# Paolo Gajo

☐ +39 345 5665 460 | @ paolo.gajo2@unibo.it | 🔗 LinkedIn | ⭘ GitHub | 📍 Forlì, Italy

PhD student in NLP at the University of Bologna. Current PhD work involves researching graph neural networks and large language models, and their use for semantic dependency parsing, information extraction, and natural language understanding of graph-structured texts. My thesis is on the application of these techniques for the automatic adaptation of culinary recipes, based on user dietary constraints.

## PUBLICATIONS

- **Paolo Gajo**, Domenic Rosati, Hassan Sajjad, and Alberto Barrón-Cedeño. Dependency Parsing is More Parameter-Efficient with Normalization. In *Proceedings of the 39th Conference on Neural Information Processing Systems (NeurIPS 2025)*, 2025.

- **Paolo Gajo** and Alberto Barrón-Cedeño. Natural vs Programming Language in LLM Knowledge Graph Construction. *Information Processing & Management*, 62(5):104195, 2025.

- **Paolo Gajo**, Daniele Polizzi, Adriano Ferraresi, and Alberto Barrón-Cedeño. LEARN: on the feasibility of Learner Error AutoRegressive Neural annotation. In *CLiC-it 2025: 11th Italian Conference on Computational Linguistics*, 2025.

- **Paolo Gajo** and Alberto Barrón-Cedeño. On Cross-Language Entity Label Projection and Recognition. In *CLiC-it 2024: 10th Italian Conference on Computational Linguistics*, 2024.

- **Paolo Gajo**, Luca Giordano, and Alberto Barrón-Cedeño. UniBO at CheckThat! 2024: Multi-lingual and Multi-label Persuasion Technique Detection in News with Data Augmentation and Sequence-Token Classifiers. In *Working Notes of CLEF 2024 - Conference and Labs of the Evaluation Forum*, CLEF 2024, pages 426–434, Grenoble, France, 2024.

- **Paolo Gajo**, Arianna Muti, Katerina Korre, Silvia Bernardini, and Alberto Barrón-Cedeño. On the Identification and Forecasting of Hate Speech in Inceldom. In Ruslan Mitkov and Galia Angelova, editors, *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, pages 373–384, Varna, Bulgaria, September 2023. INCOMA Ltd., Shoumen, Bulgaria.

- **Paolo Gajo**, Silvia Bernardini, Adriano Ferraresi, and Alberto Barrón-Cedeño. Hate Speech Detection in an Italian Incel Forum Using Bilingual Data for Pre-Training and Fine-Tuning. In Federico Boschetti and Nicole Novielli Gianluca E. Lebani, Bernardo Magnini, editors, *CLiC-it 2023: 9th Italian Conference on Computational Linguistics*, volume 3596, Venice, Italy, 2023. CEUR-WS, CEUR-WS.

- Katerina Korre, John Pavlopoulos, **Paolo Gajo**, and Alberto Barrón-Cedeño. Hate Speech According to the Law: An Analysis for Effective Detection. *arXiv preprint arXiv:2412.06144*, 2024.

## PROJECTS

### Local Dish Semantic Recommendation System                                           *2025*
- Built a local dish recommendation system, semantic-querying a novel dataset of GialloZafferano recipes, annotated with geospatial information. Code: [GitHub](#)

### GIARA: Gastronomy and AI for a welcoming Romagna Region                              *2024*
- Collaborated on the development of a text localization system for texts characterized by a high density of terms with highly contextual regional and cultural value (realia). Website: [Unibo.it](#)

### Automatic Subtitle Segmenter                                                         *2023*
- Developed a seq2seq model that automatically segments subtitles by fine-tuning on a dataset of subtitles personally built from scratch. Code: [GitHub](#)

| Programming | Python, Bash |
| --- | --- |
| Infrastructure | On-prem HPC (Slurm), AWS (EC2, S3) |
| Libraries | PyTorch, NumPy, Transformers, Pandas, Sci-Kit Learn, SpaCy, Selenium, BeautifulSoup, NetworkX |
| Software | Git, VS Code, LaTeX, Overleaf, Excel, Label Studio |
| Text processing | RegEx, Sketch Engine, AntConc, BootCaT |
| Translation | Trados, MemoQ, MateCat, MultiTerm, Subtitle Edit, Aegisub |

## Education

### Dalhousie University
Halifax, Canada

*Visiting @ Hypermatrix*
*Sep 2024 − Sep 2025*

- Researching graph neural networks and text-to-graph models. Supervisor: Prof. Hassan Sajjad.
- Courses attended: CSCI 3151 - Foundations of Machine Learning, CSCI 6908 - NLP with Deep Learning, CSCI 6516 - Deep Learning, CSCI 6904 - Deep Reinforcement Learning.

### University of Bologna
Forlì, Italy

*PhD in Natural Language Processing*
*2023 − Present*

- Researching natural language understanding for procedural texts using graph neural networks and large language models. Supervisor: Prof. Alberto Barrón-Cedeño.

### University of Bologna
Forlì, Italy

*MA in Specialized Translation (110/110 Summa Cum Laude)*
*2021 − 2023*

- 2-Year degree on translation technologies and translation in English, Spanish, and Italian.
- Thesis: *This Is My Cope: Identification and Forecasting of Hate Speech in Inceldom.* Code: GitHub

### Ca' Foscari University
Treviso, Italy

*BA in Linguistic and Cultural Mediation (110/110 Summa Cum Laude)*
*2018 − 2021*

- 3-Year degree focusing on translation and interpreting in English, Spanish, and Italian.
- Thesis: *Practical Application of Subtitling Guidelines - Intralingual and Interlingual Subtitling of the Short Film "Group B"*

## Invited Talks

### Osaka University
Osaka, Japan

*Seminar*
*September 2025*

- Topic: Dependency parsing on procedural texts.

### Expo 2025
Osaka, Japan

*Short presentation*
*September 2025*

- Three-minute pitch on culinary recipe understanding and adaptation, followed by a panel discussion.

### Dalhousie University
Halifax, Canada

*Lecture*
*March 2025*

- Lecture on Geometric Deep Learning for the CSCI 6908 - NLP with Deep Learning course.

### Cardiff University
Cardiff, Wales

*Poster presentation*
*July 2024*

- Topic: Data augmentation via MT and label projection in the culinary domain.

### University of Bari Aldo Moro
Bari, Italy

*Poster presentation*
*June 2024*

- Topic: Data augmentation via MT for the improvement of NER models in the culinary domain.

## Teaching

**Deep Learning Reading Group**
- Organizing meetings and presenting chapters from the book "Understanding Deep Learning".

*2025*

## Supervision

**MA Thesis in MT for English-Swahili**
- Co-supervising an MA thesis on the development of an EN-SW MT model for medical texts.

*2025*

**MA Thesis in MT for English-Yoruba**
- Co-supervising an MA thesis on the development of EN-YO MT models.

*2025*

## Reviewing

ACL Rolling Review (ARR)

*2024 – Present*

## Work Experience

**Transperfect**                                                                        Remote
*Freelance Translator*                                                          *2016 – Present*
- Translation of Blizzard Entertainment eSports articles, subtitling of Amazon AWS company videos, subtitling for the deaf and hard-of-hearing of over 30 Nickelodeon TV series. Mobile and desktop video game translations, feature-length film subtitling, subtitles for Netflix documentaries.

**Parole O_Stili**                                                                   Forlì, Italy
*Translator*                                                              *Jun 2022 – Jan 2023*
- Translation and localization of training material for schools and businesses concerning non-hostile and inclusive language.

## Languages

Italian: Native

English: C2 (CEFR)

Spanish: C2 (CEFR)