

Public Corpora and Research on African American Language

Tyler Kendall

University of Oregon

Charlie Farrington

University of Oregon

Jason McLarty

University of Oregon

No variety of American English has received more attention by researchers, the media, or the public than that spoken by African Americans. African American Language (AAL) has long been a central object of study by sociolinguists and dialectologists (e.g., Wolfram 1969; Labov 1972; Fasold 1972; Rickford 1999; Poplack & Tagliamonte 2001; Green 2002; Wolfram & Thomas 2002; Lanehart 2015). Much is known about many structures of AAL varieties and a large body of research has investigated its origins and current trajectories of change. Yet, important questions remain about the origin of these varieties, their current and future development, and their relationship(s) to other varieties (regional and socio-ethnic). There also continue to be important social and educational applications of enhanced knowledge about AAL.

While AAL has been extensively studied, it remains massively underrepresented in terms of publicly available datasets. Almost all AAL data collected to-date are in private collections and often have not been collected in ways that allow for wider, public sharing (Kendall, Bresnan, & Van Herk 2011), creating a situation in which the most studied variety of American English has been inaccessible to larger-scale, computational and corpus-based study.

This paper announces the release of the first public corpus of spoken AAL. We discuss the benefits of public corpora on AAL and describe our project to create this corpus. The corpus, with a core component of sociolinguistic interviews with a cross-section of African Americans from Washington DC, makes available both legacy data from the canon of sociolinguistic research on AAL (Fasold 1972) as well as a series of interviews recorded in 2015-16. Additionally, we are incorporating recordings from other locales around the U.S. (e.g. Atlanta, GA, Rochester, NY, Princeville, NC). We hope to receive important input from participants at Methods that can help steer the direction of our ongoing work.