It is an illusion that data collection one the one hand and analysis of these data and theorisation based on them are separate endeavours. In fact, what counts as data, how one sets out to gather them, from whom, and what methods are employed to represent them determines how these data can be interpreted even before the first word is elicited (Hellwig, 2010; Lüpke & Storch, 2013; Lüpke, 2009). Therefore, reflections on data collection and analysis methods should be central to field linguistics – after all, we want to describe and document the undescribed, so we need to be particularly careful to ensure that our methods turn our data into keys rather than mislead us or superficially confirm a priori assumptions. Methodological awareness is a core concern in the social sciences and in sociolinguistics, where notions like the need for triangulation (Rothbauer, 2008) – confirmation of data and analysis through replication, repeated collection and analysis by different researchers and with different speakers and corroboration through the application of different methods for collection and analysis – and ways to circumvent observer’s paradox (Labov, 1982) – the fact that the presence of an observer can have unintended consequences on the event observed and hence skew the data – are central. Sociolinguists, social psychologists and anthropologists have developed differentiated notions surrounding ethnolinguistic identities, attitudes and communities of practice. For linguistics, Himmelmann (1998, 2012) suggests the triad of observed, staged and elicited communicative events on which I draw in my (2009) paper. Seifart (2008) offers an inspiring method of designing documentary corpora so that they include these events in a representative way, while Mc Laughlin & Sall (2001) and Dimmendaal (2001) remind us how arbitrarily we often chose our “language community” and our select our “consultants” and how often employed classifications into “good”, “bad” or “semi-speakers” only seemingly help us to create a homogenous and coherent data set by weeding out variation, because they are not applicable in the multilingual and polylectal field settings with nonstandardised languages that we encounter on the ground, where speakers easily can have two mother tongues, or none at all, and change their linguistic practices in the course of their lives.

In this talk, I set out to explore the intricate relationship between method and result from an Africanist perspective, and focussing on three focal areas: the language community, the speaker, and variation within the community and within and between speakers. Drawing on examples from my own field research on Bainounk languages in Senegal, I exemplify how different ways of collecting data have a huge impact on what one can uncover through them. At the same time, I suggest, again based on my own experiences, that some methods of data collection are more ecologically valid or enriched than others, and exemplify how this approach, despite its slow character, can be taken in externally funded projects with short duration. I will exemplify how a biased way of identifying “the community” can be avoided in dynamic multilingual settings, how observer’s paradox can influence the data collected and how this can be addressed, and how instead of creating unnecessary noise, variation constitutes an important data type in itself, if the metadata to interpret it are present.

References


