

GIORGIA TROIANI

Corpus design specialist - Computational Linguist

✉ giorgiajtroiiani@gmail.com

🌐 github.com/gtroiani

📍 Madrid, Spain

🌐 /in/gtroiani

SUMMARY

- Linguistics research scientist with a strong background in design of spoken corpora for technological applications in low-resource language varieties leading to the creation of the first ever corpus of spoken Kazakh.
- Project manager with 4 years of experience supervising a team of >30 research assistants responsible for the methodological choices in the data collection and annotation processes and for the design of the training curricula of new assistants.
- Strong builder of interdisciplinary collaborations and effective science communicator that developed and maintained relationships with key stakeholders of the Kazakh corpus (academic institutions, general public, government).

EXPERIENCE

- 9/2023 – now **Postdoctoral Research Fellow** **Nazarbayev University**
- Project manager responsible for the achievement of publication goals for laboratory leading to the planning and execution of >10 research projects and dissemination of results to diverse audiences.
 - Supervisor for students' theses with experience in quality assessment of quantitative and qualitative research design
 - Instructor with strong interdisciplinary collaborations that developed 3 university-level courses aimed at developing linguistics skills for specialists in Natural Language Processing (corpus creation and annotation, bias reduction in data collection, quality evaluation of linguistic texts).
- 3/2022 – 8/2023 **Graduate Research Assistant** **Nazarbayev University**
- Design of data collection and annotation protocol for a corpus of spoken language
 - Training and management of >30 research assistants in annotation techniques
- 6/2020 – 9/2021 **Localization manager and Linguistic Data Consultant** **Rezonator**
- Management of a team of >20 translators that led to the localization of software into 10 language varieties
 - Linguist consultant to developing team focusing on data import and user interface for the development of a software to support annotation of conversational data
- 9/2017 – 6/2021 **Instructor** **University of California Santa Barbara**
- Skilled communicator of scientific concepts to non-specialists with experience in teaching a wide-ranging set of topics in linguistics
 - Expert trainer in different corpus annotation tools
- 7/2016 – 9/2016 **Videogame Localization Tester** **Electronic Arts**
- Detail-oriented analysis of linguistic material in products tailored for different platforms to ensure quality of user experience

SKILLS

Corpus creation	Project management	Python
Corpus annotation	Team management	NLTK
Qualitative research methods	Curriculum development and training	ELAN
Quantitative research methods	Conflict management	GameMaker Language
Digital resources for low-resource varieties	Public speaking	R

EDUCATION

- 9/2017 – 8/2023 **PhD in Linguistics** **UC Santa Barbara (USA)**
- Application of mixed quantitative and qualitative methods to a wide-ranging array of research projects (experimental design, qualitative surveys, statistical modeling, conversation analysis) as evidenced by doctoral dissertation.
 - Project and resource management and acquisition of knowledge of current industry trends and computational skills as demonstrated by ability to produce scientific publications and secure grants.
 - Acquisition of communication skills and ability to convey scientific concepts to a wide audience as evidenced by presentations at multiple academic conferences, extensive teaching experience, and public media outreach.

LANGUAGES

Italian - native, **Spanish/ English/ Russian** - fluent, **Kazakh** - intermediate, **Arabic** - beginner