

## **Which data for which kind of linguistics?**

### **The challenges of video recordings of naturally occurring social interactions**

Lorenza Mondada

The last decades have seen an explosion of data banks of linguistic corpora, both in quantitative and qualitative terms, with an increase not only of “big” data, but also of their diversity (regarding the languages and the situations represented). However, discussions about how to constitute repertoires of big data often heavily constrain the type of data collected – either because of technological limitations, issues of representativity, or ethical requirements. Beyond these constraints (but keeping them in mind in order to problematize them), this talk proposes a reflection discussing which kind of data we need for which kind of linguistics, or rather for which kind of conceptual and analytical approach to language. As a matter of fact, and beyond the fiction of data banks of corpora that could be universally used by any discipline of the linguistic community, the reflection elaborates on the specific requirements that specific analytical/epistemological projects might have, thus drawing a strong relation between methodology, definition of what constitutes relevant data, analytical framework and conceptualization of language and language use. The cases illustrating this reflection come from multimodal Conversation Analysis, which proposes a kind of linguistics grounded on language use by collectivities in interaction (vs isolated individual speakers), in naturalistic and diverse contexts (vs in constrained and experimental settings), as well as on the recognition that talking is done not only with the mouth but also with other parts of the body. In particular, the discussion will show the challenges of constituting adequate and relevant corpora of video data that adjust to the specificities of a diversity of contexts, as well as the consequences of these specificities for how we think about language(s) and language use.