Alice Fedotova

☐ +39 333 393 6279 | @ alice.fedotova7@gmail.com | 🔗 LinkedIn | ⭘ GitHub

OVERVIEW

Language Technology Specialist and former Research Fellow at the University of Bologna. During my fellowship, I developed tools for corpus construction, managed complex datasets, and conducted experiments with speech-to-text models to enhance transcription accuracy for linguistic analysis. Additionally, during my studies, I worked as a Research Intern on a project exploring the success factors of medical dramas. My work focused on comparing unimodal and multimodal classifiers trained on subtitles and/or images using models including BERT and MMBT, with the aim of automatically identifying the narrative content of a scene.

EXPERIENCE

University of Bologna   Forlì, Italy
Research Fellow   *Oct 2023 – Oct 2024*
- Developed a pipeline to automatically enrich EPTIC, a corpus of European Parliament translations and interpretations, by streamlining transcription, sentence alignment, video alignment, and metadata extraction
- Compared ASR models for the transcription of speeches and interpretations, using WER for evaluation and experimenting with fine-tuning to include features of spoken language such as disfluencies and pauses

University of Bologna   Bologna, Italy
Research Intern   *July 2022 – Apr 2023*
- Investigated the use of deep learning for the automatic identification of isotopies, a novel task in the field of automated content analysis aimed at reducing the cost of annotation for the study of medical dramas
- Expanded an existing supervised dataset and fine-tuned MMBT, a multimodal transformer-based model, obtaining a 7.6% higher F_1 score than an unimodal baseline trained on text by combining subtitles and keyframes

PUBLICATIONS

Fedotova, Alice, Adriano Ferraresi, Maja Miličević Petrović, and Alberto Barrón-Cedeño (2024). "*Constructing a Multimodal, Multilingual Translation and Interpreting Corpus: A Modular Pipeline and an Evaluation of ASR for Verbatim Transcription*". Presented at *CLiC-it 2024 – Tenth Italian Conference on Computational Linguistics*. URL: https://clic2024.ilc.cnr.it/wp-content/uploads/2024/12/41_main_long.pdf.

Fedotova, Alice, Alberto Barrón-Cedeño, Adriano Ferraresi, and Maja Miličević Petrović (2024). "*Expanding the European Parliament Translation and Interpreting Corpus: A Modular Pipeline for the Construction of Complex Corpora*". Presented at the *14th Conference on Language Technologies and Digital Humanities (JT-DH-2024)*. DOI: https://doi.org/10.5281/zenodo.13936433.

Antici, Francesco, Andrea Galassi, Federico Ruggeri, Katerina Korre, Arianna Muti, Alessandra Bardi, Alice Fedotova, and Alberto Barrón-Cedeño (2024). "*A Corpus for Sentence-Level Subjectivity Detection on English News Articles*". Presented at the *2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. URL: https://aclanthology.org/2024.lrec-main.25.

Fedotova, Alice and Alberto Barrón-Cedeño (2023). "*Towards the Automatic Identification of Isotopies*". In: *Investigating Medical Drama TV Series: Approaches and Perspectives. 14th Media Mutations International Conference*, edited by Stefania Antonioni and Marta Rocchi. DOI: https://doi.org/10.21428/93b7ef64.59f47006.

EDUCATION

University of Bologna   Forlì, Italy
MA in Specialized Translation, Translation and Technology Curriculum   *Oct 2021 - July 2023*
- Main Coursework: Natural Language Processing, Language Data Analysis, Machine Translation
- Thesis: *Decoding Medical Dramas: Identifying Isotopies through Multimodal Classification*
- Final Grade: 110/110, Cum Laude

University of Bologna   Forlì, Italy
BA in Intercultural and Linguistic Mediation   *Oct 2017 - Oct 2020*
- Main Coursework: Translation and Liaison Interpreting, Computer-Assisted Translation, General Linguistics
- Thesis: *Proposta di Sottotitolazione di un Estratto del Film "Neverojatnye Priključenija Ital'jantsev v Rossii"*
- Final Grade: 109/110

Skills

Languages	Python, R, SQL, Bash
Libraries	Pandas, NLTK, SpaCy, Scikit-Learn, Keras, PyTorch, Transformers, Selenium, HuggingFace
Translation	Trados, MemoQ, MateCat, MultiTerm, Aegisub, Subtitle Workshop
Text Processing	XML, HTML, RegEx, Sketch Engine, AntConc, BootCaT
Other	Git, VSCode, Overleaf, Slurm, Microsoft Excel, Knime

Academic Projects

Parallel Sentence Identification in Wikipedia Articles – GitHub
- Project within the Profession-Based Research course of the MA in Translation and Technology, addressing the task of identifying relevant Wikipedia articles in English and Italian for the purpose of domain adaptation
- Researched related work and devised a pipeline with Sentence-Transformers, extracting an in-domain comparable corpus and 10,069 parallel sentences using multilingual embeddings and cosine similarity

Isotopy Identification as Label Distribution Learning – GitHub
- Project within the Natural Language Processing course of the MA in Translation and Technology, addressing the isotopy identification task as a label distribution learning problem using continuous target values
- Trained CNNs and BERT on textual features and experimented with hyperparameter optimization, obtaining a 17% lower MAE with BERT compared to a naive baseline model predicting the mean for each sample

Investigating STTR in Mediated Texts with ANOVA – GitFront
- Project within the Language Data Analysis course of the MA in Translation and Technology, including research design, data collection and hypothesis testing to investigate lexical simplification in mediated texts
- Implemented a Python script to compute the standardized type-token ratio every 1,000 tokens, conducted statistical testing in R and wrote a report on the study setup, the analysis and the results

International Study

Petnica Research Center Petnica, Serbia
Summer School *July 2023*
- Researched argument order in Italian as part of a collaborative study on double object constructions, collecting corpus data from Sketch Engine, performing statistical testing and preparing a poster presentation of the results
- Courses: Language Variation: Investigating Double Object Constructions

Higher School of Economics Moscow, Russia
Overseas Exchange Programme *Sept 2019 - Jan 2020*
- Enhanced Russian and English language skills through courses in political debates and argumentative writing (Russian), and web writing and speech practice (English), improving both written and spoken communication
- Courses: Advanced Russian through Politics, Web Writing, Speech Practice

Other Experience

CLEF 2024 Shared Task Annotator Bologna, Italy
Annotated English and Italian news articles for a shared task on subjectivity detection *Jan 2024 - Mar 2024*
Terminologist at IN.TRA Forlì, Italy
Created and updated terminological resources for a not-for-profit language services agency *Feb 2023 - Jun 2023*
Subtitler at Concorto Film Festival Pontenure, Italy
Created subtitles and transcriptions for the Focus Russia section of Concorto Film Festival *May 2019 - Sept 2019*

Languages

Native	Italian
Advanced	English, Russian
Intermediate	French, Spanish
Basic	Japanese