

Session Title

Appalachian Languages: Data Preservation

Roundtable Abstract

The Roundtable will feature speakers exploring techniques for the digitization, preservation, and increased accessibility of archived audio recordings from Appalachia. Some participants will be describing these techniques in connection with current data collection; some will include descriptions of the detailed cataloging of collections of hundreds or thousands of audio recordings. [Abstracts are included here and are also attached as .pdf files, as requested.]

Kentucky Phono Atlas Project

Irons, Terry Lynn

The KY Phono Atlas project is currently a website (<http://www.phonoatlasky.net>) that comprises two interactive web-based software tools for the analysis of acoustic phonetic data and a primary source archive of speech data collected in field interviews in each of the five regions of the commonwealth of Kentucky. The online archive of speech samples collected as part of the field work of the Phonological Atlas of Kentucky English project preserves a picture of non-urban speech in the commonwealth of Kentucky at the beginning of the twenty-first century. In that project, a study of variation and change in non-urban areas of Kentucky, field interviews were conducted in or near each of those Kentucky communities in which interviews were conducted for the Dictionary of American Regional English in the 1960s. To complete gaps in the sampling grid in the western areas of the state, additional interviews were conducted in five communities that were included in the Linguistic Atlas of the North Central States. Each subject participated in a sociolinguistic interview in which each was asked to provide demographic information, to discuss life, experiences and issues in that region, and to read word lists and a passage. All speech samples were recorded digitally, using GoldWave v4.25, a headset microphone, and a laptop computer. This online archive of 120 field interviews constitutes an important resource for current and future generations of scholars. The entry portal to the archive is an interactive area map that has links to tables of informants for specific counties/regions. In addition to informant information, the tables include links to downloadable zipped sound files. The sound files may be used freely for any research purposes, as long as credit is given to the Phono Atlas of Kentucky project.

Additional work is currently underway digitizing recorded interviews from the Appalachian Oral History Project. Initiated in 1970 as a collaborative effort between Alice Lloyd and Lees Junior College in Eastern Kentucky, the Appalachian Oral History project has grown to a collection of more than 3,000 recordings, yet accessibility is limited. As the digitizing efforts proceeds, selected materials will also be made available at the KY PhonoAtlas Web site.

Digital Conversion of Atlas Audio Tapes

Kretzschmar, William A., Jr.

The Linguistic Atlas of the Gulf States (LAGS) surveyed the everyday speech of the Southern states (Georgia, Tennessee, Florida, Alabama, Mississippi, Arkansas, Louisiana, and Texas) in a series of 914 audio-taped interviews in the late 1960s and 1970s. Each interview extends to about

6 hours in length, mostly directed conversation about the common circumstances of life (housing, food, customs, flora and fauna, etc.). This audio archive is thus an unparalleled resource for study not only of the language of the South, but for Southern culture more generally including Appalachia. The systematic LAGS tape archive amounts to approximately 5500 hours of sound recordings. The LAGS audio tapes and other surviving Atlas recordings are now forty years old, far past the life of the tape medium. Many tapes have been copied once since the original interview (most originals are held at Emory University or the University of Georgia, with copies at the University of Georgia and University of Chicago) but even the most recent copies were made in the late 1980s and the copies are also past the life of the tape medium. For the time being, but only for the immediate future, we can still recover the sound from the audio tapes. Unless the audio-tape sound recordings can be converted to digital recordings that can be stored and backed-up on computer, this priceless record of American speech and culture will certainly be lost. The University of Georgia has recently created a Research Computing resource that can make available the mass computer storage required to store the 20 Terabytes (= 20000 Gigabytes) of digital sound recordings that will be generated by complete digital conversion of these audio files. This resource makes it possible now to consider permanent storage of the sound recordings, along with Web access to the information needed to find, select, and analyze the recordings.

Work being performed on the audio files (nearly 200 interviews so far) includes 1) digital conversion of audio tape to disk in real time, 2) an audit of the sound file to make sure that the digital conversion is clean and successful, 3) a further audit of the sound file to remove speakers' personal information from versions of the digital file to be made available publicly and to provide a rough time index for conversational topics in the interview, and 4) creation of files for public use in two formats: A) a complete .wav file, to be distributed on CD; B) a set of compressed MP3 files, each of a few minutes' duration, for the whole interview, for distribution on CD and on the Web. The .wav files preserve as much sound quality as possible for users who want to study acoustic information in the interviews. The MP3 files retain excellent quality for listeners, and have the advantage of being distributable on the Web, just as MP3 music files are distributed. We can report significant technical problems in the process, including solutions for some but not all of them.

Digital Preservation and Access Issues in Southern Mountain Speech Recordings Puckett, Anita

The lexeme of "place" runs centrally throughout much of the oral discourse of rural Southern Appalachia. It exhibits multiple semantic and pragmatic functions, and central among them is to connect local residents to the physical landscape in which they organize much of their verbal interaction with others in their communities (Allen 1990, Puckett 2000). The trope of "place" therefore organizes much of the verbal interpretive frameworks that constitute individuals' and communities' identities and histories. Audio recordings of emic, or community-based, speech about the significance of these "places" have, however, focused on "telling about" community life for the most part instead of being representations of speech-in-action or speech as being constitutive of community cultural life. That is, speech occurring as people actually ask for things, make promises, learn (or teach) others in local life patterns, preach or instruct sacred matters, joke or tell stories, convey affection or disdain, among many other ways of "doing"

things with discourse has often been omitted from the digital recording, archiving, preserving, and accessing of speech. Instead, speech about daily life—reported and referential speech—has been preferred in many cases. This presentation argues for ways of creating digital speech archives that would address these omissions and contribute to the sustainability of place-based communities in the process. It suggests a digital library model that allows scholars studying language-in-use, internet professionals, and local residents to work together to create digital libraries and archives that promote the development of an integrated set of multi-formatted materials that incorporate speech at “work” into other text and image based materials about and in southern Appalachian communities. Possible innovative results that can occur from this model include (1) a more democratic model of digital libraries in which community “memory keepers” play a critical role in the creation and maintenance of the library; and (2) an integrated approach to the use of speech that permits an interdisciplinary use of sound data by not only linguists and sociolinguists, but also historians, educators, and community members who are engaged in sustaining or creating civic space and community vitality. The presentation also discusses how practical concerns regarding speech recording standards, personnel, data management, expansion, and storage; text coding; preservation; access; and intellectual property rights will be addressed in part by drawing on the resources of the South Atlantic Humanities Center ([//www.southatlanticcenter.org](http://www.southatlanticcenter.org)).

References:

- Allen, Barbara. 1990. The Genealogical Landscape and the Southern Sense of Place. In *Sense of Place: American Regional Cultures*, ed. Barbara Allen and Thomas Schlereth. Lexington: University Press of Kentucky. Pp. 152-163.
- Puckett, Anita. 2000. *Seldom Ask, Never Tell: Labor and Discourse in Appalachia*. Oxford Studies in Anthropological Linguistics 25. New York: Oxford University Press.

Data Preservation

Shuttlesworth, Rachel

While the efforts of linguists to collect, digitize, and study audio data are ongoing, libraries are approaching similar tasks with their own goals and challenges. This presentation offers a broad survey of efforts by and discussions among libraries and library-related organizations (e.g. Council on Library and Information Resources, Digital Library Federation, Coalition for Networked Information) to digitize, preserve, and provide access to audio recordings, including those from Appalachia. Technological, scholarly, and budgetary issues will be discussed, as well as the challenges and benefits of endeavors to coordinate academic and library scholars.