

Jesús Rabanal Álvarez

NLP & Data Analysis | Computational Linguistics | Applied Machine Learning

Currently based in Copenhagen; relocating to Belgium
jesusrabanal@hotmail.com | My Portfolio | GitHub: RabanalGsus | LinkedIn: Jesús Rabanal Álvarez

Profile

MSc student in IT & Cognition at the University of Copenhagen with a background in linguistics, data analysis, programming, and NLP. My work focuses on using computational and quantitative methods to study language data, build reproducible analysis workflows, and evaluate language-processing systems. I have experience with Python, R, statistical analysis, machine learning workflows, corpus processing, experimental data, annotation workflows, and Linux-based research environments.

I am particularly interested in contributing to applied NLP and interdisciplinary research projects involving data preprocessing, annotation workflows, corpus analysis, model evaluation, research software support, and computational approaches to language. My academic work has focused on anaphora resolution and coreference-related processing, while my technical experience includes practical NLP pipeline development for transcription and subtitle generation.

Technical & Research Skills

Programming & analysis	Python, R, pandas, NumPy, scikit-learn, statistical analysis, data preprocessing, data visualization
NLP & machine learning	NLP pipelines, corpus processing, text classification, model evaluation, Hugging Face/Transformers, Sentence Transformers, Whisper, faster-whisper
Research methods	Experimental design, psycholinguistic experiments, visual-world eye-tracking, fixation data preprocessing, annotation workflows, quantitative analysis, interpretation of experimental results
Systems & workflow	Linux, Git, reproducible workflows, dependency management, documentation, server/lab technical support
Languages	Spanish: native; English: fluent/professional; German: basic/intermediate; French: intermediate; Danish: basic; Dutch/Flemish: beginner, currently learning

Selected Research & NLP Projects

MSc Thesis – Anaphora Resolution and Coreference-Related Processing University of Copenhagen, 2026
Investigated pronominal anaphora resolution using visual-world eye-tracking data across Spanish, Danish, and Danish learners of Spanish. The project involved experimental design, stimulus preparation, fixation data preprocessing, statistical analysis, visualization, and interpretation of cognitive processing patterns. The work connects linguistic theory, psycholinguistic methods, and data-driven analysis of language processing.

NLP Pipeline Development – Transcription and Subtitle Generation NLPGo, 2024
Worked on practical language-processing pipelines for transcription and subtitle generation using Whisper and faster-whisper. Focused on building maintainable end-to-end workflows in Linux environments, with attention to reproducibility, dependency management, testing, and documentation.

Political Speech Classification and Corpus Analysis Academic / Personal Project
Processed Spanish parliamentary speech data and built text classification pipelines using traditional and embedding-based NLP methods. Worked on corpus extraction, preprocessing, labeling, feature generation, model training, evaluation, and error analysis. The project explored how computational methods can be used to study political language and speaker/party-related patterns in large-scale text data.

SCAN Generalization Experiments with Transformer-Based Models Academic Project
Implemented and evaluated sequence-to-sequence learning experiments inspired by research on compositional generalization. Worked with custom PyTorch data loading, tokenization, training loops, evaluation metrics, and model comparison. The project involved both custom Transformer models and fine-tuning approaches using T5-small.

Experience

DataAnnotation

2026 – Present

Freelance Data Annotator, Remote / Denmark

- Work remotely on data annotation and evaluation tasks involving language, reasoning, and AI-generated outputs.
- Apply careful judgement to assess data quality, consistency, and task-specific criteria.
- Contribute to human-in-the-loop workflows used to evaluate and improve AI systems.
- Work independently with attention to detail, clear guidelines, and reliable task completion.

NLPGo

September 2024 – December 2024

Junior LLM Developer Intern, Remote / Spain

- Developed and maintained NLP pipelines for transcription and subtitle generation using Whisper and faster-whisper.
- Worked in Linux-based environments with attention to reproducibility, dependency management, and clear documentation.
- Built practical workflows for processing audio and language data, focusing on reliability and maintainability.
- Took a hands-on approach to debugging, testing, and improving language-processing tools.

Bilingualab, University of Granada

September 2022 – June 2023

Research Assistant, Granada, Spain

- Supported PhD researchers with the technical setup of psycholinguistic experiments.
- Assisted with experiment programming, lab operations, and research infrastructure.
- Maintained Linux-based lab/server environments under the supervision of the lab manager.
- Gained practical experience working at the intersection of experimental linguistics, technical support, and research workflows.

Education

University of Copenhagen

September 2023 – April 2026

MSc in IT & Cognition, Copenhagen, Denmark

Master's degree focused on computational linguistics, cognitive science, NLP, statistics, data analysis, and empirical language research.

- Master's thesis on pronominal anaphora resolution and coreference-related processing using visual-world eye-tracking data.
- Worked with experimental design, fixation data preprocessing, statistical analysis, visualization, and interpretation of language-processing patterns.
- Relevant areas: NLP, machine learning, data analysis, statistics, cognitive modelling, and computational approaches to language.

University of Granada

September 2018 – June 2023

BA in English Studies / Linguistics, Granada, Spain

Bachelor's degree with a strong focus on linguistics, syntactic analysis, second language processing, and experimental approaches to language.

- Bachelor's thesis examined pronominal anaphora resolution and attachment ambiguity in Spanish learners of English.
- Combined theoretical linguistic analysis with experimental interpretation.
- Received a mention of excellence in linguistics for outstanding academic performance.

Additional Academic & Technical Interests

- Applied NLP and machine learning for real-world language data.
- Coreference resolution, anaphora resolution, cross-linguistic processing, and multilingual language technology.
- Corpus analysis, annotation workflows, model evaluation, and explainable research pipelines.
- Digital humanities, educational language technology, and interdisciplinary research support.