Alberto Muñoz Ortiz

alberto.munoz.ortiz@proton.me | amunozo.com

Summary

NLP researcher with a PhD in Computer Science and over five years of experience developing NLP models and building training and evaluation workflows. First author of 8 papers at NLP venues including ACL, EMNLP, and COLING, with work on language model evaluation, multilingual and data-efficient NLP, and structured prediction.

Available immediately.

Experience

PhD Researcher Mar 2021 — Jun 2026
University of A Coruña *A Coruña, Spain*

- Published 8 first-author papers on LLM evaluation, data-efficient NLP, structured prediction, multilingual transfer, and model analysis.
- Developed sequence-labeling methods for dependency, constituent, and semantic parsing and nested named entity recognition.
- Evaluated three multilingual pre-trained models across 13 dependency and 10 constituent treebanks, examining how pre-training coverage and tokenization affect syntactic representations.
- Built pipelines comparing human and LLM-generated text across syntactic, lexical, and emotional features, with generation, annotation, statistical analysis, and reporting.
- Built reusable PyTorch and Hugging Face frameworks for model training and evaluation.
- Ran reproducible training and evaluation workflows on GPU clusters with Docker, Kubernetes, Slurm, Run:ai, and Weights & Biases.
- Contributed to the ERC Starting Grant project FASTPARSE on fast and accurate parsing.

Visiting Researcher Oct 2024 — Mar 2025
EPFL *Lausanne, Switzerland*

- Studied how transformers represent and generalize knowledge in Antoine Bosselut's NLP Lab.
- Built Docker and Run:ai workflows to queue, monitor, and log parallel experiments across multiple A100 GPUs.

Visiting Researcher Jul 2023 — Oct 2023
Ludwig-Maximilian University *Munich, Germany*

- Trained and evaluated a pixel-based language model on four 80 GB A100 GPUs for zero-shot transfer from Standard German to dialects across four tasks, outperforming token-based models by up to 26 percentage points in some settings.

Python Developer Intern Nov 2019 — Nov 2020
Alstom *Madrid, Spain*

- Replaced a manual spreadsheet process with a Python and SQLite application for reliability and maintainability data. Added validation and reporting used by engineering stakeholders.

Education

University of A Coruña A Coruña, Spain
PhD in Computer Science *Oct 2021 — Jun 2026*

- Completed with *cum laude* and International Doctorate distinctions.
- *Thesis:* "Encoding, Representation, and Interpretation of Linguistic Structure Across NLP Paradigms". Examines how linguistic structure can support efficient NLP systems and help evaluate language models.

UNED Spain
MSc in Language Technologies *2020 — 2021*

- *Thesis:* "Dependency Parsing as Sequence Labeling for Low-Resource Languages".

Complutense University of Madrid Madrid, Spain
BSc in Physics *2014 — 2019*

- *Thesis:* "Sociophysics: Physics of Complex Systems Applied to Social Systems".

Skills

- **Research areas:** LLM evaluation, model behaviour, data-efficient NLP, multilingual transfer, structured prediction, information extraction, probing, ablation studies, error analysis
- **ML stack:** Python, PyTorch, Hugging Face (Transformers, Datasets), scikit-learn, pandas, NumPy, SciPy, spaCy, NLTK
- **Research infrastructure:** Docker, Kubernetes, Slurm, Run:ai, Git, Linux, pytest, Weights & Biases, SQLite
- **Languages:** Spanish (Native), English (C2), French (B2), German (B1), Chinese (A1)

Selected Publications

- **Alberto Muñoz-Ortiz**, David Vilares, Caio Corro, and Carlos Gómez-Rodríguez. *Nested Named Entity Recognition as Single-Pass Sequence Labeling*. Findings of EMNLP 2025.
  - Developed a single-pass sequence-labeling method for nested named entity recognition.
- **Alberto Muñoz-Ortiz**, Verena Blaschke, and Barbara Plank. *Evaluating Pixel Language Models on Non-Standardized Languages*. COLING 2025.
  - Evaluated zero-shot transfer from Standard German to dialects across four tasks; pixel-based models improved over token-based models by up to 26 percentage points in some settings.
- **Alberto Muñoz-Ortiz**, Carlos Gómez-Rodríguez, and David Vilares. *Contrasting Linguistic Patterns in Human and LLM-Generated News Text*. Artificial Intelligence Review, 2024.
  - Compared human and LLM-generated news text across syntactic, lexical, and emotional features.

Additional Publications

- Inés Veiga Menéndez, **Alberto Muñoz-Ortiz**, and David Vilares. *On the Impact of Syntactic Infusion for Gender Categorization Across Contextual Dimensions*. Revista de Procesamiento de Lenguaje Natural, 2025.
- **Alberto Muñoz-Ortiz**, Carlos Gómez-Rodríguez, and David Vilares. *Demographic Background Prompting Does Not Affect Linguistic Features on LLM-Generated News Texts*. XoveTIC 2024.
- **Alberto Muñoz-Ortiz**, Carlos Gómez-Rodríguez, and David Vilares. *Assessment of Pre-Trained Models Across Languages and Grammars*. IJCNLP-AACL 2023.
- **Alberto Muñoz-Ortiz** and David Vilares. *LyS A Coruña at GUA-SPA@IberLEF2023: Multi-Task Learning with Large Language Model Encoders for Guarani-Spanish Code Switching Analysis*. IberLEF 2023.
- **Alberto Muñoz-Ortiz** and David Vilares. *Another Dead End for Morphological Tags? Perturbed Inputs and Parsing*. Findings of ACL 2023.
- **Alberto Muñoz-Ortiz**, Mark Anderson, David Vilares, and Carlos Gómez-Rodríguez. *Parsing Linearizations Appreciate PoS Tags - but Some Are Fussy about Errors*. IJCNLP-AACL 2022.
- **Alberto Muñoz-Ortiz**, Carlos Gómez-Rodríguez, and David Vilares. *Cross-lingual Inflection as a Data Augmentation Method*. 3rd Workshop on Insights from Negative Results in NLP, 2022.
- **Alberto Muñoz-Ortiz**, Michalina Strzyz, and David Vilares. *Not All Linearizations Are Equally Data-Hungry in Sequence Labeling Parsing*. RANLP 2021.

Academic Service & Peer Review

- **Conferences & Workshops:** ACL Rolling Review (ARR 2024–2025), ACL (2023), EMNLP (2023), EACL (2023), COLING (2022, 2024), RepL4NLP @ ACL 2024, ALTNLP (2022), CAEPIA (2024)
- **Journals:** Computational Linguistics (MIT Press), ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)