

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)

NLP Research Intern

Pavia, Italy

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)

NLP Specialist

Cuernavaca, Mexico

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

Linguistic Consultant

Mexico City, Mexico

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)

Field Linguist | Speech Data Specialist

Mexico City, Mexico

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

M.Sc. in Cognitive Science

Trento, Italy

2022–2023

University of the Basque Country

M.Sc. in Computational Linguistics

San Sebastián, Spain

2020–2021

National School of Anthropology and History

B.A. in Linguistics (with honors)

Mexico City, Mexico

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.

PERFIL PROFESIONAL

Lingüista Computacional con formación en ciencia cognitiva, lingüística y procesamiento del lenguaje natural (NLP). Experiencia en investigación e implementación de tecnologías de NLP, curación automática, consultoría lingüística, y colaboración interdisciplinaria. Dominio de entornos Linux, machine learning, procesamiento de datos y redacción técnica. En todos mis proyectos mantengo un enfoque riguroso y con atención al detalle. Bilingüe en español (nativo) e inglés (fluido).

EXPERIENCIA DE TRABAJO

NeTS (IUSS Pavia)

Pavia, Italia

Investigador en NLP (Pasante)

2023–2024

- Diseñé un modelo experimental de parsing para fenómenos de adjunción complejos, integrando frameworks computacionales con modelos cognitivos modernos sobre el procesamiento del lenguaje (discontinuous parsing, teoría del acceso léxico, gramáticas generativas top-down).
- Colaboré con investigadores internacionales para presentar innovaciones en parsing en el *31st Colloquium on Generative Grammar* (Palma, España), contribuyendo a discusiones sobre metodologías interdisciplinarias en NLP y al avance en la comprensión de estructuras complejas de adjunción.

Centro de Ciencias Genómicas (CCG-UNAM)

Cuernavaca, México

Especialista en NLP

2016–2019

- Implementé soluciones en NLP para RegulonDB, una base de datos de regulación genética en *E. coli* K-12, desarrollando métodos de minería de texto especializados para extraer datos de regulación genética de literatura biomédica (biocuración automática, Python, Linux/Bash, OpenIE).
- Diseñé componentes híbridos de minería de texto combinando análisis basado en reglas lingüísticas (simplificación de texto, identificación de roles temáticos) con machine learning (reconocimiento de entidades, extracción de relaciones), logrando un F-score de 0.84 en identificación de pares regulatorios, superando a los sistemas existentes comparables.
- Dichas contribuciones permitieron la adaptación transversal de pipelines de biocuración (de literatura sobre *E. coli* a *Salmonella*), demostrando la transferibilidad de los métodos desarrollados.
- Establecí vínculos entre los equipos de lingüística computacional y de genómica mediante mentoría técnica y presentaciones interdisciplinarias regulares.

Investigador en NLP (Pasante)

2016

- Brindé asesoría lingüística y contribuí al rediseño de un pipeline de biocuración automática en la recién formada división de NLP (análisis de estructura sintáctica, evaluación de métodos de simplificación de texto).
- Desarrollé prototipos de componentes para simplificación de texto que posteriormente se escalaron a sistemas de producción (Python, Bash).

Semiosfera Marketing

Ciudad de México, México

Consultor Lingüístico

2016

- Asesoré sobre el diseño y la curaduría de corpus anotados de español mexicano con datos recopilados mediante entrevistas de campo orales, estableciendo lineamientos de anotación y análisis lingüístico.
- Colaboré en el desarrollo de corpus multimodales anotados (transcripción, análisis del discurso, anotación fonética y prosódica, etiquetado pragmático, Praat) para aplicaciones de NLP en investigación de mercado.

Laboratorio de Estudios Fónicos (El Colegio de México)

Ciudad de México, México

Lingüista de campo | Especialista en datos de habla

2014

- Realicé grabaciones de campo sistemáticas de español mexicano dialectal con controles fonéticos y sociolingüísticos en múltiples regiones, generando corpus de habla para el análisis prosódico y variacionista.

HABILIDADES TÉCNICAS

Competencias principales: NLP, machine learning, procesamiento de datos, redacción técnica, investigación, consultoría.

Programación: Python, Bash, C, SQL, LaTeX, Markdown.

Herramientas y frameworks: PyTorch, scikit-learn, spaCy, Pandas, NumPy, Praat, Audacity, Linux, macOS.

Análisis lingüístico: Procesamiento del habla, lingüística de corpus, gramática generativa y de dependencia, análisis pragmático-discursivo.

Università degli Studi di Trento*M.Sc. in Cognitive Science**Trento, Italia**2022–2023***Universidad del País Vasco***M.Sc. in Computational Linguistics**San Sebastián, España**2020–2021***Escuela Nacional de Antropología e Historia***Licenciatura en Lingüística (mención honorífica)**Ciudad de México, México**2013–2017***PRESENTACIONES Y PONENCIAS**

Construyendo la proyección extendida de N: Una perspectiva derivacional descendente. Conferencia en el V Encuentro de Lingüística Formal en México (ELF). Benemérita Universidad Autónoma de Puebla, Puebla, 2024.

Towards an Analysis of Ordering Restrictions within Noun Phrases: A Top-Down Minimalist Approach. Presentación de póster en el 31st Colloquium on Generative Grammar. Universidad de las Islas Baleares, Palma, 2022.

Sentence Simplification for Biomedical Literature. Presentación para el Seminario de Genómica Computacional. CCG-UNAM, Cuernavaca, 2017.

Extraction of Regulatory Interactions and Growth Conditions. Presentación para el Programa de Genómica Computacional. CCG-UNAM, 2018.

La sintaxis de los adjetivos en la teoría generativa. Ponente invitado. ENAH, Ciudad de México, 2019.

La frase nominal y sus modificadores. Ponente invitado. ENAH, Ciudad de México, 2017.

Problemas en el diseño de una interfaz morfosintáctica de traducción automática de FNs entre español e inglés. Conferencia en el XI Coloquio de Lingüística en la ENAH. ENAH, Ciudad de México, 2016.