Nilo Pedrazzini

✉ npedrazzini@turing.ac.uk

☎ +447437476608   🌐 Website & Portfolio

🐙 npedrazzini   ⓘD 0000-0003-3757-2961

in nilopedrazzini   🐦 @nilo_pedrazzini

Skills

Computer languages	🔖 Solid Python, R and Bash. Intermediate Ruby.
Operating systems	🔖 [*In order of preference*] MacOS, Linux (Ubuntu, ChromeOS), Windows.
Cloud skills	🔖 Microsoft Azure.
Data handling and NLP	🔖 Extensive experience with Gensim, SpaCy, NLTK, Sci-kit Learn, Pandas, NumPy, BeautifulSoup.
Web Development	🔖 Solid HTML/CSS, XML. Basic JavaScript, SQL and Django (for database management).
Version Control	🔖 Solid Git knowledge.
Data Visualisation	🔖 [*Python*] Extensive experience with Matplotlib, Seaborn, Plotly (& Plotly Dash for web app deployment), Bokeh. [*R*] Base R plotting, 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), Arabic (MSA) (A1), te reo Māori (A1). Good reading knowledge of French and Bulgarian. Working knowledge of Ancient Greek, Latin, Hittite, Old Church Slavonic.
Misc.	🔖 Word embeddings and Large Language Models. Software sustainability. Collaborative software development (GitFlow, Agile/Scrum). 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). Copyediting. TEX.

Academic employment

August 2023–	🔖 **Turing Research Fellow**. The Alan Turing Institute (London) 🔗
Feb 2023–July 2023	🔖 **Research Software Engineer**. University of Oxford (*Digital Scholarship @ Oxford*) 🔗
Jan 2022–July 2023	🔖 **Research Associate in Corpus-Based Digital Humanities**. The Alan Turing Institute (London) (*Living with Machines*) 🔗👤
Sept 2021–Mar 2022	🔖 **Hourly-Paid Lecturer** in Digital Humanities, King's College London (Department of Digital Humanities).
Sept 2020-June 2021	🔖 **Tutor** in General Linguistics (Oriel College & Lady Margaret Hall, University of Oxford).
Dec 2020–Aug 2021	🔖 **Research Assistant**. Oxford School of Global and Area Studies, International Multimodal Communication Centre (IMCC) (*Depictions of Post-COVID-19 Futures in Russian International Media: Multimodal Viewpoint Analysis*) 🔗.
Feb–Jun 2020	🔖 **Research Assistant**. Department of Experimental Psychology, University of Oxford (*ReadOxford*) 🔗

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   **Assistant Manager**. Le Pain Quotidien, London & Oxford. Assistant Manager position since July 2016. Waiter and Barista until then.

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.

Education

2023   **DPhil (PhD), University of Oxford** in Linguistics, Philology & Phonetics. 👤 Thesis: *A quantitative and typological study of Early Slavic participle clauses and their competition.* ESRC full scholarship (Grand Union DTP). Supervisor: 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, Hittitology, 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.

Academic Service

2022–now   **Main organizer & Chair** 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.

2020 – 2023   **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 LaTeX-written paper templates; conducting data analysis on publication metrics; occasionally serving as Section Editor.

2019 – 2023   **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.

Academic Service (continued)

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.

Teaching

Oct-Nov 2023   🔖 *Introduction to Text Mining with Python* (MSc in Digital Scholarship, University of Oxford).

1-7 Jul 2023   🔖 *Applied Data Analysis* (Digital Humanities at Oxford Summer School (DHOxSS)).

Sept-Dec 2021   🔖 *The Cultural Web: Building a Humanities Website.* (HTML/CSS, Jekyll, Git) (King's College London).

April-June 2021   🔖 *Introduction to Historical Linguistics* (University of Oxford).

Jan-Mar 2021   🔖 *Introduction to Semantics & Pragmatics* (University of Oxford).

🔖 *Introduction to Sociolinguistics* (University of Oxford).

Sept-Dec 2020   🔖 *General Linguistics Foundations* (University of Oxford).

🔖 *Introduction to Psycholinguistics* (University of Oxford).

Further Training

Keys: </> Tech & Coding   👉 Lingustics   🔤 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. 🔗

</> Transkribus User Conference (Innsbruck, Austria). Topics covered: 1. Training of large HTR models 2. Automatic recognition of tables 3. Transkribus Text2Image tool. 🔗

👉 Corpora & Discourse International Conference (Sussex University [held online]) Main topic: corpus methods in discourse analysis. 🔗

</> Belgrade Winter School in Indo-European Linguistics (Belgrade, Serbia). Courses attended: 1. Phrygian 2. Messapic 3. Gothic. 🔗

🎓 Preparation for Learning and Teaching at Oxford (PLTO).

2019   </> Digital Editions course (Taylor Institution Library, Oxford, UK). Topics: 1. Transcription principles 2. XML (TEI) 3. Unicode.

2017   👉 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.

2015   🔤 Bulgarian language course (SSEES, University College London, UK).

| 2014 | 🔤 | Russian language course (Echo Eastern Europe School, Kiev, Ukraine). |
| 2013 | 🔤 | Spanish language course (¡Málaga Sí! Language School, Málaga, Spain). |

Awards and grants

2022 🔖 Circle U. Prize Award (PI: Barbara McGillivray), as a contributor to the project *The Language of Mechanization*, part of the *Living with Machines* 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

Publications and Presentations

Peer-reviewed papers

1 Marongiu, Paola, **Nilo Pedrazzini**, Marton Ribary, and Barbara McGillivray (Forthcoming). "Le Journal of Open Humanities Data (JOHD): enjeux et défis dans la publication de data papers pour les sciences humaines". In: *Humanités numériques et science ouverte. Vol III: Publier, partager, réutiliser les données de la recherche: les data papers et leurs enjeux*. Ed. by Joachim Schopfel and Christine Kosmopoulos. Lille: Presses Universitaires du Septentrion.

2 McGillivray, Barbara, Malvina Nissim, **Nilo Pedrazzini**, Saskia Peels-Matthey, and Silvia Stopponi (Forthcoming). "Evaluating Language Models for Ancient Greek: Design, Challenges, and Future Directions". In: *Diachronica* (Special issue "Demystifying new methods in historical linguistics").

3 Haug, Dag and **Nilo Pedrazzini** (2023). "The semantic map of *when* and its typological parallels". In: *Frontiers in Communication* 8. 🔗DOI: 10.3389/fcomm.2023.1163431.

4 Stopponi, Silvia, **Nilo Pedrazzini**, Saskia Peels, Barbara McGillivray, and Malvina Nissim (2023). "Evaluation of Distributional Semantic Models of Ancient Greek: Preliminary Results and a Road Map for Future Work". In: *Proceedings of the Ancient Language Processing Workshop associated with RANLP-2023*, pp. 49–58. 🔗URL: https://aclanthology.org/2023.alp-1.6.

5 McGillivray, Barbara, Paola Marongiu, **Nilo Pedrazzini**, Mandy Wigdorowitz, and Eleonora Zordan (2022). "Deep Impact: A study on the impact of data papers and datasets in the Humanities and Social Sciences". In: *Publications* 10.4 (39). 🔗DOI: 10.3390/publications10040039.

6 **Pedrazzini, Nilo** (2022). "One question, different annotation depths: A case study in Early Slavic". In: *Journal of Historical Syntax* (Special issue: "Annotating Historical Corpora") 6.7, pp. 1–40. 🔗DOI: 10.18148/hs/2022.v6i4-11.96.

7 **Pedrazzini, Nilo** and Barbara McGillivray (2022). "Machines in the media: semantic change in the lexical field of mechanization in 19th-century British newspapers". In: *Proceedings of the 2nd International Workshop on Natural Language Processing for Digital Humanities (NLP4DH)*. Association for Computational Linguistics (ACL), pp. 85–95. 🔗URL: https://aclanthology.org/2022.nlp4dh-1.12.

8 **Pedrazzini, Nilo** and Hanne Martine Eckhoff (2021). "OldSlavNet: A scalable Early Slavic dependency parser trained on modern language data". In: *Software Impacts*, p. 100063. 🔗DOI: 10.1016/j.simpa.2021.100063.

9 **Pedrazzini**, **Nilo** (2020). "Exploiting Cross-Dialectal Gold Syntax for Low-Resource Historical Languages: Towards a Generic Parser for Pre-Modern Slavic". en. In: *Proceedings of the Workshop on Computational Humanities Research (CHR 2020)*. Ed. by Folgert Karsdorp, Barbara McGillivray, Adina Nerghes, and Melvin Wevers. Vol. 2723. Amsterdam, Netherlands: CEUR Workshop Proceedings, pp. 237–247. 🔗URL: http://ceur-ws.org/Vol-2723/short48.pdf.

Chapters in books

1 McGillivray, Barbara, **Nilo Pedrazzini**, Arianna Ciula, Jon Lawrence, Tiffany Ong, Mia Ridge, and Miguel Vieira (Forthcoming). "Analysing the language of mechanisation in 19th-century British newspapers". In: *Living with Machines*. Ed. by Ruth Ahnert, Emma Griffin, Jon Lawrence, and the Living with Machines Team. London: University of London Press.

2 Ridge, Mia, Giorgia Tolfo, Barbara McGillivray, **Nilo Pedrazzini**, Kalle Westerling, and Christina Last (Forthcoming). "Living with Machines: Libraries crowdsourcing in collaboration with data scientists and the public". In: *Libraries, Archives, and the Digital Humanities (Routledge Companions to the Digital Humanities)*. Ed. by Harold Short, Isabel G. Russel, Nirmala Menon, and Ray Siemens. London-New York: Routledge.

Theses

1 Pedrazzini, Nilo (2023). "A quantitative and typological study of Early Slavic participle clauses and their competition". PhD thesis. University of Oxford. 🔗URL: https://ora.ox.ac.uk/objects/uuid:27636bb4-efd0-47bf-8e28-dce5a2b2ca4d.

Conference Presentations

1 "Temporal clauses in Old Church Slavonic and Ancient Greek: typological differences emerged from massively parallel corpora" (Aug. 2023). 56th Annual Meeting of the Societas Linguistica Europaea (SLE, Athens, Greece), joint work with Dag Haug. 🔗URL: https://societaslinguistica.eu/sle2023/.

2 "Metadata Enrichment in the Living with Machines Project: User-focused Collaborative Database Development in a Digital Humanities Context" (July 2023). ADHO Digital Humanities Conference (DH2023, Graz, Austria), joint work with Kalle Westerling, David Beavan, Kaspar Beelen, Mariona Coll Ardanuy, Timothy Hobson, Christina Last, Griffith Reese, and Like Hare. 🔗URL: https://dh2023.adho.org.

3 "Diachronic and diatopic word embeddings from British historical newspapers" (June 2023). 12th AIUCD (Convegno dell'Associazione per l'Informatica Umanistica e la Cultura Digitale), joint work with Barbara McGillivray. 🔗URL: http://www.aiucd2023.unisi.it.

4 "The Language of Mechanisation" (Nov. 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). 🔗URL: https://uclouvain.be/en/discover/university-transition/conference-sur-la-recherche-interdisciplinaire-et-transdisciplinaire-transition-et-developpement-durable.html.

5 "Machines in the media: semantic change in the lexicon of mechanization in 19th-century British newspapers" (Nov. 2022). The 2nd International Workshop on Natural Language Processing for Digital Humanities – NLP4DH 2022 (AACL-colocated, Taipei, Taiwan), with Barbara McGillivray. 🔗URL: https://rootroo.com/en/nlp4dh-2022/.

6 "Evaluating Language Models for Ancient Greek: Design, Challenges, and Future Directions" (Aug. 2022). International Conference on Historical Linguistics 2022 (ICHL25: Oxford), with Barbara McGillivray, Malvina Nissim, Saskia Peels-Matthey & Silvia Stopponi. 🔗URL: https://ichl.ling-phil.ox.ac.uk/abstracts/168.

7 "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. 🔗URL: `https://zenodo.org/record/6578021#.YtjqFy8w1Yh`.

8 "Deep impact: A study on the impact of data papers and datasets in the humanities and social sciences" (June 2022). SciDataCon (International Data Week, Seoul, 20-23 June 2022), with Marton Ribary, Mandy Wigdorowitz, Eleonora Zordan, Barbara McGillivray. 🔗URL: `https://www.scidatacon.org/IDW-2022/sessions/400/paper/975/#`.

9 "Le Journal of Open Humanities Data (JOHD): enjeux et défis dans la publication de data papers pour les sciences humaines" (Nov. 2021). DHNord2021 round table "Le data paper: une nouvelle forme de publication scientifique en SHS", with Paola Marongiu, Marton Ribary & Barbara McGillivray. 🔗URL: `https://meshs.fr/page/enjeux_et_defis_dans_la_publication_de_data_papers_pour_les_shs`.

10 "Tackling lack of linguistic data with HTR: A specialized model for the transcription of Serbian Church Slavonic manuscripts" (Apr. 2021). El' Manuscript Conference 2020. 🔗URL: `http://www.elmanuscript2020.uni-freiburg.de/wordpress/wp-content/uploads/2021/04/ElManuscript-2021-Booklet-of-Abstracts_09.04.2021.pdf`.

11 "Same question, different annotation depths: early Slavonic dative absolute in deeply versus shallowly annotated treebanks" (Nov. 2020). Grammar & Corpora International Conference, Kraków (Poland) [held as online event]. 🔗URL: `https://gac2020.ijp.pan.pl/GaC_2020_paper_25.pdf`.

12 "Exploiting Cross-Dialectal Gold Syntax for Low-Resource Historical Languages: Towards a Generic Parser for Pre-Modern Slavic" (Nov. 2020). Computational Humanities Research Workshop, Amsterdam (The Netherlands).

13 "Participial free adjuncts in early Slavic treebanks: The discourse properties of the dative absolute" (Sept. 2020). Slavic Linguistics Society (SLS) Conference, Bloomington (IN, USA) [held as online event]. 🔗URL: `https://slavicgf.sitehost.iu.edu/SLS/abstracts/Pedrazzini.PDF`.

14 "Challenges in historical corpus analyses of discourse relations: a case study in Old Church Slavonic" (June 2020). Corpora & Discourse International Conference, University of Sussex (UK) [held as online event].

Invited lectures and talks

1 "T*xt as data: A sneak peek into the possibilities unleashed by computational methods" (15 May 2023). Creative & Heritage Innovation Challenge Workshop (Oxford, United Kingdom).

2 "Text mining per le Humanities: un'introduzione pratica con Python" (22 March 2023). Hands-on introduction to text mining with Python for doctoral students at the University of Pavia (Italy). 🔗URL: `https://github.com/npedrazzini/pavia-text-mining`.

3 "An introduction to Word2Vec" (14 July 2022). Digital Humanities at Oxford Summer School ('Text to Tech' strand), Oxford (United Kingdom). 🔗URL: `https://github.com/Living-with-machines/dhoxss-text2tech`.

Blog posts

1 *A reluctantly modern voice from the 17th-century Russian storm: Archpriest Avvakum and the Life written by himself* (10 June 2015). (British Library, European Studies Blog, 10th June 2015). 🔗URL: `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). (British Library, European Studies Blog, 26th March 2019). 🔗URL: `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). ⊘URL: `https://hackmd.io/@turing-hds/DiscussionGroup#Topic-Open-Data`.

2 *Data processing for*: Dawson, Nicola, Yaling Hsiao, Alvin Tan, Nilanjana Banerji, and Kate Nation (2021). "Features of lexical richness in children's books: Comparisons with child-directed speech". In: *Language Development Research*. ⊘DOI: `10.34842/5we1-yk94`.