



The TIGR corpus: collection, transcription, processing and management of videorecorded data for the study of Italian talk-in- interaction

The TIGR corpus of spoken Italian includes 23.5 hours of video-recorded and transcribed discourse and was gathered by a team of linguists at USI Università della Svizzera italiana in the Swiss cantons Ticino and Grisons in 2021-2022. It is currently being prepared for being shared via the Swiss FAIR-compliant (cf. Wilkinson et al. 2016) LaRS repository.

We describe the corpus design and some changes it underwent during fieldwork; the interaction with the speakers, including their informed consent to participate in the study; the set-up of audio and video recordings; sociolinguistic characteristics of the participants; transcription conventions and techniques; workflows of transcript processing in view of qualitative analysis and annotation; de-identification measures; scenarios of data reuse; the organization of the data and metadata on LaRS; an open science approach to problem solving during the preparation of the data to be shared; future perspectives.

06.11.2024 - h.15.00



<https://site.unibo.it/laboratorio-sperimentale/it/agenda/seminario-tigr>

**Prof. Johanna
Miecznikowski-Fuenfschilling,
Elena Battaglia**
Università della Svizzera Italiana

Studio 76, LILEC, via Cartoleria 5