Bad data, incomplete data, and inappropriate data: A view from fieldwork
Keren Rice

In a recent book on fieldwork, Bowern (2015: 7) writes “So, after all that, what is optimal ‘fieldwork’? My definition is rather broad. It involves the collection of accurate data from language speakers in an ethical manner.” In this presentation, I examine what ‘accurate’ means with respect to data from fieldwork. I distinguish between bad data, incomplete data, and inappropriate data. Bad data can arise in many ways – poor transcription, misunderstanding of the meaning and thus mistranslation, conflation of data from different dialects. Incomplete data is probably inevitable, as the questions that one asks change over time, and the data available may not allow for a new question to be answered. Inappropriate data arises from the goals that one has. In order to ask about what speakers know about their language - competence - it is generally necessary to work with elicited data that allows for the study of, for instance, structural properties of the language. While such data is controlled and can be used to gain negative evidence, such data might be considered ‘bad’ in that it does not represent language use. If, on the other hand, the goal of the study is to examine language use - performance - then naturalistic language is favoured. Such data can be ‘bad’ in different ways. It is not controlled and much is missing from a corpus of natural speech. In the end, probably all people involved in linguistic fieldwork aim to avoid bad data in the first sense. Incomplete data is probably unavoidable. And inappropriate data may well be best understood as appropriate or not for the questions that are of interest to the particular researcher.