José Daniel GUTIÉRREZ NIETO

La Paz, Mexico | +52-612-158-3817 | danezoj@gmail.com | linkedin.com/in/danezoj | github.com/ezojg

PROFILE

Computational Linguist with a background in Cognitive Science, Linguistics, and Natural Language Processing (NLP). Experienced in NLP research and implementation, automatic biocuration, linguistic consulting, and interdisciplinary collaboration. Skilled in Linux-based workflows, machine learning, data processing, technical writing, and with a detail-oriented approach to problem-solving. Native Spanish speaker and fluent in English.

EXPERIENCE

NeTS (IUSS Pavia) — *Pavia, Italy*
NLP Research Intern — *2023–2024*

- Developed a novel parsing model for complex adjunction phenomena by integrating computational frameworks with contemporary cognitive models of language processing (discontinuous parsing, lexical access, top-down grammars).
- Collaborated with international researchers to present parsing innovations at the *31st Colloquium on Generative Grammar* (Palma, Spain), contributing to cross-disciplinary NLP discussions and advancing the understanding of complex adjunction structures.

Center for Genomics Sciences (CCG-UNAM) — *Cuernavaca, Mexico*
NLP Specialist — *2016–2019*

- Provided NLP solutions for RegulonDB, a database of *E. coli* K-12 genetic regulation knowledge, developing customized text-mining methods for extracting gene-regulation data from biomedical literature databases (automatic biocuration, Python, Linux/Bash, OpenIE).
- Developed hybrid text-mining components combining linguistic rule-based analysis (text simplification, thematic role identification) with machine learning (NER, relation extraction), achieving an 0.84 F-score for identifying gene regulatory pairs in biomedical literature, outperforming comparable state-of-the-art systems.
- The contributions made to this project enabled cross-domain adaptation of biocuration pipelines (from *E. coli* to *Salmonella* literature), demonstrating transferability of the text-mining framework.
- Linked computational linguistics and genomics teams through technical mentoring and interdisciplinary talks.

NLP Research Intern — *2016*

- Provided linguistic consultancy and contributed to modifications in the design of an automatic biocuration pipeline as part of a newly formed NLP division (text structure analysis, evaluation of text simplification methods).
- Prototyped initial text simplification components that were later expanded into production systems (Python, Bash).

Semiosfera Marketing — *Mexico City, Mexico*
Linguistic Consultant — *2016*

- Advised on the design and curation of dialectal Mexican Spanish corpora from field interviews, establishing annotation guidelines and linguistic analysis protocols.
- Participated in developing annotated multimodal corpora, including transcription, analysis, and annotation (discourse analysis, prosodic/phonetic annotation, pragmatic tagging, Praat), to enable NLP applications in marketing research.

Laboratory of Phonetic Studies (El Colegio de México) — *Mexico City, Mexico*
Field Linguist | Speech Data Specialist — *2014*

- Conducted systematic field recordings of dialectal Mexican Spanish with phonetic/sociolinguistic controls across multiple regions, producing speech corpora for prosodic and variationist analysis.

TECHNICAL SKILLS

Core Competencies: NLP, machine learning, data processing, technical writing, academic research, consulting.
Programming: Python, Bash, C, SQL, LaTeX, Markdown.
Tools and Frameworks: PyTorch, scikit-learn, spaCy, Pandas, NumPy, Praat, Audacity, Linux, macOS.
Linguistics and Analysis: Speech processing, generative/dependency grammar, corpus linguistics, pragmatics, discourse.

EDUCATION

University of Trento — *Trento, Italy*
M.Sc. in Cognitive Science — *2022–2023*

University of the Basque Country — *San Sebastián, Spain*
M.Sc. in Computational Linguistics — *2020–2021*

National School of Anthropology and History — *Mexico City, Mexico*
B.A. in Linguistics (with honors) — *2013–2017*

Building the Extended Projection of N: A Top-Down Derivational Perspective. Presentation at the 5th Formal Linguistics Meeting in Mexico (ELF). Benemérita Universidad Autónoma de Puebla, Puebla, 2024.

Toward an Analysis of Ordering Restrictions within Noun Phrases: A Top-Down Minimalist Approach. Poster at the 31st Colloquium on Generative Grammar (CGG). University of the Balearic Islands, Palma, 2022.

Sentence Simplification for Biomedical Literature. Presentation at the Computational Genomics Seminar. CCG-UNAM, Cuernavaca, 2017.

Extraction of Regulatory Interactions and Growth Conditions. Presentation for the Computational Genomics Project. CCG-UNAM, Cuernavaca, 2017.

The Syntax of Adjectives in Generative Theory. Invited lecture. National School of Anthropology and History (ENAH), Mexico City, 2019.

The Noun Phrase and its Modifiers. Invited lecture. National School of Anthropology and History (ENAH), Mexico City, 2017.

Challenges in Designing a Morphosyntactic Interface for Spanish-English NP Machine Translation. Presentation at the 11th Linguistics Colloquium. National School of Anthropology and History (ENAH), Mexico City, 2016.