

Génesis MONTENEGRO

CIFRE PhD Student

genesismontenegro07@gmail.com |  genesis-montenegro |  gmontenegrou | Toulouse, France

PRESENTATION

Dynamic and motivated PhD Student with a foundation in linguistics, computer science, and data science. Experienced in designing and implementing research projects, including ontology development, semantic data modeling, and multimodal data integration. Skilled in creating structured datasets from unstructured and heterogeneous data, with practical expertise in ML, NLP, Knowledge Graphs and Computer Vision.

EDUCATION

- **INRIA & Université Côte d'Azur | Berger Levrault** September 2022 - 2028
CIFRE PhD in Computer Science Toulouse-Nice, France
 - Thesis title: From raw corpus to lexical-semantic graph: a methodological framework for industrial, agile and high-performance GraphRAG.
- **Toulouse University – Jean Jaurès ** September 2022 - July 2024
Master 2 Linguistics, Computer Science, Language Technologies (Distinction) Toulouse, France
 - Research dissertations:
 - * **M2**: Compilation of a list of meronymy and cause relationship markers in Spanish. *Distinction*.
 - * **M1**: Identification of hyperonymy relationship markers in specialized Spanish corpora. *High Distinction*.
 - **M1 Internship**: Corpus construction and terminological data creation for a language teaching application, *LAIRDIL Laboratory*, Paul Sabatier University, Toulouse.
 - Supervised projects:
 - * **M2**: **DeSeMod**, managing modals in procedural texts in French. Client: Safety Data.
 - * **M1**: **Automatic classification of messages based on stance** (8 languages), Touché Challenge 2023.
 - Courses: Machine Learning, Computer Science for NLP, Applied Statistics to Language, NLP Methods and Applications, Neural Methods.
- **University of Los Andes ** September 2014 - February 2021
Bachelor's Degree in Hispanic Language and Literature (High Distinction) Mérida, Venezuela
 - Research dissertation: "Abro hilo": The beginning of threads on Twitter in Spanish. *High Distinction*.

EXPERIENCE

- **Berger-Levrault ** Mars 2025 – Present
Research Engineer Toulouse (on-site)
 - Research focused on improving LLM-assisted response engines.
 - Application of knowledge engineering to develop adapted and robust solutions.
 - Exploring innovative approaches in NLP and generative models to optimize response generation.
- **CNRS ** September 2024 – October 2024
Research Officer Toulouse (on-site)
 - Conducted a statistical analysis of +4,000 data entries with over 30 descriptors, extracting actionable insights.
 - Designed and developed more than 5 innovative classification models, including **Machine Learning** and **Transformers**-based approaches, as well as hybrid models.
- **S2RL Supplies and Services INC ** January 2022 – October 2022
Developer | Data Scientist Mexico (remote)
 - Annotated image data and created datasets for training **Computer Vision** models.
 - Refined and optimized 5+ Transformer models in **Python** for classification and image detection tasks.
 - Designed and implemented user experience in web applications for asset monitoring.
 - Analyzed visual, numerical, and linguistic data, generating **interactive visual reports with Matplotlib**.
- **Mammut.io ** January 2021 – May 2022
Data Analyst | Data Scientist Venezuela (remote)
 - Explored and analyzed over 10 datasets relevant to **3 thematic projects**.
 - Designed, structured, and manipulated data.
 - Fine-tuned more than 15 Transformer models in **Python** (token / text classification).
 - Analyzed model errors and proposed solutions to improve performance.
- **Mammut.io ** January 2018 – December 2020
Computational Linguist | Data Scientist Venezuela (remote)

- Designed, structured, and manipulated data (**SQL, NoSQL, Excel, OWL**) to generate knowledge that informed chatbot decision-making.
- Developed rule-based semantic models to enhance text analysis system accuracy.
- Implemented **+5 projects** for lexicalization, annotation, and disambiguation of corpora in English and Spanish across 3 domains (healthcare, hospitality, and law).
- **Authored technical documentation** for ongoing applications, ensuring clear understanding for users and developers.
- **Mammut.io**  May 2017 – December 2017
Intern Venezuela (remote)
 - Evaluated existing NLP implementations in Spanish and English to **identify solutions** best suited for project needs.
 - Designed and implemented annotation taxonomies (**text classification** (5 linguistic levels)).

PROJECTS

- **Mar-Rel-es: Conceptual Relationship Marker List** 2023
Tools: Python, Pandas, NLTK, Data annotation 
 - Developed a list of markers for hyperonymy, meronymy, and cause relationships in Spanish.
 - Implemented natural language processing methods to identify and classify relationship markers.
 - Created visualizations to analyze conceptual relationships in specialized corpora.
- **Teggs-me: Computer Vision Object Identification Program** 2022
Tools: Python, OpenCV, TensorFlow, hugsvision, Scikit-learn 
 - Developed a computer vision system to detect eggs in small poultry farms.
 - Implemented an object identification model with TensorFlow to trigger automatic alerts when all hens have laid eggs.
 - Created interactive reports to monitor production status in the farms.
- **Cross-domain NER: Named Entity Recognition Across Domains** 2021
Tools: Python, Spacy, Transformers 
 - Developed a Named Entity Recognition (NER) system with knowledge transfer between different domains.
 - Implemented Transformer models to improve NER performance in specialized and heterogeneous corpora.
 - Created visualizations for the analysis of detected entities and their relevance in each domain.

PUBLICATIONS

C=COMMUNICATION, J=JOURNAL

- [C.1] G. Montenegro (July 2024). **Constitution d’une liste de marqueurs de relations de méronymie et de cause en espagnol**. 5th Printemps des Jeunes Chercheur-e-s CLLE.
- [C.2] G. Montenegro. **Communication “Éduquer à l’IA”**. Preparing the 7th Colloquium: *L’Intelligence Artificielle dans tous ses états* [Artificial Intelligence in all its states].
- [J.1] G. Montenegro (2019). **El sufijo diminutivo en el habla de hombres y mujeres de Mérida**. *Lengua y Habla*, Vol. 23, Issue 1, pp. 569–578.
- [J.2] I. González, G. Montenegro, R. Pulido, M. Riveros, H. Romero, L. Urdaneta (2019). **Modelado computacional de preposiciones en inglés y español: aproximación cognitiva**. *Lengua y Habla*, Vol. 23, Issue 1, pp. 496–505.
- [J.3] I. González, G. Montenegro, R. Pulido, M. Riveros, H. Romero, L. Urdaneta (2019). **Prefijación en español: una propuesta de formalización computacional basada en modelos cognitivos**. *Lengua y Habla*, Vol. 23, Issue 1, pp. 506–516.

SKILLS

- **Programming Languages:** Python, R, SQL, NoSQL, HTML, JavaScript, CSS.
- **Data Science:** TensorFlow, Scikit-learn, NLTK, Keras, Spacy, Transformers, Matplotlib.
- **Task Management Tools:** Confluence, Asana, Trello, Jira
- **Research Skills:** Data Annotation (manual and automatic), Machine Learning Model Fine-tuning, Report Writing, Corpus Building, Statistical Text Analysis
- **Linguistics:** Lexical Semantics, Morphology, NLP, Corpus Linguistics.
- **Specialized Area:** Textual Data Analysis, Ontology Creation (OWL, Protégé), Corpus Creation (XML, json)
- **Software:** Tableau, Git, GitLab, GitHub, MS Excel, Google Sheets, Google Colab.
- **Languages:** Native: Spanish, Fluent: French, English