



Lucía Ormaechea Grijalba

PH.D. CANDIDATE · NLP RESEARCHER · COMPUTATIONAL LINGUIST

☎ (+41) 78 220 49 79 | 📧 lucía-ormaechea-grijalba | 🌐 luciaormaechea.com | 📱 lormaechea
| ✉ Lucia.OrmaecheaGrijalba@unige.ch | 🎓 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** [↗](#):
 - Compilation of **complex-simple sentence pairs** from comparable corpora to be further used for the **generation of simplified texts**.
 - Front-end development and maintenance of the **project's website**.
- 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 – UNIVERSITÉ GRENOBLE-ALPES

Jun. 2022 - present

- Currently pursuing a **joint Ph.D.** between the **Département de Translation Technology** [↗](#) at UNIGE, and the **GETALP team** [↗](#), at UGA.
- **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

- 2021-2023 **M.A. in Translation, M.A. in Multilingual Communication Technology** – *Teaching Assistant*, Localization (L. Morado Vázquez [2021, 2022], S. Girletti [2023]), 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					<i>Native speaker</i>
English	C1	C1	C1	C1	<i>Advanced Certificate in English (CAE)</i>
French	C1	C1	C1	C1	<i>Diplôme Approfondi de Langue Française (DALF)</i>
Italian	B1	B1	A2	A2	

Software Skills

Programming	Python, Bash (<i>proficient</i>) – Perl, SQL (<i>comfortable</i>) – C++, Java (<i>some experience</i>)
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
Software	Microsoft Office – Gate – Trados – Passolo – Adobe Photoshop – GIMP

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

Selected Publications

- 2023 **Lucía Ormaechea**, Nikos Tsourakis, Didier Schwab, Pierrette Bouillon and Benjamin Lecouteux. *Simple, Simpler and Beyond: A Fine-Tuning BERT-Based Approach to Enhance Sentence Complexity Assessment for Text Simplification* [↗](#). In: Proceedings of the 6th International Conference on Natural Language and Speech Processing [↗](#) (ICNSLP), Trento (Italy).
- 2023 **Lucía Ormaechea** and Nikos Tsourakis. *Extracting Sentence Simplification Pairs from French Comparable Corpora Using a Two-Step Filtering Method* [↗](#). In: Proceedings of the 8th Swiss Text Analytics Conference [↗](#), Neuchâtel (Switzerland).
- 2023 **Lucía Ormaechea**, Pierrette Bouillon, Maximin Coavoux, Emmanuelle Esperança-Rodier, Johanna Gerlach *et al.* *PROPICTO: Developing Speech-to-Pictograph Translation Systems to Enhance Communication Accessibility* [↗](#). In: Proceedings of the 24th Annual Conference of The European Association for Machine Translation [↗](#) (EAMT), Tampere (Finland).
- 2022 **Lucía Ormaechea**, 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** and Adrien Pupier. *Une chaîne de traitements pour la simplification automatique de la parole et sa traduction automatique vers des pictogrammes* [↗](#). In: 29^{ème} Conférence sur le Traitement Automatique des Langues Naturelles et Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues [↗](#) (TALN/RÉCITAL), Avignon (France).
- 2021 **Lucía Ormaechea**, Pierrette Bouillon, Johanna Gerlach, Benjamin Lecouteux, Didier Schwab and Hervé Spechbach. *Reconnaissance vocale du discours spontané pour le domaine médical* [↗](#). In: Journée Commune Technologies du Langage Humain [↗](#) (AFIA/TLH), Remote event.

Toolkits & Datasets

- 2023 **WiViCo (Wikipedia Vikidia Corpus)** [↗](#) – A general-purpose parallel sentence simplification dataset for French, automatically extracted by using a two-step filtering method to mine register-diversified comparable corpora.
- 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.

Academic Service

- 2023 **Bruit de fond ou valeur ajoutée ? Gérer le bruit lors des traitements informatiques des corpus linguistiques** [↗](#) – Seminar jointly organized by Université Grenoble-Alpes and Sapienza Università di Roma – **Member of Scientific Committee**
- 2018 **5th Forum of the Association for Discourse Linguistics** [↗](#) – Held at Instituto de Cultura y Sociedad, Universidad de Navarra – **Local Organizer**

Certifications

3rd Advanced Language Processing School (ALPS) [↗](#)

UNIVERSITÉ GRENOBLE-ALPES – NAVER LABS EUROPE

Winter school featuring advanced lectures, social and poster sessions covering a wide range of hot topics in NLP.

Winter School

January 2023

Research Methods in Translation & Interpreting Studies (ReMeTIS II): Specific Research & Scientific Communication Skills [↗](#)

UNIVERSITÉ DE GENÈVE

Second part of the ReMeTIS programme: Corpus Construction – Quantitative Data Analysis – Career & Publication Strategies.

Online Course

December 2022

Research Methods in Translation & Interpreting Studies (ReMeTIS I): Foundations & Data Analysis [↗](#)

UNIVERSITÉ DE GENÈVE

First part of the ReMeTIS programme: Qualitative Methods – Mixed Methods – Research Ethics.

Online Course

December 2022

Sequence Models [↗](#)

COURSERA, DEEPLARNING.AI

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

Online Course

November 2022

Structuring Machine Learning Projects [↗](#)

COURSERA, DEEPLARNING.AI

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

Online Course

October 2022

Improving Deep Neural Networks: Hyperparameter Tuning, Regularization & Optimization [↗](#)

COURSERA, DEEPLARNING.AI

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

Online Course

October 2022

Neural Networks & Deep Learning [↗](#)

COURSERA, DEEPLARNING.AI

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

Online Course

October 2022

33rd European Summer School in Logic, Language & Information (ESLLI) [↗](#)

NATIONAL UNIVERSITY OF IRELAND GALWAY

Interdisciplinary summer school involving courses and workshops in logic, linguistics and computer science.

Summer School

August 2022

2nd Advanced Language Processing School (ALPS) [↗](#)

UNIVERSITÉ GRENOBLE-ALPES – NAVER LABS EUROPE

Winter school featuring advanced lectures, social and poster sessions covering a wide range of hot topics in NLP.

Winter School

January 2022

Personal Data Protection [↗](#)

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

MOOC Course introducing data confidentiality regulations and privacy protection.

MOOC Course

May 2020

Introduction to Corpus Linguistics [↗](#)

UNIVERSITÉ GRENOBLE-ALPES (UGA)

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

MOOC Course

June 2019

Other Achievements

Projet Boîte à Outils [↗](#)

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

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

Master's Project

May 2019

Projet Documents Structurés [↗](#)

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

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

Master's Project

May 2019