL LINGUISTICS • UNIVERSITY OF LEIPZIG • GERMANY

2006 - 2011

Teaching and Research Assistant

- Built data export and transformation pipelines, analyzed data, contributed to scientific publications, and presented at conferences and workshops
- Mentored students, graded homework and exams

AREAS OF EXPERTISE

Data Science and Statistical Analysis

- **Skills:** data wrangling and visualization, multilevel Bayesian models, stochastic simulations, machine learning classifiers, phylogenetic methods, geospatial analysis, clustering, time series analysis, clustering, graph structure analysis, mathematical optimization, supervised and unsupervised learning
- **Tools:** R (tidyverse, brms, sf/s2/terra), Shiny, Python (pandas, scikit-learn, PyTorch, JAX), Stan, GDAL/OGR

Programming and Software Engineering

- **Skills:** system and API design, optimization, design patterns, algorithms and data structures, data-oriented programming, functional programming, parallel and concurrent programming, virtualization/containers, application development, Web standards, Cloud Computing, HPC, GPGPU, SIMD, shell scripting, testing, version control, continuous integration, agile methodology
- **Tools:** Python, C, C++, Rust, R, Swift, JavaScript/TypeScript, Lua, bash/zsh, Svelte, React, SwiftUI, GTK, HTML5/CSS3, CUDA, Metal, OpenGL, OpenCL, Docker, Slurm, OpenStack, AWS, GCP, git, GitHub/GitLab CI

Database Design and Integration

- **Skills:** data and relationship modeling, database theory and design, data-as-code solutions, domain-specific language design, data queries, API integration, geospatial processing
- **Tools:** R, Python, SQL (various dialects), RESTful APIs, GraphQL, Google BigQuery, MongoDB, Neo4J, QGIS, phpMyAdmin, FileMaker

OTHER SKILLS Communication and Stakeholder Engagement • Project management • Research support • IT Operations and Services • Technical Writing • Presentation • Instruction • Digital Humanities • Linguistic Theory • Corpus Linguistics • Quantitative Linguistic Typology

EDUCATION **PhD in Quantitative Linguistics • 2015 • University of Zurich • Switzerland**
Thesis "Digging Deep — Mining Corpora for Typological Patterns and Beyond"

MA in Linguistics, Mathematics, and Computer Science • 2009 • University of Leipzig • Germany