

Taras Zakharko

CURRICULUM VITAE

Department of Comparative Language Science • University of Zurich
Affolternstrasse 56 • CH-8050 Zurich

✉ taras.zakharko@uzh.ch
in [linkedin.com/in/tzakharko](https://www.linkedin.com/in/tzakharko)
🐙 github.com/tzakharko

A methodologically focused linguist, data scientist, and software architect with a proven track record of solving complex challenges in cutting-edge research. Passionate about developing open digital resources and tools that empower researchers.

EMPLOYMENT

- since 2011** **University of Zurich, Department of Comparative Language Science**
Data Scientist & Scientific Programmer (since 2014)
Mandated Lecturer at the Faculty of Arts and Social Sciences (since 2013)
Consultant IT & Cloud Computing (since 2020)
Manager IT Resources and Strategy (2014-2020)
PhD Student (2011-2014)
- 2006-2011** **University of Leipzig, Department of Linguistics**
Teaching and research assistant

EDUCATION

- 2015** **PhD in General Linguistics (summa cum laude)**
Dissertation: *Digging deep: mining corpora for typological patterns and beyond*
University of Zurich
Supervisors: Balthasar Bickel, Volker Gast
- 2009** **Magister Artium General Linguistics, Computer Science, and Mathematics**
MA thesis: *Aspects of structure in Chintang child-surrounding speech*
University of Leipzig

LANGUAGES

English (fluent)
German (fluent)
Ukrainian (native)

RESEARCH PROJECTS

- since 2021** **The Evolution of Linguistic Units**
Consultant data modeling and analysis, implemented the database model
- since 2017** **TYMBER: database of nominal number**
Consultant data modeling and analysis, implemented the domain-specific language for data entry
- since 2009** **AUTOTYP Typological Databases Network**
Database architect and technical lead, implemented data processing and transformation pipelines, carried out quantitative analyses, led the conversion from a proprietary file format to an openly accessible dataset

- since 2009 Chintang Language Research Program**
Curated the corpus database and language materials, implemented tools and data pipelines, carried out corpus analyses, typeset the community dictionary
- 2016-2019 Linguistic morphology in time and space**
Consultant data modeling, carried out statistical and geospatial analyses, supervised a team of research assistants
- 2015-2018 The Zurich Database of Agreement in Italo-Romance**
Co-wrote the project proposal, designed and implemented the database model and the web query backend, developed corpus annotation tools, handled data import and archiving
SNF project grant 156530, **CHF 666'064***
- 2014-2019 Acquisition Processes in Maximally Diverse Languages**
Corpus database consultant, implemented tools and data pipelines
- 2012-2015 The Relative Frequencies of Nouns, Pronouns, and Verbs**
Co-applicant (with Frank Seifart and others)
Co-wrote the project proposal, carried out statistical analyses
VolkswagenStiftung project grant, **EUR 144'992***
- 2012-2015 Referentiality**
Co-applicant (with Christian Lehmann and others)
Contributed to the project proposal, supervised a team of research assistants
VolkswagenStiftung project grant, **EUR 502'700***
- 2011-2014 Towards a Corpus-Based Typology of Clause Linkage**
PhD research project (main applicant Balthasar Bickel)
Co-wrote the project proposal, supervised a team of research assistants, developed the annotation software, conceptualized and carried out the research
SNF project grant 140881, **CHF 234'195***

* grant size indicated for projects as applicant or with significant contribution to the project proposal

PROFESSIONAL AND ACADEMIC SERVICE

- 2024** Consultant linguistic research infrastructure for working group *Future of Cloud Computing at the University of Zurich*
- since 2020** Consultant scientific computational infrastructure, liaison to the supercomputing services group at the Department of Comparative Linguistics, University of Zurich
- since 2019** Coordinator study module group *Quantitative and Computational Methods*
- 2022** Teaching personnel representative in appointment committee for *Assistant Professor with Tenure Track in Descriptive and Anthropological Linguistics*, University of Zurich
- 2020** Member of the *Technology Platform Taskforce* for the creation of the *University of Zurich Linguistic Research Infrastructure* group, advisor to the board
- 2018** Member of the Linguistics Programme reform committee *Bologna 2020*, responsible for developing new curricula for quantitative and computational methods in linguistics
- 2015** Linguistics Department representative in the working group *Cooperative IT Support at the University of Zurich*
- 2012-2015** Founding member for the *zling* online publishing platform for the Linguistics PhD Programme, responsible for web interface and database design

TEACHING

Courses as an instructor at the University of Zurich

Data Science for Linguists. Spring semester 2020, 2021, 2022, 2023, 2024

Structuring Linguistic Data. Fall semester 2019, 2020, 2021, 2023

Quantitative Methods. Spring semester 2016, 2017, 2018, 2019

Study and Research Skills. Fall semester 2015, 2016, 2018

Semantics and Pragmatics. Spring semester 2013, 2015

Summer schools (instructor)

Data transformation for PhDs and PostDocs in Linguistics. Summer School *Quantitative Analysis of Linguistic Data*, Zurich. 2014.

Introduction to corpus analysis with R. Workshop on *Current Tendencies and Potentials in Language Documentation*, Campinas. 2013.

Introduction to R for quantitative data analysis. CLARA *Advanced Resource Creation, Archiving, and Usage*. Nijmegen. 2010

Thesis supervision

Huber Meili. 2024. Typology of Noun Phrase Structures: a multivariate analysis. MA Thesis, Zurich.

Ruben Mögel. 2020. Visualising typological dissimilarity on global scale. BA Thesis, Zurich.

Valerie Ledermann. 2019. Recursion in linguistics. BA Thesis, Zurich.

Katrin Rechsteiner. 2015. Evidentialität in den germanischen Sprachen. BA Thesis, Zurich.

PUBLICATIONS AND DATASETS

Peer-reviewed

Graff, Anna; Chousou-Polydouri, Natalia; Inman, David; Skirgård, Hedvig; Lischka, Marc; Zakharko, Taras; Barbieri, Chiara; Bickel Balthasar. Submitted. Curating global datasets of structural linguistic features for independence. *Scientific Data*.

Graff, Anna; Lischka, Marc; Zakharko, Taras; Furrer, Reinhard; Bickel, Balthasar. Submitted. densify: An R package to prune sparse data frames of typological linguistic data. *Journal of Open Source Software*.

Ivani, Jessica K. & Zakharko, Taras. In press. Phasal polarity in Suansu. *Linguistics of the Tibeto-Burman Area*.

Witzlack-Makarevich, Alena; Nichols, Johanna; Hildebrandt, Kristine; Zakharko, Taras; Bickel, Balthasar. Managing AUTOTYP Data: Design Principles and Implementation. 2022. In: Berez-Kroeker, Andrea L; McDonnell, Bradley; Koller, Eve; Collister, Lauren B. *The Open Handbook of Linguistic Data Management*. Cambridge: MIT Press, 632-642.

Witzlack-Makarevich, Alena; Zakharko, Taras; Bierkandt, Lennart; Zúñiga, Fernando; Bickel, Balthasar. 2016. Decomposing hierarchical alignment: Co-arguments as conditions on alignment and the limits of referential hierarchies as explanations in verb agreement. *Linguistics*, 54(3):531-561.

Bickel, Balthasar; Witzlack-Makarevich, Alena; Zakharko, Taras; Iemmolo, Giorgio. 2015. Exploring diachronic universals of agreement: alignment patterns and zero marking across person categories. In: Fleischer, Jürg; Rieken, Elisabeth; Widmer, Paul. *Agreement from a diachronic perspective*. Berlin: De Gruyter Mouton, 29 - 51.

- Stoll, Sabine; Zakharko, Taras; Moran, Steven; Schikowski, Robert; Bickel, Balthasar. 2015. Syntactic mixing across generations in an environment of community-wide bilingualism. *Frontiers in Psychology*, 6:82.
- Bickel, Balthasar; Zakharko, Taras; Bierkandt, Lennart; Witzlack-Makarevich, Alena. 2014. Semantic role clustering: an empirical assessment of semantic role types in non-default case assignment. *Studies in Language*, 38(3):485 - 511.
- Bickel, Balthasar; Witzlack-Makarevich, Alena; Zakharko, Taras. 2014. Typological evidence against universal effects of referential scales on case alignment. In: Bornkessel-Schlesewsky, Ina; Malchukov, Andrej; Richards, Marc. *Scales and Hierarchies: a cross-disciplinary perspective on referential hierarchies*. Berlin: De Gruyter Mouton, 7-44.
- Bickel, Balthasar; Iemmolo, Giorgio; Zakharko, Taras; Witzlack-Makarevich, Alena. Patterns of alignment in verb agreement. 2013. In: Bakker, Dik; Haspelmath, Martin. *Languages across boundaries. Studies in memory of Anna Siewierska*. Berlin: De Gruyter Mouton, 15-36.

Manuscripts

- Zakharko, Taras. 2017. *Digging deep: mining corpora for typological patterns and beyond*. Doctoral Dissertation. University of Zurich, Faculty of Arts.

In progress

- Bickel, Balthasar; Ledermann, Valerie; Zakharko, Taras. Language death predicts conflict risk.
- Bickel, Balthasar [...] Zakharko, Taras et al. Dictionary of Chintang. *Dictionaria*.
- Ivani, Jessica K. & Zakharko, Taras. Nominal plurality in animacy-based systems.

Datasets and databases (selected)

- Graff, Anna; Chousou-Polydouri, Natalia; Inman, David; Skirgård, Hedvig; Lischka, Marc; Zakharko, Taras; Barbieri, Chiara; Bickel Balthasar. Submitted. 2023 . Curating global datasets of structural linguistic features for independence. <https://github.com/annagraff/crossling-curated>
- Bickel, Balthasar; Nichols, Johanna; Zakharko, Taras; Witzlack-Makarevich, Alena; Hildebrandt, Kristine; Rießler, Michael; Bierkandt, Lennart; Zúñiga, Fernando & Lowe, John B. 2022. *The AUTOTYP database*. <https://doi.org/10.5281/zenodo.6255206>
- Ivani, Jessica K. & Zakharko, Taras. 2019. Tymber, Database on nominal number marking constructions [Dataset]. Available on GitHub. <https://github.com/jkivani/tymber>
- Loporcaro, Michele; Paciaroni, Tania; Pescarini, Diego; Idone, Alice; Romagnoli, Serena; Zanini, Chiara; Zakharko, Taras. 2018. *The Zurich Database of Agreement in Italo-Romance (DAI)*. University of Zurich. <https://www.dai.uzh.ch/new/#/public/home>
- Bickel, Balthasar; Taras Zakharko. 2018. Recodings of WALS Online data. <https://github.com/IVS-UZH/WALS-recodings>
- Moran, Steven; Hyzi, Casim; Zakharko, Taras. 2016. TypDB — a collection of typological databases. <https://github.com/IVS-UZH/TypDB>

Software and scripts (selected)

- Zakharko, Taras. CorpusReaders, an R package for importing and manipulating multi-level annotated corpus data. *In development*. <https://github.com/tzakharko/CorpusReaders>
- Zakharko, Taras. pipestrator, an R package for building sustainable data processing pipelines. *In development*. <https://github.com/tzakharko/pipestrator>
- Zakharko, Taras. Exploring scalable matrix extension performance of the Apple M4 chip. <https://github.com/tzakharko/m4-sme-exploration>

Graff, Anna; Lischka, Marc; Zakharko, Taras. densify, an R package for densifying (sparse) matrices. <https://github.com/annagraff/densify>

Zakharko, Taras. Lightweight precondition and postcondition checking (aka assertions) for R. <https://github.com/tzakharko/precondition>

Zakharko, Taras. A Pandoc plugin for typesetting linguistic examples and interlinear glosses. <https://github.com/tzakharko/pandoc-glossify>

Zakharko, Taras. Data aggregations for the AUTOTYP database. <https://github.com/autotyp/autotyp-data>

Zakharko, Taras. git-toolbox, the git extension for integration with Field Linguist's Toolbox. <https://github.com/git-toolbox-integration/git-toolbox>

Zakharko, Taras. Corpus annotation tool for agreement relations in Italo-Romance. <https://github.com/IVS-UZH/DAI-anno-tool>

Zakharko, Taras; Bickel, Balthasar. FamilyBias - software for determining typological biases in language families. <https://github.com/IVS-UZH/familybias>

Zakharko, Taras. ToolboxSearch - a query tool for Toolbox corpora. 2011. <https://github.com/tzakharko/ToolboxSearch>

CONFERENCE PRESENTATIONS AND INVITED TALKS (selected)

(with Jessica K. Ivani and Chundra Cathcart) Nominal plurality in Sino-Tibetan: a diachronic account. *Joint Conference on Language Evolution (JCoLE)*, Kanazawa, 2022.

Mining corpora for form-meaning associations: perspectives for corpus-driven typology. *The 50th Annual Meeting of the Societas Linguistica Europaea*, Zurich, 2017.

Annotating semantics of complementizers in corpora: the case of interpretative backgrounds. *Variation in space and time: clausal complementation in South Slavic*, Zurich, 2016.

(with Balthasar Bickel and Alena Witzlack-Makarevich) Better Data with Late Aggregation: AUTOTYP and beyond. *46th Poznań Linguistic Meeting*, 2016.

(with Steven Moran and Balthasar Bickel) Phonological features from a quantitative typological perspective. *Diversity Linguistics: Retrospect and Prospect*. May 1-3. Leipzig, Germany, 2015.

(with Balthasar Bickel, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich and Frank Seifart) Noun-to-verb ratio and word order. *Diversity Linguistics: Retrospect and Prospect*, Max Planck Institute for Evolutionary Anthropology, May 1-3, 2015.

(with Søren Wichmann, Dan Dediu, Michael J. Dunn, Harald Hammarström, Gerhard Jäger, Damian Blasi and Balthasar Bickel) Simulating language, family, and feature evolution: a review of the state of the art. *Association for Linguistic Typology 10th Biennial Conference (ALT 10)*, Leipzig, August 15-18, 2013.

ToolboxSearch: an R package for working with Toolbox Corpora. *Exploring data from language documentation*, ZAS Berlin, 2013.

Corpus-based typology of clause linkage: a language-independent annotation schema of linguistic predication. *Current Tendencies and Potentials in Language Documentation*, University of Campinas, 2013.

(with Alena Witzlack-Makarevich and Balthasar Bickel) Typological evidence against universal effects of referential scales on agent case marking. *The 45th Annual Meeting of the Societas Linguistica Europaea*, Stockholm, 2013.

(with Steven Moran and Balthasar Bickel) Estimating diachronic preferences of phonological features cross-linguistically. *The 10th biennial Conference of the Association for Linguistic Typology*, Leipzig, 2013.

(with Balthasar Bickel) Synchronic and diachronic biases in language families. *The 45th Annual Meeting of the Societas Linguistica Europaea*, Stockholm, 2012.

(with Lennart Bierkandt, Alena Witzlack-Makarevich, Fernando Zúñiga and Balthasar Bickel) Decomposing hierarchical alignment: co-arguments as conditions on alignment. *Referential hierarchies in morphosyntax: description, typology, diachrony*, University of Zurich, 2012.

(with Balthasar Bickel, Alena Witzlack-Makarevich and Lennart Bierkandt) Semantic Role Clustering: An Empirical Assessments of Semantic Role Types. *Workshop on Role Complexes (New) approaches to defining semantic roles*, Zurich, 2011.