

British Sign Language Corpus Project: Open Access Archives and the Observer's Paradox

Adam Schembri
(UCL)

The British Sign Language Corpus Project is a new three-year project (2008-2010) that aims to create a machine-readable digital corpus of spontaneous and elicited British Sign Language (BSL) collected from deaf native and near-native signers across the United Kingdom. In the field of sign language studies, it represents a unique combination of methodology from variationist sociolinguistics and corpus linguistics. The project aims to conduct quantitative studies of sociolinguistic variation, language change and language contact simultaneously with the creation of a corpus. The recruitment of participants will be balanced for gender and age, with 240 signers filmed in 8 key regions across the UK. Furthermore, participant recruitment will rely on deaf community fieldworkers, using 'network sampling'. The data will be limited in terms of situational varieties, focusing mainly on conversational and interview data, together with narratives and some elicitation tasks. Unlike previous large-scale sociolinguistic projects, however, the dataset will be partly annotated and tagged using ELAN software, given metadata descriptions, and will be made accessible and searchable on-line. This means, however, that participants must consent to having the video data of their sign language use made public. This seems to put at risk the authenticity of the data collected, as signers may monitor their production more carefully than might otherwise occur. While recognising that the concept of the vernacular variety studied by sociolinguists represents an abstraction, we discuss the possibility of minimizing this special case of the observer's paradox by making some of the conversational data password protected for use by academic researchers only, while making other parts of the corpus publicly available as part of a dual access archive of BSL.