Deutscher Platz 6
04103 Leipzig, Germany
frederic_blum@eva.mpg.de
https://fredericblum.codeberg.page
ORCID
GitHub
Codeberg
Mastodon

# Frederic Blum

*Curriculum Vitae*

## Education

2026 – **Postdoc in *Language Evolution***
*Max Planck Institute for Evolutionary Anthropology*
Quantitative Models for Language Evolution

2022 – 2026 **PhD in *Computational Linguistics* (Awarded 17/07/2026 with *summa cum laude*)**
*Max Planck Institute for Evolutionary Anthropology, University of Passau*
Thesis: *The regularity of sound correspondences in computer-assisted language comparison.*
Main Supervisor: Prof. Dr. Mattis List

2019 – 2022 **MA *Linguistics***
*Humboldt-Universität zu Berlin*, (Final Grade: 1,1)
MA Thesis: *Initial lengthening of consonants in 25 typologically diverse languages*
Supervisors: PD Dr. Frank Seifart, Dr. Susanne Fuchs

2017 – 2018 **Exchange scholarship *Linguistics***
*Pontificia Universidad Católica del Perú (PUCP)*, Lima
Exchange studies for two semesters, supported by scholarships from DAAD and FU Berlin

2015 – 2019 **BA *Language and Society (major) & Spanish philology (minor)***
*Freie Universität Berlin*, Final Grade: 1,2

## Scholarships and Grants

| | | |
|---|---|---|
| 2026 | Max Planck Foundation | *Two weeks writing residence for excellent researchers* |
| 2021 | DAAD training grant | *2x500€ alumni grant for advanced training in data visualization* |
| 2017 | DAAD scholarship | *One-year scholarship to study at PUCP* |

## Five Key Publications

[1] **Frederic Blum** & Johann-Mattis List. 2026. Using correspondence patterns to identify irregular words in cognate sets through leave-one-out validation. In *The Proceedings for the 6th International Workshop on Computational Approaches to Language Change (LChange'26)*, 75–86. Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.lchange-1.6.

[2] **Frederic Blum**. 2026. The over-representation of phonological features in basic vocabulary doesn't replicate when controlling for spatial and phylogenetic effects. *Linguistic Typology* 30(2). 245–269. https://doi.org/10.1515/lingty-2025-0050.

[3] **Frederic Blum**, Carlos Barrientos, Johannes Englisch, Robert Forkel, Simon J. Greenhill, Christoph Rzymski & Johann-Mattis List. 2025. Lexibank 2: pre-computed features for large-scale lexical data. *Open Research Europe* 5(126). 1–27. https://doi.org/10.12688/openreseurope.20216.2.

[4] **Frederic Blum**, Steffen Herbold & Johann-Mattis List. 2025. From isolates to families: Using neural networks for automated language affiliation. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova & Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 17915–17927. Vienna, Austria: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.876/.

[5] **Frederic Blum**, Ludger Paschen, Robert Forkel, Susanne Fuchs & Frank Seifart. 2024. Consonant lengthening marks the beginning of words across a diverse sample of languages. *Nature Human Behaviour* 8. 2127–2138. https://doi.org/10.1038/s41562-024-01988-4.

## Teaching: University Seminars

Winter 2024/2025 **Computer-assisted language comparison**
*University of Passau (full semester)*

Summer 2023 **Language Evolution: Historical linguistics and historical language comparison**
*University of Passau (full semester)*

## Teaching: Workshops and Summer Schools

2025 **Computer-assisted language comparison with Lexibank**
*Chana Winter School, Lima, Peru*

2024 **A standardized workflow for computer-assisted language comparison**
*Data in Historical Linguistics (online seminar series)*

2024 **Computer-assisted historical linguistics**
*Verein Junge Sprachwissenschaft e.V.*

## Invited Talks and Keynotes

[1] **Frederic Blum** & Johann-Mattis List. 2026. *Using correspondence patterns to identify irregular words in cognate sets through leave-one-out validation*. Invited talk presented at *qc-leste: Quantitative Cross-Linguistic Explorations of Space, Time, and Evolution*. Paris.

[2] **Frederic Blum**, Pilar Valenzuela & Roberto Zariquiey. 2025. *Tracing the evolution of Panoan languages in parallel with archaeological changes in the Ucayali Basin*. Talk presented at the workshop *New Advances in Phylolinguistics*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

[3] **Frederic Blum**. 2024. *Using cognate reflex prediction during fieldwork as hypothesis test for genealogical relations between language families*. Invited talk presented at the workshop *Advancing the studies of Peruvian indigenous languages: Psycholinguistic, experimental, descriptive and computational research at the Chana Research Station*. Zürich: University of Zürich.

[4] **Frederic Blum**. 2023. *Una evaluación computacional del hipótesis Pano-Tacana ("A computational Evaluation of the Pano-Tacana hypothesis")*. Invited talk at the *Pontificia Universidad Católica del Perú*. Lima.

## Fieldwork Experience

2025 Shipibo-Konibo (two weeks, Pucallpa, Perú)

2024 Shipibo-Konibo (two weeks, Pucallpa, Perú)

2023 Isconahua (one week, Callería, Perú)

## Mentorship and supervision

2026 - ongoing Mentorship of a Ghanaian PhD student as part of the ARTEMIS programme of the Max Planck Society. The project entails the reciprocal exchange on a joint research project and the transfer of skills needed in academia.

2025 - ongoing Independent supervision of two student assistants for the digitization and annotation of Lexibank datasets.

## Organization of Conferences and Workshops

Workshop at ICHL 2025 Main organizer for a workshop on *Challenges for Computer-Assisted Language Comparison* during the ICHL 27 Conference 2025. [URL]

ETT 2025 Organizing coordinator for the *3rd Emerging Topics in Typology* conference. [URL]

ETT 2024 Part of the organizing team for the *2nd Emerging Topics in Typology* conference. [URL]

StuTS 2020 Part of the organizing team for the *68. Studentische Tagung Sprachwissenschaft*. [URL]

## My Research in the Media

- Passauer Neue Presse: Sprachfamilien im Amazonas

- MDR: Woher wissen wir, wann ein neues Wort anfängt?
- Der Standard: 13 Tausendstelsekunden, die einen wichtigen Unterschied beim Sprechen machen
- Deutschlandfunk Nova: Wissensnachrichten
- Evangelische Zeitung: Längere Konsonanten zeigen Wortanfang an
- Welt: 13-Tausendstel Sekunden machen es möglich, dass wir uns verstehen
- Tagesspiegel: Wie Sprache funktioniert: Das Geheimnis des ersten Konsonanten
- NRC: Een medeklinker aan het begin van een woord klinkt 13 milliseconden langer

## Review Activities

- Proof-Reading: Language Science Press (8)
- Royal Society B (2016: 1)
- Journal of Open Source Software (2026: 1)
- Oceanic Linguistics (2026: 1)
- Scientific Reports (2025: 1, 2026: 1)
- Scientific Data (2025: 1)
- Research Grants Council of Hong Kong (2025: 1)
- Linguistic Typology at the Crossroads (2025: 1)
- Association for Computational Linguistics (ARR 2025: February, ARR 2025 July)
- Humanities and Social Sciences Communications (2025: 1)
- Journal of Open Humanities Data (2024:1, 2026: 1)
- Open Research Europe (2024: 1)
- Language Resources and Evaluation Conference (2024)
- Proof-Reading: Language Science Press (9)

## Peer-reviewed Articles in Journals


[1] **Frederic Blum**. 2026. The over-representation of phonological features in basic vocabulary doesn't replicate when controlling for spatial and phylogenetic effects. *Linguistic Typology* 30(2). 245–269. https://doi.org/10.1515/lingty-2025-0050.

[2] **Frederic Blum**, Carlos Barrientos, Johannes Englisch, Robert Forkel, Simon J. Greenhill, Christoph Rzymski & Johann-Mattis List. 2025. Lexibank 2: pre-computed features for large-scale lexical data. *Open Research Europe* 5(126). 1–27. https://doi.org/10.12688/openreseurope.20216.2.

[3] Roberto Zariquiey, **Frederic Blum** & Pilar Valenzuela. 2026. The forgotten mid vowels: reconstructing the history of the Proto-Pano vowel system. *Diachronica* 43(4). 554–591. https://doi.org/10.1075/dia.25015.zar.

[4] **Frederic Blum**, Ludger Paschen, Robert Forkel, Susanne Fuchs & Frank Seifart. 2024a. Consonant lengthening marks the beginning of words across a diverse sample of languages. *Nature Human Behaviour* 8. 2127–2138. https://doi.org/10.1038/s41562-024-01988-4.

[5] **Frederic Blum**, Carlos Barrientos, Adriano Ingunza & Johann-Mattis List. 2024b. Cognate reflex prediction as hypothesis test for a genealogical relation between the Panoan and Takanan language families. *Scientific Reports* 14(30636). 1–12. https://doi.org/10.1038/s41598-024-82515-3.

[6] **Frederic Blum**, Carlos Barrientos, Roberto Zariquiey & Johann-Mattis List. 2024c. A comparative wordlist for investigating distant relations among languages in Lowland South America. *Scientific Data* 11(92). 1–9. https://doi.org/10.1038/s41597-024-02928-7.

[7] Johann-Mattis List, Nathan W. Hill, **Frederic Blum** & Cristian Juárez. 2024. Grouping sounds into evolving units for the purpose of historical language comparison. *Open Research Europe* 4(31). 1–21. https://doi.org/10.12688/openreseurope.16839.2.

[8] **Frederic Blum**, Carlos Barrientos, Adriano Ingunza & Zoe Poirier. 2023. A phylolinguistic classification of the Quechua language family. *INDIANA - Anthropological Studies on Latin America and the Caribbean* 40(1). 29–54. https://doi.org/10.18441/IND.V40I1.29-54.

[9] Onur Özsoy & **Frederic Blum**. 2023. Exploring individual variation in Turkish heritage speakers' complex linguistic productions: Evidence from discourse markers. *Applied Psycholinguistics* 44(4). 534–564. https://doi.org/10.1017/s0142716423000267.

## Peer-reviewed Conference Proceedings

[1] **Frederic Blum**, Steffen Herbold & Johann-Mattis List. 2025. From isolates to families: Using neural networks for automated language affiliation. In Wanxiang Che, Joyce Nabende, Ekaterina Shutova & Mohammad Taher Pilehvar (eds.), *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 17915–17927. Vienna, Austria: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.876/.

[2] **Frederic Blum**, Johannes Englisch, Alba Hermida-Rodríguez, Rik van Gijn & Johann-Mattis List. 2024. Resource acquisition for understudied languages: Extracting wordlists from dictionaries for computer-assisted language comparison. In *Proceedings of the 3rd Annual Meeting of the Special Interest Group on Under-Resourced Languages @LREC-COLING 2024*, 300–306. Torino: ELRA & ICCL. https://aclanthology.org/2024.sigul-1.36/.

[3] **Frederic Blum** & Johann-Mattis List. 2023. Trimming phonetic alignments improves the inference of sound correspondence patterns from multilingual wordlists. In *Proceedings of the 5th Workshop on Research in Computational Linguistic Typology and Multilingual NLP*, 52–64. Dubrovnik: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.sigtyp-1.6.

[4] Johann-Mattis List, Nathan W. Hill, Robert Forkel & **Frederic Blum**. 2023. Representing and computing uncertainty in phonological reconstruction. In Nina Tahmasebi, Syrielle Montariol, Haim Dubossarsky, Andrey Kutuzov, Simon Hengchen, David Alfter, Francesco Periti & Pierluigi Cassotti (eds.), *Proceedings of the 4th Workshop on Computational Approaches to Historical Language Change*, 22–32. Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.lchange-1.3.

[5] **Frederic Blum**. 2022. Evaluating zero-shot transfers and multilingual models for dependency parsing and pos tagging within the low-resource language family Tupían. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, 1–9. Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-srw.1.

[6] Roberto Zariquiey, Claudia Alvarado, Ximena Echevarría, Luisa Gomez, Rosa Gonzales, Mariana Illescas, Sabina Oporto, **Frederic Blum**, Arturo Oncevay & Javier Vera. 2022. Building an endangered language resource in the classroom: Universal Dependencies for Kakataibo. In *Proceedings of the Language Resources and Evaluation Conference*, 3840–3851. Marseille, France: European Language Resources Association. https://aclanthology.org/2022.lrec-1.409.

## Tutorials Published as Blogposts

[1] **Frederic Blum**. 2026. How to visualize language polygons with QGIS (How to Do X in Linguistics 15). *Computer-Assisted Language Comparison in Practice: Tutorials on Computational Approaches to the History and Diversity of Languages* 9(1). 41–53. https://doi.org/10.15475/CALCIP.2026.1.5.

[2] **Frederic Blum**. 2025. How to run EDICTOR 3 locally. *Computer-Assisted Language Comparison in Practice* 8(1). 1–8. https://doi.org/10.15475/CALCIP.2025.1.1.

[3] **Frederic Blum** & Carlos Barrientos. 2023. A new dataset with phonological reconstructions in CLDF. *Computer-Assisted Language Comparison in Practice* 6(1). 43–51. https://doi.org/10.15475/calcip.2023.1.6.

[4] **Frederic Blum**. 2022. Creating custom commands in CLDF: From Lexibank to Nexus files. *Computer-Assisted Language Comparison in Practice* 5(11). 1–4. https://doi.org/10.58079/m6lq.

[5] **Frederic Blum**. 2021. Data gathering in times of a pandemic: upcycling Constenla Umaña's data on the Chibchan, Lencan and Misumalpam language families. *Computer-Assisted Language Comparison in Practice* 4(5). 1–6. https://doi.org/10.58079/m6l4.

## Talks at Conferences and Workshops

[1] **Frederic Blum** & Rebecca Lethgo. 2026. *Computational resources for a new reconstruction of Proto-Pano-Takana*. Talk presented at the *11th Conference on Indigenous Languages of Latin America*. Austin.

[2] **Frederic Blum**. 2026. *Recursos computacionales para una nueva reconstrucción de Proto-Pano-Takana (Computational resources for a new reconstruction of Proto-Pano-Takana)*. Talk presented at the *21st International Congress of the Association of Latin American Linguistics and Philology*. Lima: Pontificia Universidad Católica del Perú.

[3] **Frederic Blum** & Miri Mertner. 2026. *Temperature does not shape sonority*. Talk presented at the *16th International Conference of the Association for Linguistic Typology*. Lyon.

[4] **Frederic Blum**, Steffen Herbold & Johann-Mattis List. 2025. *From isolates to families: Using neural networks for automated language affiliation*. Talk presented at the *Annual Conference of the Association for Computational Linguistics*. Vienna.

[5] **Frederic Blum**, Pilar Valenzuela & Roberto Zariquiey. 2025. *Tracing the evolution of Panoan languages in parallel with archaeological changes in the Ucayali Basin*. Invited talk presented at the workshop *Uncovering the past with linguistic data: methodological, typological and area-specific studies*. Riga, Latvia.

[6] Roberto Zariquiey, Pilar Valenzuela & **Frederic Blum**. 2024. *Tracing the evolution of Panoan languages in parallel with archaeological changes in the Ucayali Basin*. Invited talk presented at the workshop *Multidisciplinary dialogues on the human past of the Urubamba/Ucayali basin: towards a new synthesis*. Leiden: Leiden University.

[7] **Frederic Blum**. 2024a. *Generalizability and reproduction in quantitative typology: A case study on sound symbolism*. Emerging Topics in Typology. Stockholm: Stockholm University.

[8] **Frederic Blum**. 2024b. *Using lexical and grammatical data to automatically affiliate language isolates and orphans*. Keynote in the workshop on language isolates held at the *57th Annual Meeting of the Societas Linguistica Europaea*. Helsinki.

[9] **Frederic Blum**. 2024c. *Using cognate reflex prediction during fieldwork as hypothesis test for genealogical relations between language families*. Invited talk presented at the workshop *Advancing the studies of Peruvian indigenous languages: Psycholinguistic, experimental, descriptive and computational research at the Chana Research Station*. Zürich: University of Zürich.

[10] **Frederic Blum**. 2024d. *Replicating sound symbolism: Overcoming bias and generalizability challenges in quantitative typology*. Talk presented at the *15th International Conference of the Association for Linguistic Typology*. Singapore: Association for Linguistic Typology.

[11] **Frederic Blum** & Johann-Mattis List. 2023a. *Using Lexibank and Grambank to Automatically Affiliate Languages to Families*. Talk presented at the *Grambank Workshop*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

[12] **Frederic Blum** & Johann-Mattis List. 2023b. *A computational evaluation of regularly recurring sound correspondences*. Talk presented at the *26th International Conference on Historical Linguistics*. Heidelberg: International Society for Historical Linguistics.

[13] **Frederic Blum**. 2023. *Re-examining the proposed genetic relationship(s) of Panoan and Tacanan*. Talk presented at the *56th Annual Meeting of the Societas Linguistica Europaea*. Athens: Societas Linguistica Europea.

[14] **Frederic Blum** & Ludger Paschen. 2022. *Lengthening at Domain Boundaries*. Talk presented at the *Workshop for Spoken- and Signed-Language Corpus Studies in Linguistic Typology* during the *14th Conference of the Association for Linguistic Typology*. Talk.

[15] **Frederic Blum**, Carlos Barrientos Ugarte, Zoe Poirier & Adriano Ingunza. 2021. *Una Aproximación Filoingüística a la Clasificación Interna del Quechua*. Talk presented at the *5.ª Reunión de la Red para el Estudio de las Lenguas Andinas (RELA)*. Talk.