

## **Studying talk-in-interaction: methodological and theoretical challenges when working with corpora**

*Marlou Rasenberg*

Working with spoken interaction data involves making countless small decisions—some about managing and organizing the data, and others that involve important theoretical commitments. Either way, these choices affect how the data can be used and the overall outcome of the research project. In this keynote, I will walk through the full arc of building a corpus of spoken interaction—from the early conceptualization stages through collection, processing, annotation, transcription, and finally sharing and publishing—drawing from extensive hands-on experience with multimodal, multi-speaker data. I will illustrate these steps with concrete examples, including complex recording setups, multi-device synchronization, and processing workflows using video editing and annotation software.

Rather than offering a one-size-fits-all approach, I will highlight decision points where practical constraints shape theoretical commitments (and vice versa). This includes issues of segmentation, the limits of turn-level transcription, and annotation tier design. I will discuss often overlooked challenges like inter-rater agreement and how file and tier structures affect long-term usability of a corpus. Finally, I will reflect on the evolving role of team science and the increasing need to embed robust documentation and reproducibility practices into every layer of the research process.