

Nilo Pedrazzini

✉ npedrazzini@turing.ac.uk
☎ +447437476608 🌐 Website & Portfolio
🐙 npedrazzini 🆔 0000-0003-3757-2961
🌐 nilopedrazzini 🐦 @nilo_pedrazzini

Skills

Computer languages	📖 Solid programming skills in Python, R and Bash. Intermediate skills in Ruby.
Operating systems	📖 [In order of preference] macOS, Linux (Ubuntu, ChromeOS), Windows.
Cloud skills	📖 Experience working with Microsoft Azure virtual machines, storage/blobs.
Data handling and NLP	📖 In no particular order, good knowledge of Gensim (e.g. Word2Vec/KeyedVectors and Idamodell), SpaCy, NLTK, Sci-kit Learn, Pandas, NumPy, BeautifulSoup.
Web Development	📖 Solid skills in HTML/CSS, and XML. Introductory skills in JavaScript. Basic experience with SQL and Django for database management.
Version Control	📖 Solid Git knowledge.
Data Visualisation	📖 Extensive experience with several Python libraries, particularly Matplotlib, Seaborn and Plotly (& Plotly Dash for web app deployment). Solid skills in base R plotting and ggplot2.
Languages	📖 Italian (native), English (C2), Russian (C1), Serbian (reading C1, speaking and listening B1), Spanish (Dele B2), German (Goethe-Zertifikat B1), Norwegian (Bokmål) (A2-B1), Korean (A1-A2), te reo Māori (A1). Good reading knowledge of French and Bulgarian. Working knowledge of Ancient Greek, Latin, Hittite, Old Church Slavonic.
Misc.	📖 Software sustainability. Collaborative software development (GitFlow). Testing (Continuous Integration). Corpus building and linguistic annotation. Linguistic Data Analysis. Development of digital and computational tools for historical languages. Teaching (Linguistics and Digital Humanities). Copy-editing. T _E X.

Teaching and research positions

Research

- February 2023– 📖 **Research Software Engineer.** University of Oxford - Within the project *Digital Scholarship @ Oxford* 🔗
- January 2022– 📖 **Research Associate in Corpus-Based Digital Humanities.** The Alan Turing Institute (London) - Within the project *Living with Machines* 🔗👥
- 2020–2021 📖 **Research Assistant.** Oxford School of Global and Area Studies - Largely within the project *Depictions of Post-COVID-19 Futures in Russian International Media: Multimodal Viewpoint Analysis*, International Multimodal Communication Centre (IMCC). 🔗 - I carried out annotation and correlation analyses of speech-gesture co-occurrences in Russian and American media.
- 2020 📖 **Guest Researcher.** Universitetet i Oslo (Norway). ESRC-funded (Overseas Institutional Visit award). - Collaborative project (with Prof. Dag T. T. Haug 👤) on the temporal semantics of non-finite clauses and their competition in Ancient Greek and Early Slavic.

Teaching and research positions (continued)

- **Research Assistant.** Department of Experimental Psychology, University of Oxford. - I collaborated with the ReadOxford research group [🔗](#), whose aim is to answer different questions related to child literacy development. - I developed R scripts to make corpus data reproducible and statistically analysable for morphological complexity and lexical variation. - My major contribution has been scripting an R code to automatically calculate the Average Reduced Frequency (ARF) of combined lemmata/parts of speech in the Oxford Children Corpus [🔗](#) and Childes treebank [🔗](#). Code freely available in my GitHub repository [🔗](#).

Teaching

- July 2022 ■ **Invited lecturer** (Digital Humanities at Oxford Summer School). Introduction to *Word2Vec*.
- Sept 2021–Jan 2022 ■ **Hourly-Paid Lecturer** in Digital Humanities, King's College London (Department of Digital Humanities). Module 7AAVDM14: *The Cultural Web: Building a Humanities Website*. (HTML/CSS, Jekyll, Git)
- April–June 2021 ■ **Tutor** (Oriol College & Lady Margaret Hall, University of Oxford). Introduction to Historical Linguistics.
- Jan–Mar 2021 ■ **Tutor** (Oriol College & Lady Margaret Hall, University of Oxford). Introduction to Semantics & Pragmatics; Introduction to Sociolinguistics.
- Sept–Dec 2020 ■ **Tutor** in Linguistics, Oriol College, University of Oxford. General Linguistics Foundations; Introduction to Psycholinguistics.





Other Employment

- 2018 – now ■ **Undergraduate Admissions Interviewer.** Lady Margaret Hall, University of Oxford. I interview candidates applying for admission to the BAs in Modern Languages & Linguistics and in Psychology, Philosophy & Linguistics.
- 2018 – 2020 ■ **IT/AV Support Assistant.** Somerville College, University of Oxford. I staffed the college IT Helpdesk to assist with software issues. I also set up mics, projectors and other AV equipment at college events.
- 2014 – 2019 ■ **Supervisor & Assistant Manager.** Le Pain Quotidien, London & Oxford Supervisor position since July 2016. Waiter/Barista until then. Acting as Assistant General Manager between December 2016 and January 2018.
- 2015 ■ **Trainee Assistant Curator-Cataloguer.** Slavonic Collections, The British Library. I enhanced the online catalogue of all Slavonic early-printed Cyrillic books held at the BL, in accordance with LC Subject Headings, correcting mistakes, checking items *de visu*, adding information on binding and ownership, doing provenance research where appropriate and disseminating the results in social media. I acquired knowledge of MARC21, LMS (ALEPH), rare-book handling and copy-specific information structure.







Academic Service

- 2022–now ■ **Main organizer** of the Open Humanities Seminar Series (OHSS). [🔗](#)
- 2021–now ■ **Fellow.** Reproducibility Research Oxford (RROx) - UK Reproducibility Network (UKRN) - I am particularly involved in the discussion around best Open Humanities practices.





Academic Service (continued)







- 2020 – now  **Editorial Assistant & Copyeditor.** Journal of Open Humanities Data (JOHD), Ubiquity Press. I am involved in a wide range of tasks, including contacting potential collaborators for research events; creating and moderating satisfaction surveys; providing support with L^AT_EX-written paper templates; conducting data analysis on publication metrics; occasionally serving as Section Editor.
- 2019 – now  **Corpus building & annotation.** Tromsø Old Russian and Old Church Slavonic Treebank (TOROT)  - I am involved in various stages of corpus building and annotation, from text preprocessing to manual correction of automatically-performed morphological analysis, and manual tagging of dependency relations.
- June 2021  **Supervision** (with Dr Barbara McGillivray) of Malithi Alahapperuma (MSc Digital Humanities, University College London), as part of her work placement at the Alan Turing Institute, providing mentorship on the collection and correlation analysis of metrics for the Journal of Open Humanities Data.













Education







- 2019 – now  **DPhil (PhD), University of Oxford** in Linguistics, Philology & Phonetics.  Thesis title: *Early Slavic participle clauses and their competition: a computational approach*. ESRC full scholarship  Supervisor: Dr. Hanne M. Eckhoff.
- 2019  **MSt, University of Oxford** in Slavonic Studies (Distinction). Bosnian-Croatian-Serbian language (Serbian standard); history of Church Slavonic; Cyrillic palaeography.
- 2015 – 2017  **Graduate Program, Università di Pavia** in Linguistics. Project title: *Secondary uses of Russian second-genitive: a corpus-driven quantitative diachronic analysis*. Supervisor: Prof. Silvia Luraghi. I have earned 70% of the credits towards an MA in Theoretical and Applied Linguistics, completing exams in computational linguistics, first-order logic, advanced phonology & phonetics, linguistic typology, Celtic linguistics, Hittology, advanced Slavic philology, advanced Russian language, advanced Russian literature (in Russian). Average grade: 29.2/30 (predicted Distinction).
- 2015  **BA, Università di Pavia** in Modern Languages and Cultures (Russian and English). Thesis title: *La poesia sillabica dei Vecchi Credenti e il Barocco russo: lamenti anonimi della prima metà del XVIII secolo* [Old Believers' syllabic poetry and the Russian Baroque: Anonymous lamentations from the first half of the 18th century]. Supervisor: Prof. Maria C. Bragone. Grade: 110/110 Summa cum Laude.

Further Training

Keys:  Tech & Coding  Linguistics  Languages  Teaching

- 2021  Digital Humanities & Research Software Engineering Summer School (Alan Turing Institute, London [held online]). Main topics: 1. GitFlow 2. Software sustainability 3. Good data visualisation 4. Software development life cycle 5. Historical language modelling; MyBinder. 
- 2020  Mediterranean Machine Learning Summer School (Milan, Italy [held online]). Sessions attended: 1) Reinforcement learning 2) Causal inference 3) Low-resource machine translation 4) Data-efficient optimization with Bayesian optimization. 
-  Digital Humanities at Oxford Summer School (Oxford [held online]). Strand: *From Text to Tech* (Python for NLP). I received a bursary for attending. 

- | | |
|------|--|
| | <p> Transkribus User Conference (Innsbruck, Austria). Topics covered: 1. Training of large HTR models 2. Automatic recognition of tables 3. Transkribus Text2Image tool. </p> <p> Corpora & Discourse International Conference (Sussex University [held online]) Main topic: corpus methods in discourse analysis. </p> <p> Belgrade Winter School in Indo-European Linguistics (Belgrade, Serbia). Courses attended: 1. Phrygian 2. Messapic 3. Gothic. </p> <p> Preparation for Learning and Teaching at Oxford (PLTO).</p> |
| 2019 | <p> Digital Editions course (Taylor Institution Library, Oxford, UK). Topics: 1. Transcription principles 2. XML (TEI) 3. Unicode.</p> |
| 2017 | <p> 15th New York-St. Petersburg Institute of Linguistics, Cognition and Culture (St. Petersburg, Russia). Courses attended: 1. Generative syntax 2. Morphology 3. Phonology 4. Formal semantics.</p> |
| 2015 | <p> Bulgarian language course (SSEES, University College London, UK).</p> |
| 2014 | <p> Russian language course (Echo Eastern Europe School, Kiev, Ukraine).</p> |
| 2013 | <p> Spanish language course (¡Málaga Sí! Language School, Málaga, Spain).</p> |

- | | |
|-----------|--|
| 2022 |  Circle U. Prize Award (PI: Barbara McGillivray), as a contributor to the project <i>The Language of Mechanization</i> , part of the <i>Living with Machines</i> project (The Alan Turing Institute), for innovative and novel approaches to inter- and trans-disciplinary research  |
| 2020 |  Overseas Institutional Visit Award (Economic and Social Research Council), for a two-month visiting research period at the University of Oslo, Norway
 Digital Humanities Summer School (DHOx2020) Bursary |
| 2019-2023 |  Economic and Social Research Council (Grand Union DTP) Full PhD Scholarship |
| 2019 |  Somerville College Archibald Jackson Prize |

Peer-reviewed papers

Updated on Feb 27, 2023

Natural Language Processing for Digital Humanities (NLP4DH) (pp. 85–95). Association for Computational Linguistics (ACL).

- 6 **Pedrazzini**, N. (2020a). Exploiting cross-dialectal gold syntax for low-resource historical languages: Towards a generic parser for pre-modern Slavic. In F. Karsdorp, B. McGillivray, A. Nerges, & M. Wevers (Eds.), *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)* (Vol. 2723, pp. 237–247). Amsterdam, Netherlands: CEUR Workshop Proceedings. Retrieved from <http://ceur-ws.org/Vol-2723/short48.pdf>

Conference Presentations

- 1 *The Language of Mechanisation*. (November 2022). Presenting this Circle-U award-winning project at a round table at the Conference on interdisciplinary and transdisciplinary research for sustainable development (UCLouvain, Louvain-la-neuve, Belgium). Retrieved from <https://uclouvain.be/en/discover/university-transition/conference-sur-la-recherche-interdisciplinaire-et-transdisciplinaire-transition-et-developpement-durable.html>
- 2 *Machines in the media: semantic change in the lexicon of mechanization in 19th-century British newspapers*. (November 2022). The 2nd International Workshop on Natural Language Processing for Digital Humanities – NLP4DH 2022 (AACL-colocated, Taipei, Taiwan), with Barbara McGillivray. Retrieved from <https://rootroo.com/en/nlp4dh-2022/>
- 3 *Evaluating Language Models for Ancient Greek: Design, Challenges, and Future Directions*. (August 2022). International Conference on Historical Linguistics 2022 (ICHL25: Oxford), with Barbara McGillivray, Malvina Nissim, Saskia Peels-Matthey & Silvia Stopponi. Retrieved from <https://ichl.ling-phil.ox.ac.uk/abstracts/168>
- 4 *Historic machines from 'prams' to 'Parliament': new avenues for collaborative linguistic research*. (June 2022). DH Benelux 2022: RE-MIX. Creation and alteration in DH, with Mia Ridge, Giorgia Tolfo, Kalle Westerling and Barbara McGillivray. Retrieved from <https://zenodo.org/record/6578021#.YtjqFy8w1Yh>
- 5 *Le Journal of Open Humanities Data (JOHD): enjeux et défis dans la publication de data papers pour les sciences humaines*. (November 2021). DHNord2021 round table "Le data paper: une nouvelle forme de publication scientifique en SHS", with Paola Marongiu, Marton Ribary & Barbara McGillivray. Retrieved from https://meshs.fr/page/enjeux_et_defis_dans_la_publication_de_data_papers_pour_les_shs
- 6 *Participial free adjuncts in early Slavic treebanks: The discourse properties of the dative absolute*. (September 2020). Slavic Linguistics Society (SLS) Conference, Bloomington (IN, USA) [held as online event]. Retrieved from <https://slavicgf.sitehost.iu.edu/SLS/abstracts/Pedrazzini.PDF>
- 7 *Tackling lack of linguistic data with HTR: A specialized model for the transcription of Serbian Church Slavonic manuscripts*. (April 2021). El' Manuscript Conference 2020. Retrieved from http://www.elmanuscript2020.uni-freiburg.de/wordpress/wp-content/uploads/2021/04/ElManuscript-2021-Booklet-of-Abstracts_09.04.2021.pdf
- 8 *Same question, different annotation depths: early Slavonic dative absolute in deeply versus shallowly annotated treebanks*. (November 2020). Grammar & Corpora International Conference, Kraków (Poland) [held as online event]. Retrieved from https://gac2020.ijp.pan.pl/GaC_2020_paper_25.pdf
- 9 *Exploiting Cross-Dialectal Gold Syntax for Low-Resource Historical Languages: Towards a Generic Parser for Pre-Modern Slavic*. (November 2020). Computational Humanities Research Workshop, Amsterdam (The Netherlands).
- 10 *Challenges in historical corpus analyses of discourse relations: a case study in Old Church Slavonic*. (17-19 June 2020). Corpora & Discourse International Conference, University of Sussex (UK) [held as online event]. Retrieved from <https://corporadiscourse.com/historical-discourse-videos/>

Blog posts

- 1 A reluctantly modern voice from the 17th-century Russian storm: Archpriest Avvakum and the Life written by himself. (10 June 2015). Retrieved from <https://blogs.bl.uk/european/2015/06/a-reluctantly-modern-voice-from-the-17th-century-russian-storm.html>
- 2 Fairytales on trial: the Good and the Beautiful in early-Soviet children's literature. (26th March 2019). Retrieved from <https://blogs.bl.uk/european/2019/03/fairytales-on-trial.html>

Other contributions

- 1 Organizer (with Barbara McGillivray) of: "Open Data", session on Open Humanities Data at the Humanities & Data Science discussion group at The Alan Turing Institute. (25th May 2022). Retrieved from <https://hackmd.io/@turing-hds/DiscussionGroup#Topic-Open-Data>
- 2 *Data processing for*: Dawson, N., Hsiao, Y., Tan, A., Banerji, N., & Nation, K. (2021). Features of lexical richness in children's books: Comparisons with child-directed speech. *Language Development Research*. [doi:10.34842/5we1-yk94](https://doi.org/10.34842/5we1-yk94)