Nilo Pedrazzini

✉ nilo.pedrazzini@ling-phil.ox.ac.uk
📞 +447437476608   🌐 npedrazzini.github.io
🔗 npedrazzini   ⓘD 0000-0003-3757-2961
in nilopedrazzini   🐦 @nilo_pedrazzini

Skills

Computer languages	🔖 Solid skills in Python, R, XML, Bash, HTML/CSS and TEX. Intermediate skills in Ruby, JavaScript and SQL.
Data handling and NLP	🔖 Experience with Pandas, NumPy, BeautifulSoup, SpaCy, NLTK, CLTK.
Data Visualisation	🔖 Experience with Plotly (& Plotly Dash for web app deployment) and ggplot2. Good skills in Seaborn and Matplotlib.
Languages	🔖 Italian (native), English (C2), Russian (C1), Serbian (reading C1, speaking and listening B2), Spanish (Dele B2), German (Goethe-Zertifikat B1), Norwegian (Bokmål) (A2-B1). Good reading knowledge of French and Bulgarian. Working knowledge of Ancient Greek, Latin, Hittite, Old Church Slavonic.
Misc.	🔖 Corpus building and linguistic annotation. Linguistic Data Analysis. Development of digital and computational tools for historical languages. Teaching (Linguistics and Digital Humanities). Copyediting.

Teaching and research positions

Research

January 2022–	🔖 **Research Assistant in Corpus-Based Digital Humanities**. 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.
	🔖 **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

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** (Oriel College & Lady Margaret Hall, University of Oxford). Introduction to Historical Linguistics.

Teaching and research positions (continued)

Jan-Mar 2021   **Tutor** (Oriel College & Lady Margaret Hall, University of Oxford). Introduction to Semantics & Pragmatics; Introduction to Sociolinguistics.

Sept-Dec 2020   **Tutor** in Linguistics, Oriel College, University of Oxford. General Linguistics Foundations; Introduction to Psycholinguistics.

Other Employment

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.

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.

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.

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

2019 – now   **Corpus builder & annotator**. 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 of 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.

Education (continued)

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.

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).

Publications and Presentations

Journal Articles

1 **Pedrazzini**, N. (Forth.). One question, different annotation depths: A case study in Early Slavic. *Journal of Historical Syntax* (Special issue: "Annotating Historical Corpora").

2 **Pedrazzini**, N., & Eckhoff, H. M. (2021a). OldSlavNet: A scalable Early Slavic dependency parser trained on modern language data. *Software Impacts*, 100063. 🔗doi:10.1016/j.simpa.2021.100063

Conference Proceedings

1 **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. Nerghes, & 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

2 Marongiu, P., **Pedrazzini**, N., Ribary, M., & McGillivray, B. (Forthcoming). Le Journal of Open Humanities Data (JOHD): enjeux et défis dans la publication de data papers pour les sciences humaines, [To be published by Presses Universitaires du Septentrion].

Conference Presentations

1 *Le Journal of Open Humanities Data (JOHD): enjeux et défis dans la publication de data papers pour les sciences humaines.* (November 2020). To be presented at 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

2 *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

3 *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

4 *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

5 *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).

6 *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/

Open software and datasets

1 **Pedrazzini**, N., & Eckhoff, H. M. (2021b). Reproducible Capsule for: 'OldSlavNet: A scalable Early Slavic dependency parser trained on modern language data'. *Code Ocean* (Software). 🔗doi:10.24433/CO.3462707.V1

2 **Pedrazzini**, N. (2020b). Data from 'Dative absolutes in discourse: the value of deeply versus strategically annotated treebanks'. *Figshare* (Dataset). 🔗doi:10.6084/m9.figshare.12894035

3 **Pedrazzini**, N. (2020c). Data from 'Exploiting cross-dialectal gold syntax for low-resource historical languages: towards a generic parser for pre-modern Slavic'. *Figshare* (Dataset).
🔗doi:10.6084/m9.figshare.12950093

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 *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