

DIGITAL TOOLS AND CORPUS-BASED APPROACHES IN (ARABIC) SOCIOLINGUISTICS: METHODS, CHALLENGES, AND CROSS-DISCIPLINARY INSIGHTS

INTERNATIONAL ONLINE WORKSHOP

Department of Humanities (DISUM)
University of Catania

WORKSHOP PROGRAMME

→ 8:45 **Welcoming participants**
Rosa Pennisi (University of Catania): Introduction to the workshop

1ST SECTION (9:00 – 10:30)

→ 9:00 – 9:30 Veronika Laippala (University of Turku)
Tracing universals of register variation in masses of web data – a machine learning approach.

→ 9:30 – 10:00 May Rostom (Iremam, University of Aix-Marseille)
Methods, challenges, and strategies adopted to analyse and represent mixed Arabic repertoire employed in digital contexts: reflections on a case study.

→ 10:00 – 10:30 Rosa Pennisi (University of Catania)
From Posts to Podcasts: Corpus-Based Methods for Mixed (Moroccan) Arabic Across Written and Oral Data.

→ Pause (10:30-10:45)

2ND SECTION (10:45 – 12:15)

→ 10:45 – 11:15 Ibraam Abdelsayed (University for Foreigners of Siena)
Building the LAPE Corpus: Challenges in Transcription, Lemmatization, and Statistical Processing of Spoken Egyptian Arabic.

→ 11:15 – 11:45 Marco Venuti (University of Catania)
Exploring identity on 4chan: the view of the people on Migration.

→ 11:45 – 12:15 Antonino Andrea Belfiore & Ludovica Blangiforti (University of Catania)
Ideological lexicon and textual frequencies: A data-driven approach to comparing Arabic and Western Media.



Date
11 Feb 2026



Time
Start at 8:45 AM



Place
online

This workshop is part of the activities of the SABIRANET Project (CUP E63C24001920006), funded by the European Union – NextGenerationEU under Italy's National Recovery and Resilience Plan (PNRR), Young Researcher 2024 – SoE line, administered by the Italian Ministry of University and Research (MUR).

LINK TEAMS



<https://teams.microsoft.com/meet/39301106414845?p=X2ktWITNkYRyP8ISK>