

# Roberta Rocca, PhD

## Short bio

Hi there! I am a researcher working at the intersection between cognitive science, NLP, and data science, 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 language, cognition, and social interactions.

I love contributing to open-source projects, and I thrive in impact-driven teams with ambitious interdisciplinary agendas.

✉ [rbrccc@gmail.com](mailto:rbrccc@gmail.com)  
🌐 [rbroc](https://rbroc.github.io)  
🌐 [rbroc.github.io](https://rbroc.github.io)

---

## Skills

**Programming languages:** Python; SQL; R; Matlab; bash; PERL

**Main tools and libraries:** TF; PyTorch; HuggingFace; sklearn; numpy/scipy; pandas; Git; Docker

**Areas of Expertise:** Machine Learning; NLP; Data Science; Experimental Methods

**Languages:** Italian (native); English (fluent); Danish (fluent); French (intermediate)

---

## Experience

### Tenure-track Assistant Professor

December 2022 - ongoing

**Department of Culture, Cognition, and Computation, Aarhus University, Denmark**

**Focus areas:** Natural Language Processing; Machine Learning; Cognitive Science

### 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](#)
  - QA testing and [documentation](#)
- Developed language models for text-based user encoding
- Engineered and evaluated NLP models for language-based inference of psychiatric disorders
- Conducted research on model evaluation in machine learning and cognitive science
- Published research outputs in peer-reviewed scientific journals (see publications)

### 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 partners for pilot on data-driven cholera response
- Provided strategic recommendations on causal modeling for humanitarian needs assessment
- Disseminated findings through a [technical report](#), a blog post, conference talks, and webinars

### Postdoctoral Researcher (project-based collaboration)

December 2020 - November 2022

**Interacting Minds Centre, Aarhus University, Denmark**

**Focus areas:** Natural Language Processing, Computational Social Science

**Key achievements**

- Collected and analyzed large-scale Twitter datasets using multilingual language models
  - Developed methodologies for large-scale semantic modeling of multilingual text data
  - Coordinated online data collection for international social science consortium
- 

## Education

**PhD, Cognitive Neuroscience**

September 2016 - October 2019

Aarhus University, Denmark

**MSc, Cognitive Science (Track: 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

### Peer-reviewed publications

- Hansen, L., **Rocca, R.**, Simonsen, A., Olsen, L., Parola, A., Bliksted, V., ... & Fusaroli, R. (2023). Speech-and text-based classification of neuropsychiatric conditions in a multidagnostic setting. *Nature Mental Health*, 1-11.
- Coventry, K. R., Gudde, H. B., Diessel, H., Collier, J., Guijarro-Fuentes, P., Vulchanova, M., ..., **Rocca, R.**, ... & Incel, O. D. (2023). Spatial communication systems across languages reflect universal action constraints. *Nature Human Behaviour*, 1-12.
- 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, *Cognitive Science*
- **Rocca, R.**, Tamagnone, N., Contla, X., Bove, J.B., Rekabsaz, N. (2023), 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*
- Palaniyappan, L., Benrimoh, D., Voppel, A., **Rocca, R.** (2023). Studying psychosis using Natural Language Generation: A review of emerging opportunities. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*.
- Corona Hernández, H., Corcoran, C., Achim, A. M., de Boer, J. N., Boerma, T., Brederoo, S. G., ..., **Rocca, R.**, ... & Palaniyappan, L. (2023). Natural language processing markers for psychosis and other psychiatric disorders: emerging themes and research agenda from a cross-linguistic workshop. *Schizophrenia bulletin*, 49(Supplement\_2), S86-S92.
- **Rocca, R.**, Yarkoni, T. (2022), Language as a fingerprint: A self-supervised approach to text-based user modeling using transformers, *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, *Proceedings of the 3<sup>rd</sup> Workshop on Evaluation and Comparison of NLP Systems (Eval4NLP) @ ACL-IJCNLP*
- 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 44<sup>th</sup> 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. & Wiltzsche H. (eds), *Analytical and Continental Philosophy: Methods and Perspectives. Paper of the 37th Wittgenstein Symposium*

## Preprints

- Kruse, L., Wallentin, M., **Rocca, R.** (2023). Inferring depression and its semantic underpinnings from simple lexical choices., PsyArXiv: <https://osf.io/preprints/psyarxiv/5xsjw/>
- **Rocca, R.**, Lawall, K., Tsakiris, M., & Cram, L. (2023). From economy to identity: A data-driven analysis of the European Commission's Twitter communication between 2010 and 2022. SocArXiv: <https://osf.io/preprints/socarxiv/a7udw/>

## 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 other languages (including journals, newspapers, and magazines)

- Conversazioni umane e artificiali, available [here](#) (with Marco Carapezza)
- 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, 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 related to my publications is shared on GitHub: <https://github.com/rbroc>.

## Teaching

I am currently teaching the following courses (10 ECTS each):

- **Natural Language Processing** (MSc in Cognitive Science, Aarhus University)
- **Data Science** (MSc in Cognitive Science, Aarhus University)
- **Applied Cognitive Science** (BSc in Cognitive Science, Aarhus University)

I have previously taught courses in 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).

## Invited talks, conferences and outreach

I have recently given invited talks at:

- Harvard-MIT Speech and Language Biomarkers Interest Group (2022)

- Humanitarian Networks and Partnerships Week (HNPW) in Geneva (2023)
- Odense workshop of the Nordic Network for the Science of Science (2023)
- Seminar series organized by Karolinska Institute, McGill University, University of Toronto (2021)

I have been an invited participant at two Lorentz workshops ("Crosslinguistic speech patterns: biosocial markers of psychiatric disorders": <https://www.lorentzcenter.nl/crosslinguistic-speech-patterns-biosocial-markers-of-psychiatric-disorders.html> and "Cognitive modeling of complex behavior": <https://www.lorentzcenter.nl/cognitive-modeling-of-complex-behavior.html>).

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 piano and bass, write fiction, and watch art-house movies. I enjoy learning (natural and programming) languages, and I am counting on expanding my current repertoire. I am also into high-stamina road trips and long hikes.