The Spoken BNC2014: design, compilation and...bad language

Robbie Love

Thursday, 30 November 2017

This lecture introduces the Spoken British National Corpus 2014 (Spoken BNC2014), an 11.5-million-word corpus of orthographically transcribed conversations among L1 speakers of British English from across the UK, recorded in the years 2012–2016 (Love et al. 2017). I discuss the compilation of the corpus and address some of the methodological issues we encountered. Next, I demonstrate how the Spoken BNC2014 may be used from a corpus-based sociolinguistic perspective by reporting on a project about bad language. Using McEnery (2005) as a basis, I discuss the sociolinguistic comparability of the corpus with the original Spoken BNC. I also outline some of the research projects being undertaken by recipients of the Spoken BNC2014 Early Access Data Grants (McEnery et al. 2017, Brezina et al. forthcoming), and reveal a few funny examples from the data with a live demonstration in CQPweb (Hardie 2012).

References


