Lucía **Ormaechea Grijalba**

Ph.D. Candidate · NLP Researcher · Computational Linguist

☐ (+41) 22 37 98684 | ✉ l.ormaechea@hotmail.com | 🏠 luciaormaechea.com | lormaechea
| in lucía-ormaechea-grijalba | 🎓 Lucía Ormaechea Grijalba

Experience

Research & Teaching Assistant
Geneva, Switzerland

Université de Genève
Dec. 2020 - Present

- Contributed to **BabelDr** ⧉ :
  - Developed and enhanced a **specialized Automatic Speech Recognition (ASR)** system based on Kaldi.
  - Deployed a **multi-container Docker application** to perform ASR inside the **BabelDr medical translation device**.
- Participated in **PROPICTO** ⧉ :
  - Front-end development and maintenance of the **project's website**.
  - **Data preparation** and **compilation** for future studies (e.g. creation of a corpus of open source images regarding medical terms).
- Collaborated as an assistant in **Localization** ⧉ (M.A. Course): preparation of course materials, assignment grading, assistance and support to students during practical sessions.

NLP Research Intern
Grenoble, France

Laboratoire d'Informatique de Grenoble (LIG)
Feb. 2020 - Sep. 2020

- Participated in BabelDr: development of a **Kaldi-based ASR system for medical-related applications**.
- **Main tasks**:
  - Created a complete pipeline for injecting **grammar-based language models** into the Kaldi Speech Recognition Toolkit.
  - Trained **HMM-DNN-based acoustic models** using open-source corpora.
  - Developed a **Kaldi web-server API** for ASR applications, containerized ASR-related tools using Docker.
  - Conducted a prototype testing and evaluation.

Translation Intern
Remote Working

New York Habitat
Jul. 2018 - Sep. 2018

- Translated **ENG > ESP** commercial texts: video transcriptions – travel articles – apartment reviews – client testimonials.
- **Selected publications**:
  - *Bajo la sombra del cine: una guía del cine en Nueva York (parte 1: Manhattan)* ⧉ , (2018).
  - *El cine de los distritos: un recorrido por el cine de Nueva York (parte 2)* ⧉ , (2018).
  - *Guía de viaje de verano: nuestros mejores planes en Londres* ⧉ , (2018).

Undergraduate Research Assistant
Pamplona, Spain

Universidad de Navarra
Sep. 2016 - Jun. 2017

- Collaborated as a research student at the **Department of Philology** ⧉
- **Main tasks**:
  - Development of educational materials aiming to automatically evaluate knowledge of Spanish language in L2 learners.
  - One-on-one tutorial assessment on academic writing.
  - Proofreading of academic papers, document classification.

Education

Ph.D. in Multilingual Information Processing
Geneva, Switzerland

Université de Genève
Jun. 2022 - present

- **Ph.D. Project**: *Exploring Automatic Spoken Language Simplification in Low-Resource Conditions* ⧉

M.Sc. in Natural Language Processing
Paris, France

Institut National des Langues et Civilisations Orientales (INALCO) / Université Sorbonne-Nouvelle (Paris 3)
Sep. 2018 - Sep. 2020

- **Graduated with High Honors.**
- **Coursework**: Scripting (Bash, Python, Perl) – Object-Oriented Programming (C++, Java) – Statistical Methods for Corpus Exploitation – Text Mining – Convolutional Neural Networks (Keras) – Corpus Linguistics – Mark-up Languages (XML, XSLT) – Databases (SQL, Neo4J).
- **Master's Thesis**: *Mise en place d'un système robuste de reconnaissance automatique de la parole appliqué au domaine médical* ⧉ **GPA: 19/20**

B.A. in Hispanic Philology
Pamplona, Spain

Universidad de Navarra
Sep. 2014 - Jun. 2018

- **Extraordinary End-of-Degree Award Nominee.**
- **Coursework**: Lexicology – Semantics – Phonetics – Phonology – Morphology – Syntax – Discourse Analysis.

Teaching & Assistantships

2021-2022   **M.A. in Translation, M.A. in Multilingual Communication Technology,** *Teaching Assistant* - Localization (L. Morado Vázquez), Université de Genève

2017   **Workshops on Spanish Language,** *Instructor* - Spanish grammar courses for non-native undergraduate students, Universidad de Navarra

Languages

	Listen	Read	Speak	Write	
Spanish	Native	Native	Native	Native	
English	C1	C1	C1	C1	*Advanced Certificate in English (CAE)*
French	C1	C1	C1	C1	*Diplôme Approfondi de Langue Française (DALF)*
Italian	A2	A2	A2	A2	

Software Skills

Programming	Python, Bash (*proficient*) – Perl, SQL (*comfortable*) – C++, Java (*some experience*)
Libraries	OpenFST – OpenSSH – Flask – Keras – Speechbrain – Pandas – NLTK – SpaCy – OpenNMT
Web Development	HTML – CSS – Jekyll – XML – XSLT – JavaScript
Tools	Kaldi – SRILM – Git – Vim – LaTeX – Docker – Praat – Nginx
Operating Systems	GNU/Linux (Ubuntu, Debian, Linux Mint) – Windows – Android

Scholarships & Awards

2020   *Candoc* **Research Fellowship,** Granted by the Swiss National Science Foundation (SNSF)

2013   **Runner-up at the 4th Inter-University Debate Tournament** ⬈ **,** Universidad Pública de Navarra

Publications

2022   Jonathan Mutal, Pierrette Bouillon, Johanna Gerlach, Magali Norré and **Lucía Ormaechea Grijalba,** *A Neural Machine Translation Approach to Translate Text to Pictographs in a Medical Speech Translation System – The BabelDr Use Case* ⬈ . In: Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (AMTA), Remote Event.

2022   **Lucía Ormaechea Grijalba,** Benjamin Lecouteux, Pierrette Bouillon and Didier Schwab. *A Tool for Easily Integrating Grammars as Language Models into the Kaldi Speech Recognition Toolkit* ⬈ . In: ESSLLI 2022 - Workshop on Bridges and Gaps between Formal and Computational Linguistics, Galway (Ireland).

2022   Cécile Macaire, **Lucía Ormaechea Grijalba,** and Adrien Pupier. *Une chaîne de traitements pour la simplification automatique de la parole et sa traduction automatique vers des pictogrammes* ⬈ . In: 29ème Conférence sur le Traitement Automatique des Langues Naturelles (TALN/RÉCITAL), Avignon (France).

2022   **Lucía Ormaechea Grijalba,** *Integrating Grammar-Based Language Models into Domain-Specific ASR Systems* ⬈ . In: Second Advanced Language Processing School (ALPS), Remote Event.

2021   **Lucía Ormaechea Grijalba,** Pierrette Bouillon, Johanna Gerlach, Benjamin Lecouteux, Didier Schwab and Hervé Spechbach. *Reconnaissance vocale du discours spontané pour le domaine médical* ⬈ . In: Journée Commune AFIA/TLH, Remote Event.

Toolkits & Datasets

2021   **kaldi-grammar-compiler** ⬈ **,** Language modeling tool designed to transform fixed grammars into compiled Finite State Transducers (FSTs). This allows to make them readable as language models by the Kaldi Speech Recognition Toolkit.

2021   **MeDiCo (A Medical Discourse Corpus in French)** ⬈ **,** Crowdsourced artificial spoken dataset including utterances related to the medical consultation domain.

Other Achievements

Sequence Models   *Online Course*

Coursera, DeepLearning.AI   *November 2022*

Deep Learning Specialization (V): Recurrent Neural Networks – Long Short Term Memory – Attention Models.

Structuring Machine Learning Projects

Online Course

COURSERA, DEEPLEARNING.AI

October 2022

Deep Learning Specialization (III): Error Analysis in Machine Learning Systems – Performance Improvement Strategies.

Improving Deep Neural Networks: Hyperparameter Tuning, Regularization & Optimization

Online Course

COURSERA, DEEPLEARNING.AI

October 2022

Deep Learning Specialization (II): Optimization Algorithms – Hyperparameter Tuning.

Neural Networks & Deep Learning

Online Course

COURSERA, DEEPLEARNING.AI

October 2022

Deep Learning Specialization (I): Deep Learning Foundations – Neural Network Architectures.

Personal Data Protection

MOOC Course

INSTITUT NATIONAL DE RECHERCHE EN SCIENCES ET TECHNOLOGIES DU NUMÉRIQUE (INRIA)

May 2020

MOOC Course introducing data confidentiality regulations and privacy protection.

Introduction to Corpus Linguistics

MOOC Course

UNIVERSITÉ GRENOBLE-ALPES (UGA)

June 2019

MOOC Course introducing corpus linguistics and its many implementations in lexicology, discourse analysis and NLP.

Projet Boîte à Outils ⬀

Master's Project

INSTITUT NATIONAL DES LANGUES ET CIVILISATIONS ORIENTALES (INALCO) / UNIVERSITÉ SORBONNE-NOUVELLE (PARIS 3)

May 2019

A quantitative research focused on morphosyntactic patterns found on the French newspaper Le Monde.

Projet Documents Structurés ⬀

Master's Project

INSTITUT NATIONAL DES LANGUES ET CIVILISATIONS ORIENTALES (INALCO) / UNIVERSITÉ SORBONNE-NOUVELLE (PARIS 3)

May 2019

A set of programming exercises related to XML and information extraction.