Roberta Rocca, PhD

Short bio

Hi there! I am a researcher working at the intersection between cognitive science, NLP, and informatics, with experience in both European and US research institutions and publications in several scientific journals. My current research focuses on using machine learning models to study of human language, cognition, and social interactions, on developing open-source software, and on clinical applications of NLP. I thrive in teams with ambitious interdisciplinary agendas and when working on projects with a focus on real-world impact.

rbrrcc@gmail.com
rbroc
RockBerta
rbroc.github.io

Skills

Programming languages: Python; SQL; R; Matlab; bash; PERL
Main libraries, frameworks and tools: TF/Keras; PyTorch; Scikit-learn; numpy/scipy; pandas; Git; Docker
Areas of Expertise: NLP; Statistical Modeling; fMRI; Open Science; Experimental Methods
Languages: Italian (native); English (fluent); Danish (fluent); French (intermediate)

Experience

Tenure-track Assistant Professor
December 2022 – ongoing
School of Culture and Society, Aarhus University, Denmark
Focus areas: Clinical NLP, Cognitive Science and Neuroscience, Computational Social Science

Postdoctoral Researcher
December 2020 – November 2022
School of Culture and Society, Aarhus University, Denmark
Focus areas: Natural Language Processing, Machine Learning for Health, Computational Social Science
Key achievements
- Engineered and evaluated NLP models for language-based inference of psychiatric disorders
- Created and analyzed multilingual datasets of social media text through APIs and ML pipelines
- Developed methodologies for large-scale semantic modeling of multilingual text data
- Coordinated online survey-based data collection for an international social science consortium

Note: I was on leave between 06/2021 – 07/2021 and 09/2021 – 03/2022

Postdoctoral Researcher
December 2019 – March 2022
Psychoinformatics Lab, University of Texas at Austin, USA
Focus areas: Neuroinformatics, Deep Learning, Natural Language Processing
Key achievements
- Co-developed Neuroscout, an open-source Python platform for end-to-end analysis of brain imaging data
  - Developed and maintained the feature extraction library pliers
  - Authored Neuroscout's release paper and its open code repository
  - Performed QA-testing and contributed to Neuroscout's documentation
- Developed transformer models for text-based author encoding
- Conducted research on model evaluation in machine learning and cognitive science
- Published research outputs in peer-reviewed scientific journals (see publications)

Note: I was on leave from this role between 12/2020 and 09/2021

Predictive Analytics Data Fellow
June – August 2021
Centre for Humanitarian Data, United Nations
Focus area: Predictive Modeling, Complex Systems; Data-Driven Policy-Making
Key achievements
- Identified methods, data requirements, and partnerships for pilots on data-driven cholera response
- Provided strategic recommendations on introducing causal modeling into humanitarian needs assessment
- Disseminated findings through a technical report, a blog post, conference talks, and webinars

Education

PhD, Cognitive Neuroscience
September 2016 – October 2019
Aarhus University, Denmark

MSc, Cognitive Science (Focus: Computational Linguistics)
September 2014 – July 2016
University of Trento, Italy

Additional research experience

Visiting Researcher, Department of Applied Mathematics and Computer Science; Technical University of Denmark, 2018-19
Visiting Researcher, Institute of Cognitive Science and Technologies; National Research Council, Italy, 2018
Research Assistant, Department of Experimental Psychology, University College London, United Kingdom, 2016
Research Assistant, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, 2015-16

Publications

Preprints

- Hansen, L., **Rocca**, **R.**, Simonsen, A., Parola, A., Bliksted, V., Ladegaard, N., Bang, D., Tylén, K., Weed, E., Østergaard, S. D., Fusaroli, R. (2023), Automated speech- and text-based classification of neuropsychiatric conditions in a multidiagnostic setting, https://doi.org/10.48550/arXiv.2301.06916

- Fusaroli, R., Weed, E., **Rocca**, **R.**, Fein, D., Naigles, L. (2023), Repeat after me? Both children with and without autism commonly align their language with that of their caregivers, https://doi.org/10.31234/osf.io/m8fhk

- Palaniyappan, L., Benrimoh, D., Voppel, A., **Rocca**, **R.** (2023), Studying psychosis using Natural Language Generation: A review of emerging opportunities, https://doi.org/10.31234/osf.io/rdn3k

Peer-reviewed publications

- **Rocca. R.**, Tamagnone, N., Contla, X., Bove, J.B., Rekabsaz, N. (accepted), Natural language processing for humanitarian action: challenges, opportunities, and the path towards a humanitarian NLP community, *Frontiers in Big Data*

- Fusaroli, R., Weed, E., **Rocca**, **R.**, Fein, D., Naigles, L. (2023), Caregiver linguistic alignment to autistic and typically developing children: A natural language processing approach illuminates the interactive components of language development, *Cognition*

- **Rocca**, **R.**, Yarkoni, T. (2022), Language as a fingerprint: A self-supervised approach to text-based user modeling using transformers, to appear in *Findings of the Association for Computational Linguistics: EMNLP 2022*

- **Rocca**, **R.**, de la Vega, A. (2022), Evaluating the role of non-lexical markers in LLMs' language modeling behavior, to appear in *Proceedings of the 4th Workshop on Evaluation and Comparison of NLP Systems*

- de la Vega, A.*, **Rocca**, **R.* (co-first)**, Blair, R., Mentch, J., Markiewicz, C., Ghosh, S., Poldrack, R., Yarkoni, T. (2022), Neuroscout: a unified platform for generalized and reproducible fMRI research, *eLife*

- **Rocca R.**, Tylén, K. (2022), Cognitive diversity promotes collective creativity: an agent-based simulation, *Proceedings of the 44th Annual Meeting of the Cognitive Science Society*

- Todisco, E., **Rocca**, **R.**, Wallentin, M. (2022). Aqueix caught in the middle. A Demonstrative Choice Task Study of Catalan Demonstratives. *Probus*

- **Rocca**, **R.**, Yarkoni, T. (2021). Putting psychology to the test: rethinking model evaluation through benchmarking and prediction, *Advances in Methods and Practices in Psychological Science*

- Todisco, E., **Rocca**, **R.**, Wallentin, M. (2021). The semantics of spatial demonstratives in Spanish: a demonstrative choice task study. *Language and Cognition*

- **Rocca**, **R.**, Coventry, K. R., Tylén, K., Staib, M., Lund, T. E., & Wallentin, M. (2020). Language beyond the language system: dorsal visuospatial pathways support processing of demonstratives and spatial language during naturalistic fast fMRI, *NeuroImage*

- **Rocca**, **R.**, Wallentin, M. (2020). Demonstrative reference and semantic space: a large-scale demonstrative choice task (DCT) study. *Frontiers in Psychology*

- Wallentin, M., **Rocca**, **R.** (2020). The semantics of spatial demonstratives. *Proceedings of the 42nd Annual Meeting of the Cognitive Science Society*

- **Rocca**, **R.**, Wallentin, M., Vesper, C., & Tylén, K. (2019). This is for you: social modulations of proximal vs. distal space in collaborative interaction, *Scientific Reports*

- **Rocca**, **R.**, Tylén, K., Wallentin, M. (2019), *This* shoe, *that* tiger: Semantic properties reflecting manual affordances of the referent modulate demonstratives use, *PlosOne*

- Wallentin, M., **Rocca**, **R.**, Stoustrup S. (2019), Grammar, gender and demonstratives in lateralized imagery for sentences, *Journal of Psycholinguistic Research*

- Wallentin, M., **Rocca**, **R.**, Stoustrup, S. (2018), Lateralized imagery for sentence content: Testing grammar, gender and demonstratives, *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*

- **Rocca**, **R.**, Wallentin, M., Vesper, C. & Tylén, K. (2018). *This* and *that* back in context: grounding demonstrative reference in manual and social affordances, *Proceedings of the 40th Annual Meeting of the Cognitive Science Society*

- Carapezza, M., & **Rocca**, **R.** (2017). In-seguire la Regola: Giochi Linguistici e Arti Performative. *Rivista Italiana di Filosofia del Linguaggio*, *11*(2) (in Italian)

- **Rocca**, **R.** & Augustin, M. (2016), Gesturing in L2: Evidence for cross-linguistic transfer in the visual modality, in *TwistX Proceedings of the 10th Linguistics Student Conference*

- Carapezza, M. & **Rocca, R.** (2014), Ceci n'est pas une ontologie: a contribution to a quasi-resolute reading of the *Tractatus Logico-Philosophicus*, in Rinhofner-Kreidl S. & Wiltsche H. (eds), *Analytical and Continental Philosophy: Methods and Perspectives. Paper of the 37th Wittgenstein Symposium*

Technical reports

- **Rocca, R.**, (2021) Complex Systems Modeling for Humanitarian Action: Methods and Opportunities, *Research Report for the United Nations' Centre for Humanitarian Data*, pdf available here

Selected articles in newspapers and magazines (in Italian)

- Algoritmi di classe, *Doppiozero*, available here
- Chi ha paura dei data scientists? Numeri e pandemia, *Doppiozero*, available here
- Dati, miti, stati, *L'identità di Clio*, available here

Open-source contributions

Since 2019, I have been part of team developing Neuroscout: https://neuroscout.org, a fully open-source unified platform for the analysis of naturalistic neuroimaging data. An overview of the platform and its applications can be found in our recent *eLife* paper: https://elifesciences.org/articles/79277.

I am one of the main contributors and maintainers of the open-source feature extraction library *pliers* https://github.com/PsychoinformaticsLab/pliers, which is used by Neuroscout to extract visual, auditory and linguistic features from naturalistic fMRI stimuli.

I am co-first author of the Neuroscout release paper mentioned above, and I contributed to designing and executing the meta-analytic validation studies presented there. We shared all the underlying code in this GitHub repository: https://github.com/neuroscout/neuroscout-paper, which is also available as a fully executable Jupyter book: https://neuroscout.github.io/neuroscout-paper/intro.html.

Code for analyses reported on my publications is shared on GitHub. As an example, this repository: https://github.com/rbroc/benchmarks_paper contains the code for Rocca & Yarkoni (2022), "Putting psychology to the test: rethinking model evaluation through benchmarking and prediction", published on *Advances in Methods and Practices for Psychological Science*.

You can check my GitHub profile for more examples of open-source code related to ongoing projects. For data protection reasons, code and data for projects using clinical data is stored in private repositories.

Presentations, teaching and outreach

I have recently been an invited participant at the workshop on "Crosslinguistic speech patterns: biosocial markers of psychiatric disorders", which has taken place at the Lorentz Centre in Leiden, NL, between 31/10/2022 and 5/11/2022 (https://www.lorentzcenter.nl/crosslinguistic-speech-patterns-biosocial-markers-of-psychiatric-disorders.html).

I have taught courses in Applied Cognitive Science (BSc in Cognitive Science); Cognitive Science (MA in Cognitive Semiotics); Cognitive Neuroscience (BSc and MSC in Cognitive Science); Experimental Methods (BSc in Cognitive Science), as well as workshops and guest lectures in R programming (Staff course at Aarhus University); Social and Cultural Dynamics (BSc in Cognitive Science).

I have presented my work at (among other venues): the Annual Meeting of the Organization for Human Brain Mapping (Rome, 2019; Glasgow, 2022); the Annual Conference of the Society for the Neurobiology of Language (Québec City, 2018; Helsinki, 2019); the Annual Meeting Cognitive Science Society (Madison – WI, 2018; Online, 2020; Toronto/Hybrid, 2022), the Conference of the Society for Complex Systems (Lyon/Hybrid, 2021). I have been invited to present my work at seminar series held by the Speech and Hearing Group and MIT/Harvard; McGill University, Karolinska Institute and University of Toronto, and at the United Nations' Centre for Humanitarian Data.

In 2019, I organized a workshop on Natural Language Processing at Aarhus University, bringing together academic experts in the field from several countries. I have also co-organized seminar series, conferences (SALC7) and other scientific events (Workshop: "From fieldwork to modelling: Explaining the variability of linguistic spatial referencing systems" at ICSC 2018).

Other interests

When I am not doing science, I play bass in a band, write fiction, and watch art-house movies. I enjoy learning new languages, and I am counting on expanding my current repertoire. I spend most of my holiday time going on high-stamina road trips and long hikes.