5

Engagement Through Data Management and Preservation: The North Carolina Language and Life Project and the Sociolinguistic Archive and Analysis Project

Tyler Kendall and Walt Wolfram

1 Introduction

In this chapter we consider the ways that enhanced data management and preservation practices improve our engagement processes and procedures. Data management and engagement are often considered separate professional enterprises with different goals and methods. Despite their disparate traditions, we argue that there is indeed an underlying intersection between these activities. We examine this possibility by demonstrating a model established for North Carolina, one of the most linguistically diverse states in the United States. For linguists, this natural linguistic diversity raises a number of empirical questions about the nature of language variation and change. At the same time, it also provides a captivating window

T. Kendall (✉)
University of Oregon, Eugene, OR, USA

W. Wolfram
North Carolina State University, Raleigh, NC 27695, USA

K.P. Corrigan, A. Mearns (eds.), *Creating and Digitizing Language Corpora*, DOI 10.1057/978-1-137-38645-8_5

for the public into the intersection of language and society, with social and educational implications for students, the citizens of North Carolina, and the wider public. For more than two decades, the North Carolina Language and Life Project (NCLLP) has strived to align the scientific study of language variation and change with the need for promoting public knowledge about language variation (Wolfram et al. 2008).

Documenting the linguistic landscape of North Carolina—or of any place—is an arduous and time-consuming task. The formal research we conduct, however, is often obscured in scholarly journals and in library archives, usually inaccessible to people who do not have graduate degrees in linguistics. But socially responsible professionals have now been challenged by positions that include the *principle of debt incurred* (Labov 1982) and the *principle of linguistic gratuity* (Wolfram 1993), and inspired by the edict 'if knowledge is worth having, it is worth sharing' (Cameron et al. 1992: 24). The NCLLP's work has taken numerous shapes, ranging from publicly oriented books and audio CDs (for example, Wolfram et al. 2002), to documentary films (for example, Hutcheson 2004; Rowe and Grimes 2006), to museum exhibits and public exhibitions (for example, Vaughn and Grimes 2006), and finally to public school education materials (Reaser and Wolfram 2007).

Along with an interest in and devotion to public engagement, the NCLLP has developed a sense of duty to the data which form the backbone of its sociolinguistic research and public engagement. Thus, over the past decade, the NCLLP has increasingly been engaged in issues of corpus development, audio preservation, and, what we can broadly term, speech data management. This intersection has led to a partnership with the North Carolina State University Libraries to develop the Sociolinguistic Archive and Analysis Project (SLAAP; Kendall 2007, 2008).[1] SLAAP has been designed as a web-based repository and software toolkit for managing and working with sociolinguistic recordings and their related data (transcripts, variable tabulations, research notes, and so forth). It was primarily conceived of as a tool for sociolinguistic researchers, seeking to provide better management and preservation options for scholars and to develop new tools and approaches for the analysis

[1] See http://ncslaap.lib.ncsu.edu.

of sociolinguistic data (Kendall 2007, 2008). In this regard, we believe that SLAAP has been, and continues to be, successful. It has an increasing user-base of researchers, houses a growing collection of data, and has aided in research projects from students' theses (for example, Kohn 2008) and dissertations (for example, Carter 2009) to published research (for example, Kendall 2013a; Mallinson and Kendall 2009; Thomas 2010; Kohn 2014) and textbooks (Thomas 2011). However, the archive and its tools have also been increasingly useful in the classroom and in other nonresearch-oriented ways and, as the project develops, its potential as a public resource becomes clearer. For example, SLAAP has enabled easier dissemination of information about dialect diversity to journalists and others, ranging from novelists and scriptwriters to dialect coaches in acting contexts.

We begin this chapter by outlining the features and design of the SLAAP website (Sect. 2). We next discuss some of the outreach projects conducted by the NCLLP (Sect. 3), and then focus in some detail on a recent book, *Talkin' Tar Heel: How our Voices Tell the Story of North Carolina* (Wolfram and Reaser 2014), which targets a public audience and makes extensive use of audio and video enhancements from our archived collection (Sect. 4). *Talkin' Tar Heel* embeds quick response codes (QRs) directly in its pages, incorporating new technologies into traditional, print formats. We end (in Sect. 5) by considering the link between our data management and outreach efforts.

2 The Sociolinguistic Archive and Analysis Project

The Sociolinguistic Archive and Analysis Project (SLAAP) began in earnest in 2005, originally as a digitization and preservation effort (Kendall 2007) seeking to move the NCLLP's extensive collection of audio tapes to digitized formats, to provide for future preservability and to make the recordings more accessible to researchers. Preserving and creating databases of the many recordings collected over its years of work has been an important concern for the NCLLP. The second author had amassed several large cabinets of cassette tapes over several decades of research,

both in North Carolina and from earlier in his career. These tapes were catalogued and accessible to students and colleagues through the Sociolinguistics Lab at North Carolina State University but, due to their status as analog, physical tapes in cabinets, they required local access, were hard to work with and, even worse, were degrading in quality as time passed. In its initial design, SLAAP was first envisioned as a resource specific to the NCLLP's materials (and, as indicated in Kendall 2007, it was originally titled NC SLAAP, denoting its focus on North Carolina). Over time, SLAAP has, however, grown to become a more broadly used speech data management system and recording archive. SLAAP increasingly seeks to provide a central repository for sociolinguistic recordings from outside the NCLLP and is adding large collections of non-NCLLP materials.

The specific goals behind SLAAP are multiple. On a practical level, as mentioned, the project seeks to digitize and preserve a large collection of interviews. It also aims to provide researchers with better access to and interfaces for their data through a variety of web-based features (see Kendall 2007). At a theoretical level, SLAAP questions and rethinks current linguistic and sociolinguistic conceptions of the nature of speech data, its representations, and the sorts of questions that can be asked of it (see Kendall 2008).

2.1 Design and Features of SLAAP[2]

SLAAP centers on a web-based archive and analytic toolset for sociolinguistic data collections, but simultaneously encompasses a broader effort to explore new approaches to storing, managing, and interacting with natural speech data. SLAAP looks to some extent like some of the other corpus development projects discussed in the recent literature (such as the ONZE Corpus discussed by Gordon et al. 2007 and the LANCHART database discussed by Gregersen 2009). However, SLAAP was developed to fill a gap in terms of sociolinguistic practice more than it sought to create a particular corpus (Kendall 2008, 2013b). In terms of Poplack's

[2] Parts of this section, including the specific examples (and screenshots from SLAAP), are based on Kendall (2013a).

(2007: xi) explanation of corpora design as oriented towards either *end-product* or *tool*, SLAAP was designed as a tool with no envisioned end-product. It is conceptualized as a *speech data management system* designed to house and organize an expanding collection of audio recordings. The archive continues to grow, as new recording collections are added to the database and new transcripts are developed. As of April 2015, the SLAAP digital archive contains over 4250 interviews and over 3700 hours of audio. Over 105 hours have associated time-aligned transcripts, making a transcript collection of over 1 million words. The projects housed in SLAAP are primarily from work conducted by the NCLLP in North Carolina—for example Ocracoke, NC (see Wolfram and Schilling-Estes 1995), Robeson County, NC (Wolfram and Dannenberg 1999), Hyde County, NC (Wolfram and Thomas 2002), Raleigh, NC (Dodsworth and Kohn 2012)—but also includes recordings from farther afield, for example, a study of the Northern–Midland boundary in Ohio (Thomas 2010) and research on Mexican American English in south Texas (Thomas 2016), as well as 'historical' recordings that form part of the sociolinguistic canon (for example, Fasold 1972; Wolfram and Christian 1976; Butters 1981). SLAAP also stores the large collection of longitudinal recordings of African American English collected by the Frank Porter Graham Project at the University of North Carolina (see Van Hofwegen and Wolfram 2010; Kohn and Farrington 2012; Kohn 2014). Most of the collections are (American) English recordings, but the archive also houses several collections of American Spanish, as well as collections from the Caribbean (for example, Reaser 2004; Myrick 2013) and even a collection of Burushaski recordings from Pakistan (Khan 2014). Finally, SLAAP increasingly includes collections from other current research groups, such as the recordings of the West Virginia Dialect Project (see, for example, Hazen 2008) and projects undertaken in Oregon (for example, McLarty et al. 2014). SLAAP is a member of OLAC, the Open Language Archives Community, and many of the collections it houses are catalogued and discoverable through the OLAC website.[3]

By incorporating all of these recordings into a centralized, digital repository, we have put into dialogue the many diverse collections of soci-

[3] See http://www.language-archives.org.

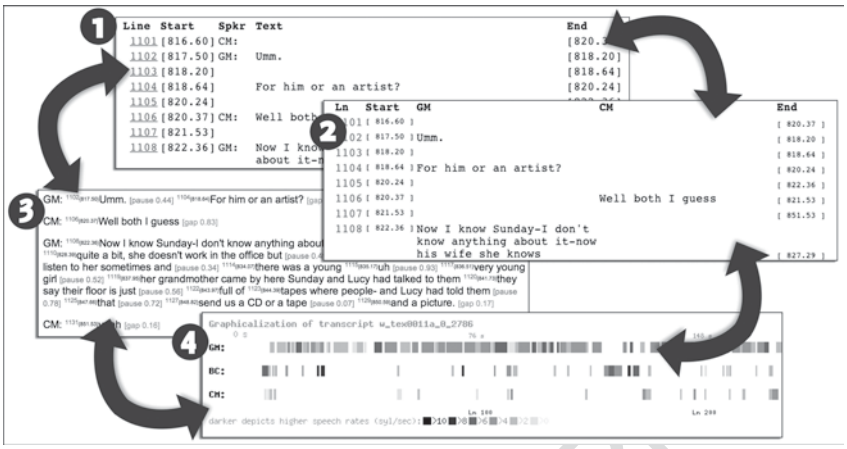

Fig. 5.1 Four presentations available in SLAAP of the same transcript data (from Kendall 2007)

olinguistic data. The descriptive metadata—the information stored about each interview, speaker, and research project—along with transcripts and researcher notes are all searchable both within and across projects. Older materials and metadata are just as easily retrieved as new materials, so tapes that formerly lay dormant for many years are now readily findable and can be put to new uses. For example, Kendall (2013a) mined transcripts and recordings from a range of projects contained in SLAAP to examine variability in speech timing features across several regional sites in the USA.

Since its beginning, we have tried to use SLAAP to explore ways that hypertext and other web technologies can enhance linguistic transcription and annotation more broadly. Our work on SLAAP has sought to make transcript information dynamic and flexible and linked to its source audio. Through SLAAP's software, the same transcript can be viewed in a *vertical format* (as in (1) in Fig. 5.1; Edwards 2001) or a *column-based format* (as in (2) in Fig. 5.1; Ochs 1979; Edwards 2001), or even in what is referred to in SLAAP as a *paragraph format* (as in (3) in Fig. 5.1). Alternatively, that same transcript can be transformed in various ways, such as into purely visual formats. The view shown in (4) of Fig. 5.1, called a *graphicalization* (Kendall 2007), displays speakers'

utterances within the complete interaction in a way that gives analysts a simple visual overview of the unfolding of the speech event. Each speaker's talk is displayed on its own tier. Shading indicates speech rate, with darker shading indicating faster speech, and pauses and speaker overlap are also accurately depicted. Analysts can generate graphicalizations with or without transcript text and can click on a passage to move to a page which provides analytic views of the transcript information (as discussed momentarily, and shown below in Fig. 5.3). Kendall (2013a) presents a second visual transformation of transcript information available in SLAAP, a so-called *Henderson graph*, which allows for the generation of quantitative metrics of a speaker's or interaction's hesitancy (see also Henderson et al. 1966; Thomas 2011: 186–7).

Transcript data in SLAAP are stored in database tables. Each transcript is a table in the database, and each line is an entry in the database table representing an utterance by a speaker. Transcripts for SLAAP are built using the TextGrid features of Praat (Boersma and Weenink 2015) to obtain highly accurate start- and end-times for each utterance. Each speaker is orthographically transcribed in his or her own TextGrid tier so that the temporal record accurately records the times of that specific speaker's contributions. The central unit of the transcript is the phonetic utterance—a stretch of speech bounded by pauses. Pauses are delimited separately from the speech.

Figure 5.2 displays the Praat editor window for the same transcript displayed in Fig. 5.1 above. This represents the 'source' transcript before it is added to SLAAP. The example shows three utterances for the interviewee GM (the full text for the third utterance is shown by Praat although the actual audio, waveform, and spectrogram run off-screen to the right). The second and third tiers house the transcriptions for the two interviewers BC and CM, although in the 8-second window shown only CM speaks, with a single utterance. The interval boundaries accurately capture the start- and end-times of each utterance and, in doing so, accurately delimit the pauses. In the Praat window shown, the 442 millisecond pause between GM's utterance 'Umm' And 'For him or an artist?' is selected. Kendall (2013a) describes the transcription model at greater length and discusses the ways that these time-aligned, databased transcripts can be mined for sociolinguistic research purposes.

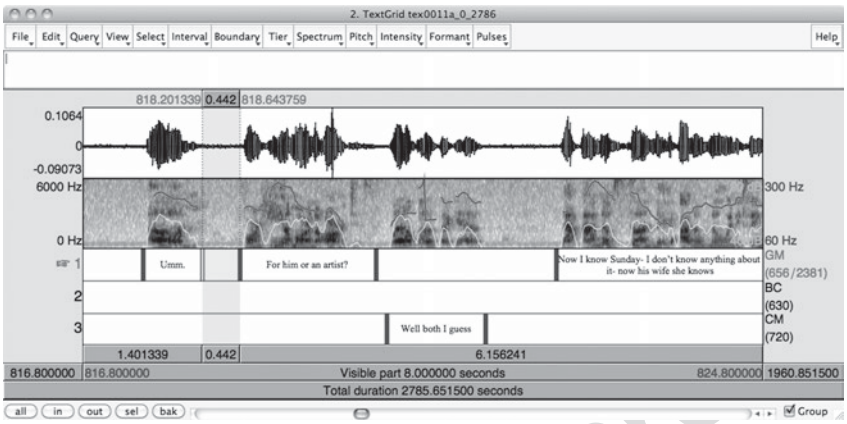

Fig. 5.2 Praat TextGrid for the transcript shown in Fig. 5.1

As this discussion illustrates, the fundamental components of SLAAP's databased transcript model are quite simple. In such a transcript model, the only data required for a complete transcription unit are: (a) a reference as to which speaker in the interaction is speaking, (b) the utterance's start-time, (c) an orthographic representation of the utterance, and (d) the utterance's end-time (Kendall 2007, 2013a). Through specially designed software like SLAAP, this very simple data model is quite powerful. SLAAP creates links between the transcript data and the audio file from which the transcript is based, and phonetic software (such as Praat in the case of SLAAP) can be integrated into the transcript interface software to allow for real-time phonetic analysis from within the transcript. With the start- and end-times for each utterance captured in the database and a linkage maintained with the audio, much of the other information that is often tagged or coded (for example latching, overlap, pause length, and so on) is unnecessary and can be reconstructed from the audio itself.

At the same time, an approximation of standard orthography (see Chafe 1993: 34; Tagliamonte 2007: 211–15) is often sufficient for the transcript text because pronunciation features (for example, vowel qualities, *r*-vocalization, and so on) can be listened for or examined instantly via a spectrogram. This simple orthography makes the transcripts easier to read than more complex systems. The use of standard orthography also

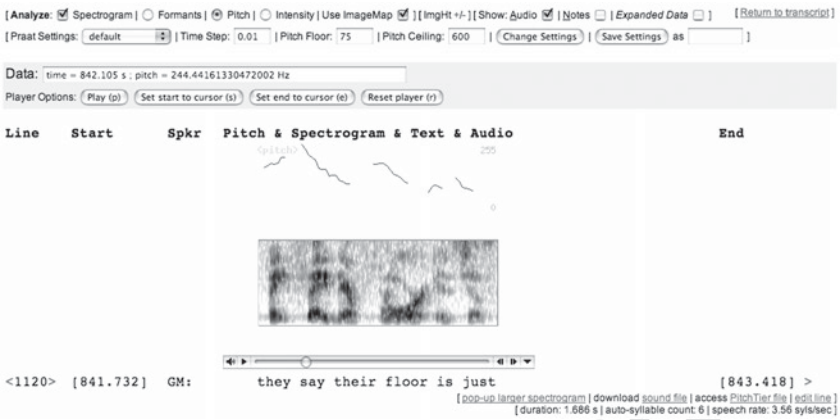

Fig. 5.3 SLAAP screenshot showing a transcript line with phonetic data

allows for easier searching and for more straightforward concordancing and other corpus-based extraction measures (Kendall 2013a). And, finally, the simple, orthographic transcription conventions and dynamic and searchable interfaces allow for flexible modes of access from a variety of different users, both linguistic analysts and nonlinguists who may be interested in the data for their content or for other purposes.

As an illustration of what can be done with this simple, databased transcript data, Fig. 5.3 shows a screenshot from the SLAAP software demonstrating an in-depth view of one transcript line. This example shows a pitch plot as well as a spectrogram, though other data views are available. Note also that the audio for the line can be listened to through an embedded audio player and that numerical data (in Fig. 5.3 this entails acoustic measurements of pitch) can be obtained at the click of the mouse. Additionally, multiple transcript lines can be displayed in this detailed format on the same page, allowing for easy comparison between utterances.

2.2 Libraries, Partnerships, and Databases

We were fortunate in the beginning phases of SLAAP to receive substantial support—moral, technical, and financial—from the Libraries at North Carolina State University (NCSU). The Libraries viewed the

preservation of the NCLLP materials as an important goal in their role as stewards of scholarly work and also viewed the project of archiving and databasing spoken language recordings as a valuable research project in library science (Kendall and French 2006; see also Smith et al. 2004). The Libraries' involvement in the project has been a crucial part of its success, and, consequently, we would like to pause briefly here to acknowledge more fully the support we have received from the NCSU Libraries. Our experiences suggest that university libraries and information technology groups are key partners in any archiving enterprise. As described in Kendall (2013b):

> A persistent issue in the long-term preservation and accessibility of research recordings is the problem of institutionalization, which presents a larger hurdle than the availability of specific tools and methods or any of the technical problems of data preservation. Many sociolinguistic data collections depend on their original collector to maintain them, and many researchers create impressive websites about their work and may even maintain their own data in a web-accessible format. However, these kinds of resources take extensive time (and cost) to maintain. Traditionally, these activities have not been evaluated as a part of researchers' academic 'credit' for advancement, so we are often, in fact, disincentivized to spend the extensive and sustained effort required to ensure that our materials are accessible to others and maintained in the long term. (p. 202)

It has become increasingly clear to us that archives and projects like SLAAP cannot succeed in the long term without broader institutional support.

3 The North Carolina Language and Life Project and its Outreach Endeavors

While the preceding sections detail the preservation and analysis potential of the SLAAP system, this phase of the project might exist autonomously and be limited to its research utility. However, it is also possible to use the resources of language diversity in such a data management system with efforts to represent and exemplify speech in an outreach program

connected to public education. The following sections show how the resources of SLAAP have enhanced and enabled an outreach program focused on public education in the US state of North Carolina.

The North Carolina Language and Life Project (NCLLP) was established at North Carolina State University in 1993 to form an umbrella program under which research and outreach could be linked. North Carolina is a convenient state for both research and engagement since it reflects a wide variety of regional and sociocultural English dialects as well as an assortment of ancestral and immigrant languages. North Carolina's language ecology is as diverse as its physical topography and climate—the most varied of any US state east of the Mississippi River. The research-outreach mission is integral to the goals of NCLLP: (a) to gather basic research information about language varieties in order to understand the nature of language variation and change; (b) to document language varieties in North Carolina and beyond as they reflect varied cultural traditions; (c) to provide information about language differences for public and educational interests; and (d) to use research material for the improvement of educational programs about language and culture.

A state-focused model for research and engagement was adopted for political, cultural and educational reasons. Though language diversity in the USA is typically immune to state political boundaries, political governance, social and cultural institutions, as well as educational programs and standards are bounded by the state. These range from museums, the North Carolina State Fair, and other cultural organizations to the statewide standards and objectives for the North Carolina Department of Public Instruction. Perhaps just as importantly, North Carolina is a state which takes considerable pride in its historical and cultural resources. As Wolfram and Reaser note (2014: 1), 'North Carolinians like their state a lot, and so do the 50 million yearly visitors', adding that 'North Carolina is not shy about marketing its resources as one of the top states for both living and vacationing'. In this social context, it seems advantageous to link language heritage to other cultural resources since 'language reflects where people come from, how they have developed and how they identify themselves regionally and socially' (Wolfram and Reaser 2014: 1). Although the NCLLP focuses on language variation within a political state, its goals extend further—to serve as a national model for integrating

community-based research and engagement. While a state moniker is beneficial for branding, the SLAAP archive (as discussed above), the NCLLP's research work, and the larger outreach efforts are not, of course, limited to the state of North Carolina.

3.1 Public Outreach Efforts

Since its inception, the NCLLP has been devoted to the intersection of research on language variation and change and public interests. Over the past 20 years, the NCLLP has produced 11 documentaries broadcasted on statewide, regional, and national television, constructed 6 museum exhibits, produced a half-dozen oral history audio CDs, developed a public school curriculum on language and dialect awareness, and written several trade books for popular audiences. In this section, we review some of these outreach projects, highlighting ways in which SLAAP, and our database more generally, have contributed to the efforts. Most recently, the NCLLP has published a publicly oriented book on language in North Carolina, *Talkin' Tar Heel* (Wolfram and Reaser 2014), which uses QRs to create an enhanced audiovisual experience for readers, despite its existence as a traditional print book. We focus on this most recent project in Sect. 4.

NCLLP video productions range from TV programs that have aired nationally, regionally, or on the state affiliate of the United States Public Broadcasting Service (PBS) to those produced primarily for community organizations, though a focus on local and broader-based audiences is not mutually exclusive. Examples of the former include: *Indian by Birth*: *The Lumbee Dialect* (Hutcheson 2001), *Mountain Talk* (Hutcheson 2004), *Voices of North Carolina* (Hutcheson 2005), *The Carolina Brogue* (Hutcheson 2008), *Spanish Voices* (Cullinan 2011), *First Language*: *The Race to Save Cherokee* (Cullinan and Hutcheson 2016), while the latter include: *The Ocracoke Brogue* (Blanton and Waters 1996), *Hyde Talk*: *The Language and Land of Hyde County* (Torbert 2002). Fortunately, high-quality video recording equipment and editing software that are portable, user friendly, and quite affordable are now available, making video documentary production quite feasible for students and faculty. Moreover,

based on the positive responses to our programs from the public, we have even convinced the administration at NCSU to fund a full-time videographer as a core component of our research-engagement initiative.

Invariably, passages from sociolinguistic recordings archived in SLAAP are used in our documentary productions along with the collection of high-quality video recordings for each documentary. Simultaneously, the audio from video recordings captured for documentary videos have been incorporated into SLAAP. One of the challenges for the future of SLAAP is to archive and manage the hundreds of hours of video footage now collected in the production of video documentaries. At the same time, it is noteworthy that some of the video footage collected primarily for documentary production has led to new research (for example, Wolfram et al. 2014), as we have found that outreach and engagement can lead to both basic and so-called 'engaged research' (Reaser 2006). The relationship between research and outreach can, and should be, bilateral and synergistic rather than one-dimensional—from research to outreach.

The compilation of oral histories on CDs is yet another way that we can share the diverse voices of communities where we have conducted sociolinguistic research. Based on sociolinguistic interviews that are archived in SLAAP, and with the assistance of community members, we have compiled a number of collections of stories that reminisce, celebrate, and entertain. For example, the NCLLP staff partnered with the Ocracoke Preservation Society to produce two such compilations a decade apart, *Ocracoke Speaks* (Childs et al. 2001) and *Ocracoke Still Speaks* (Reaser et al. 2011). Staff extracted recorded stories from the archived material on SLAAP and reviewed them with community members from the Preservation Society so that the final passages for the compilation could be selected together. A similar project, *Voices of Texana*, in Texana, North Carolina, was compiled by Christine Mallinson and Becky Childs with community members in a small, isolated African American community in the Smoky Mountains (Mallinson et al. 2006), based on the archived sociolinguistic interviews in SLAAP.

The museum exhibit is another productive venue for the use of archived audio recordings from SLAAP. With community-based preservation societies and museums, it is possible to construct permanent exhibits that highlight language diversity, as well as limited-time exhibits

on history and culture, which include the voices of community residents. An exhibit titled *Freedom's Voice*: *Celebrating the Black Experience on the Outer Banks* (Vaughn and Grimes 2006) included images, a documentary, interactive audiovisuals, artifacts, and audio clips first recorded for sociolinguistic interviews and reappropriated oral histories to complement informational panels that highlighted African Americans' involvement in the history of coastal North Carolina. This exhibition, which ran for over a year, brought together history, culture, and language through narrating the story of the previously overlooked contributions of African Americans on coastal Carolina, particularly on Roanoke Island, the site of the 'Lost Colony'.

One of the most successful exhibits for the NCLLP is an annual booth at the North Carolina State Fair that has a yearly attendance exceeding 1 million people. Video vignettes and free dialect badges or 'buttons' as they are known in the USA (with phrases like 'bless your heart', 'dingbatter', 'I speak North Cackalacky', 'sigogglin') are always very popular with attendees. In addition, an interactive touch screen monitor allows visitors to guess the regional voices of speakers from the archival recordings. We even had attendees' pronounce different place names and added them to SLAAP for future research use. In fact, one of the most popular QRs in Wolfram and Reaser's book (2014: 11; see Sect. 4) is based on speakers recorded at the State Fair who correctly and incorrectly pronounced the names of different place names in North Carolina that symbolized insider–outsider status.

3.2 Public Education

One of the most ambitious and essential outreach programs involves the development of formal curricular materials on language diversity in the public schools. Unfortunately, formal education about dialect variation is still relatively novel and somewhat controversial, and school-based programs have still not progressed beyond a pilot stage (Reaser 2006; Sweetland 2006). The examination of dialect differences offers great potential for students to investigate the interrelation between linguistic and social diversity, including diversity grounded in geography, history, and cultural beliefs and practices.

The dialect awareness curriculum developed by Reaser and Wolfram (2007), the first program endorsed by a state Department of Public Instruction, fits in with the standard course of study for the state of North Carolina's eighth grade social studies curriculum.[4] This language and dialect awareness program aligns with the curricular themes of 'cultures and diversity', 'historic perspectives', and 'geographical relationships' as they relate to North Carolina. In addition, the dialect awareness curriculum helps fulfill social studies competency goals such as 'Describe the roles and contributions of diverse groups, such as American Indians, African Americans, European immigrants, landed gentry, tradesmen, and small farmers to everyday life in colonial North Carolina' (Competency Goal 1.07) or 'Assess the importance of regional diversity on the development of economic, social, and political institutions in North Carolina' (Competency Goal 8.04). Students are not the only ones who profit from the study of dialect diversity. Teachers also find that some of their stereotypes about languages are challenged and that they too become more knowledgeable and informed about dialect diversity. Many of the activities are extracted from the SLAAP archive of audio recordings and video footage collected by the NCLLP. In fact, these audio and video vignettes are integral to the majority of activities as students listen to and watch clips on regional, social, and ethnic varieties integral to their workbook activities.

More information about the public engagement efforts of the NCLLP are detailed in Wolfram et al. (2008) and Wolfram (2011, 2013).

4 Digitally Reinforced Print Media: Talkin' Tar Heel

Writing about sociolinguistics for nonspecialized audiences is often difficult for linguists, and few seem to have creative talent for writing general books and articles for broad-based audiences. Tannen (1990, 2006) and Rickford and Rickford (2000) are notable exceptions. With varying degrees of success, the staff of NCLLP have authored several trade books aimed at these nonlinguistic audiences. For example, Wolfram and

[4] See http://www.ncpublicschools.org/curriculum/socialstudies/scos.

Schilling-Estes' *Hoi Toide on the Outer Banks*: *The Story of the Ocracoke Brogue* (1997) is aimed at tourists and Wolfram et al.'s *Fine in the World*: *Lumbee Language in Time and Place* (2002) is useful for residents of the community, educators, and others curious about the language variety of the Lumbee American Indians, the largest group of American Indians East of the Mississippi River (population *c.*55,000) and the largest non-reservation American Indian group in the USA. However, so-called 'science accommodation'—translating highly specialized technical knowledge about science into accessible descriptions for lay people—has proven to be a rhetorical and discursive challenge for those attempting to extend their descriptions beyond the academy, including the current authors. Typically, sociolinguists have used genres readily at their disposal, such as journal articles, textbooks, invited media interviews, and occasional editorial opinion articles in newspaper and popular magazines to present their perspective.

Linguists have highly specialized analytical skills and metalanguage in the subject matter, and they need not be apologetic about this expertise. By the same token, their own attitudes associated with this knowledge can potentially lead to disrespect for and the dismissal of lay-based observations and understanding about language, and they need to be sensitive to the perspectives and ideologies of nonlinguists, including attitudes about language by highly trained colleagues in other social science and scientific fields. As Sally Johnson (2001: 592) notes, 'scientists themselves have much to learn from the reception of their ideas by those outside their area of expertise'.

Wolfram and Reaser's recent book, *Talkin' Tar Heel* (2014), represents our most recent attempt to write a book for non-linguists by focusing on language diversity within a political state. The authors note that 'after more than two decades of interviewing and recording thousands of residents and shooting hundreds of hours of video footage, we feel that we would be remiss if we did not share the rich assortment of North Carolina voices with a broader audience'. Walt Wolfram's admitted goal was 'to have my wife, Marge, read this book, listen to the audios, and see the people for herself'. Marge Wolfram, a nonlinguist, is an avid reader but finds linguistics books boring. At the same time, much like many nonlinguists, she finds language differences curious and captivating in

her everyday interactions with people. The challenge, then, is to capture and present the inherent language intrigue that language differences hold for the general public without drowning it in metalinguistic jargon.

An added feature to the print is more than 130 audiovisual enhancements. Readers get to experience the language of North Carolina as they read through the extensive use of Quick Response Codes (QRs). In Fig. 5.4,

FIGURE 4.4. Author Martha Pearl Villas discusses the rapid changes she has seen in Charlotte. (Photograph by Neal Hutcheson)

Fig. 5.4 Example of QR use with printed text

we include a copy of a page with quotes from a long-time, elderly resident of the city of Charlotte about traditional attitudes towards Northerners who move to the South. Instead of simply reading the quote in orthographically altered script, however, readers can scan the QR using any portable Internet device, like a smartphone or tablet, to go directly to the quote and see and hear the speaker for themselves. Many of the audio and video enhancements are extracted from our archive of audio and video footage to allow the voices and people to speak directly to the reader.

Though the integration of QRs is somewhat novel in the field of linguistics—*Talkin' Tar Heel* is the first book published by our publisher to use QRs and, to our knowledge the first book making extensive use of this format in linguistics—the clips seem as natural as language diversity itself. We want the general reader to experience language and dialect rather than imagine it, and accessing the enhancements is easy. Each enhancement has a brief description so that the reader knows what he or she might hear or see. All the reader has to do is navigate any web browser to the provided URL, or use a smartphone or any device with a QR reader to snap a picture of the QR code to access the media directly. In the enhanced e-Reader version of the book, these enhancements are embedded within the text itself so a reader simply has to click on the icon. The website and Chapter Media page organization for intuitive navigation of the media clips on the website are displayed in Fig. 5.5.

A number of outreach opportunities have derived from the publication of the book. For example, in the months following the book's publication, Wolfram and Reaser gave more than 20 book readings at independent and chain bookstores around the state where they invariably demonstrated the integration of the audiovisual QR enhancements in the book. They have also been invited to speak at numerous civic organizations, preservation societies, and other state- and community-based events, and conducted television and radio interviews where they demonstrated how the audiovisual examples are integrated into the print text. These opportunities have extended outreach in ways that even they did not imagine. One customer was so enthusiastic about the potential of the use of QRs for public school education that he purchased $2500 in books for teachers and contributed $5000 to develop education materials related to the book for middle-school students. Responses such as these have exceeded our expectations

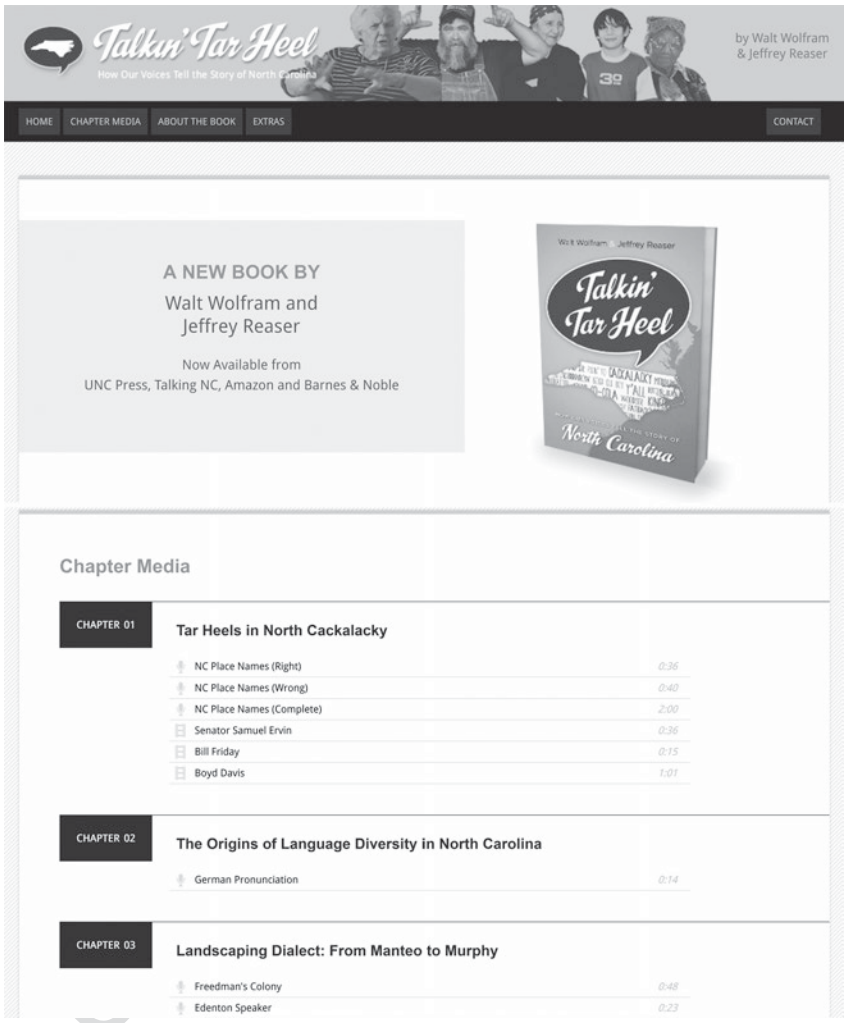

Fig. 5.5 Website and Chapter Media for *Talkin' Tar Heel*

in terms of the potential for research and outreach to enable each other. Furthermore, they underscore the potential for interdisciplinarity that includes not only the collaboration of academic fields but technical and professional collaboration as well. In our outreach efforts, we work with designers, producers, programmers, marketing experts, and other professionals who enable both research and engagement.

5 Conclusion: Engagement and Data Management

On the face of it, the connection between the NCLLP's engagement work on the one hand and the archival and data management work of SLAAP may seem like a contrived, tenuous one. It is certainly true that the primary connection is through the data, as both endeavors pivot on the core data. The data that form the backbone and *raison d'être* for SLAAP are necessary for any empirically based research effort, but they are also critical for dialect education and outreach efforts. In fact, the two sets of activities focused on in this chapter are just two points of a troika, with the third being the center of interest for most academic linguists—the study of language (and language variation and change) itself. While data management was not discussed in Wolfram's (1993) original formulation of the principle of linguistic gratuity, which stated that 'investigators who have obtained linguistic data from members of a speech community should actively pursue positive ways in which they can return linguistic favors to the community', or much of the other early literature on research ethics (Labov 1982; Cameron et al. 1992) it is nonetheless the case that being good stewards of the data we collect is a way to 'return linguistic favors' to the communities we study. It is indeed possible to enrich both the communities from whom we obtain our data and to preserve that data for the study of and celebration of language variation.

References

Books and Articles

Blanton, Phyllis and Karen Waters, producers. 1996. *The Ocracoke Brogue*. Raleigh, NC: North Carolina Language and Life Project.

Butters, Ronald R. 1981. Unstressed vowels in Appalachian English. *American Speech* 56(2): 104–110.

Cameron, Deborah, Elizabeth Frazer, Harvey Penelope, M.B.H. Rampton, and Kay Richardson. 1992. *Researching Language: Issues of Power and Method*. London/New York: Routledge.

Carter, Phillip. 2009. *Speaking Subjects*: *Language*, *Subject Formation*, *and the Crisis of Identity*. Doctoral Dissertation. Duke University.

Chafe, Wallace. 1993. Prosodic and functional units of Language. In *Talking Data: Transcription and Coding in Discourse Research*, eds. Jane Edwards, and Martin D. Lampert, 33–43. Hillsdale, NJ: Lawrence Erlbaum.

Childs, Becky, Walt Wolfram and Ellen Fulcher Cloud, producers. 2001. *Ocracoke Speaks*. CD. Raleigh, NC: North Carolina Language and Life Project.

Cullinan, Danica, producer. 2011. *Spanish Voices*. Raleigh, NC: North Carolina Language and Life Project.

Cullinan, Danica and Neal Hutcheson, producers. 2016. *First Language*: *The Race to Save Cherokee*. Raleigh, NC: North Carolina Language and Life Project.

[AU1]

Dodsworth, Robin, and Mary Kohn. 2012. Urban rejection of the vernacular: The SVS undone. *Language Variation and Change* 24(2): 221–245.

Edwards, Jane. 2001. The transcription of discourse. In *Handbook of Discourse Analysis*, eds. Deborah Tannen, Deborah Schiffrin, and Heidi Hamilton, 321–348. Malden, MA/Oxford: Blackwell.

Fasold, Ralph W. 1972. *Tense Marking in Black English: A Linguistic and Social Analysis*. Washington, DC: Center for Applied Linguistics.

Gordon, Elizabeth, Margaret Maclagan and Jennifer Hay. 2007. The ONZE Corpus. In *Creating and Digitizing Language Corpora*. *Volume 2*: *Diachronic Databases*, Joan C. Beal, Karen P. Corrigan and Hermann L. Moisl (eds), 82–104. New York/Basingstoke, Hampshire: Palgrave Macmillan.

Gregersen, Frans. 2009. The data and design of the LANCHART study. *Acta Linguistica Hafniensia* 41: 3–29.

Hazen, Kirk. 2008. (ING): A vernacular baseline for English in Appalachia. *American Speech* 83(2): 116–140.

Henderson, Alan, Frieda Goldman-Eisler, and Andrew Skarbek. 1966. Sequential temporal patterns in spontaneous speech. *Language and Speech* 9: 207–216.

Hutcheson, Neal, producer. 2001. *Indian by Birth*: *The Lumbee Dialect*. Raleigh, NC: North Carolina Language and Life Project.

Hutcheson, Neal. producer. 2004. *Mountain Talk.* Raleigh, NC: North Carolina Language and Life Project.

———. 2005. *Voices of North Carolina*. Raleigh, NC: North Carolina Language and Life Project.

Hutcheson, Neal. 2008. *The Carolina Brogue*. Raleigh, NC: North Carolina Language and Life Project.

603 Johnson, Sally. 2001. Who's misunderstanding whom? Sociolinguistics debate
604 and the media. *Journal of Sociolinguistics* 5: 591–610.
605 Kendall, Tyler. 2007. The Sociolinguistic Archive and Analysis Project: empow-
606 ering the sociolinguistic archive. *Penn Working Papers in Linguistics* 13(2):
607 15–26.
608 ———. 2008. On the history and future of sociolinguistic data. *Language and*
609 *Linguistics Compass* 2(2): 332–351.
610 ———. 2013a. *Speech Rate*, *Pause*, *and Sociolinguistic Variation*: *Studies in*
611 *Corpus Sociophonetics*. In *New York/Basingstoke*. Hampshire: Palgrave
612 Macmillan.
613 ———. 2013b. Data preservation and access. In *Data Collection in*
614 *Sociolinguistics*: *Methods and Applications*, Christine Mallinson, Becky Childs
615 and Gerard Van Herk (eds), 195–205. New York: Routledge.
616 Kendall, Tyler and Amanda French. 2006. Digital audio archives, computer-
617 enhanced transcripts, and new methods in sociolinguistic analysis. Paper pre-
618 sented at Digital Humanities (ALLC/ACH) 2006. Paris, France. July.
619 Khan, Samina. 2014. *Burushaski Recordings*. Raleigh, NC: North Carolina
620 Language and Life Project.
621 Kohn, Mary. 2008. *Latino English in North Carolina*: *A Comparison of Emerging*
622 *Communities.* Masters Thesis. North Carolina State University.
623 ———. 2014. "*The Way I Communicate Changes but How I Speak Don't*: *A*
624 *Longitudinal Perspective on Adolescent Language Variation and Change.*
625 Publication of the American Dialect Society 99. Durham: Duke University
626 Press.
627 Kohn, Mary, and Charlie Farrington. 2012. Speaker normalization: evidence
628 from longitudinal child data. *Journal of the Acoustical Society of America*
629 131(3): 2237–2248.
630 Labov, William. 1982. Objectivity and commitment in linguistic science.
631 *Language in Society* 11: 165–201.
632 Mallinson, Christine, Becky Childs, and Zula Cox. 2006. *Voices of Texana*.
633 Texana, NC: Texana Committee on Community History and Preservation.
634 Mallinson, Christine, and Tyler Kendall. 2009. "The way I can speak for
635 myself": the social and linguistic context of counseling interviews with
636 African American adolescent girls in Washington, DC. In *African American*
637 *Women's Language*, ed. S.L. Lanehart, 110–126. Newcastle upon Tyne:
638 Cambridge Scholars Press.
639 McLarty, Jason, Charlie Farrington, and Tyler Kendall. 2014. Perhaps we used
640 to but we don't anymore: the habitual past in Oregonian English. *Penn*
641 *Working Papers in Linguistics* 20(2): 111–119.

Myrick, Caroline. 2013. Big-time variation in small-time communities: sociolinguistic diversity on Saba. Paper presented at the Southeastern Conference on Linguistics (SECOL) 80. Spartanburg, SC.

NC [North Carolina] Standard Course of Study. Social Studies 2006: Eighth grade North Carolina: creation and development of the state. Online document. http://www.ncpublicschools.org/curriculum/socialstudies/scos/2003-04/050eighthgrade (accessed 22 May 2014).

Ochs, Elinor. 1979. Transcription as theory. In *Developmental Pragmatics*, eds. Elinor Ochs, and Bambi Schieffelin, 43–72. New York: Academic Press.

Poplack, Shana. 2007. Foreward. In *Creating and Digitizing Language Corpora. Volume 1*: *Synchronic Databases*, Joan C. Beal, Karen P. Corrigan and Hermann L. Moisl (eds), ix–xiii. New York/Basingstoke, Hampshire: Palgrave Macmillan.

Reaser, Jeffrey. 2004. A quantitative sociolinguistic analysis of Bahamian copula absence: morphosyntactic evidence from Abaco Island, the Bahamas. *Journal of Pidgin and Creole Languages* 19: 1–40.

———. 2006. *The Effect of Dialect Awareness on Adolescent Knowledge and Attitudes*. Durham: Duke University dissertation.

Reaser, Jeffrey, and Walt Wolfram. 2007. *Voices of North Carolina: Language and Life from the Atlantic to the Appalachians*. Raleigh, NC: North Carolina Language and Life Project.

Reaser, Jeffrey, Paula Dickerson Bunn, Walt Wolfram, DeAnna Locke, Chester Lynn, and Phillip Howard. 2011. *Ocracoke Still Speaks: Reflections Past and Present*. Raleigh, NC: North Carolina Language and Life Project.

Rickford, John R., and Russell J. Rickford. 2000. *Spoken Soul: The Story of Black English*. New York: John Wiley & Sons.

Rowe, Ryan and Drew Grimes, producers. 2006. *This Side of the River*. Raleigh, NC: North Carolina Language and Life Project.

Smith, Abby, David Allen, and Karen Allen. 2004. *Survey of the State of Audio Collections in Academic Libraries*. Washington, DC: Council on Library and Information Resources.

Sweetland, Julie. 2006. *Teaching Writing in the African American Classroom: A Sociolinguistic Approach*. Palo Alto, CA: Stanford University dissertation.

Tagliamonte, Sali. 2007. Representing real language: consistency, trade-offs and thinking ahead! In *Creating and Digitizing Language Corpora. Volume 1*: *Synchronic Databases*, Joan C. Beal, Karen P. Corrigan, and Hermann L. Moisl (eds), 205–40. New York/Basingstoke, Hampshire: Palgrave Macmillan.

Tannen, Deborah. 1990. *You Just Don't Understand: Women and Men in Conversation*. New York: Ballantine.

———. 2006. *You're Wearing That? Understanding Mothers and Daughters in Conversation*. New York: Random House.
Thomas, Erik R. 2010. A longitudinal analysis of the durability of the Northern-Midland dialect boundary in Ohio. *American Speech* 85(4): 375–430.
——— 2011. *Sociophonetics*: *An Introduction*. New York/Basingstoke, Hampshire: Palgrave Macmillan.
——— 2016. *Mexican American English*: *Substrate Influence and the Birth of an Ethnolect*.
Torbert, Benjamin, producer. 2002. *Hyde Talk*: *The Language and Land of Hyde County*, *NC.* Raleigh, NC: North Carolina Language and Life Project.
Van Hofwegen, Janneke, and Walt Wolfram. 2010. Coming of age in African American English: a longitudinal study. *Journal of Sociolinguistics* 14(4): 427–455.
Vaughn, Charlotte, and Drew Grimes. 2006. *Freedom's Voice*: *Celebrating the Black Experience on the Outer Banks*. In *Museum Exhibit*. Manteo, NC: Outer Banks History Center and North Carolina Language and Life Project.
Wolfram, Walt. 1993. Ethical considerations in language awareness programs. *Issues in Applied Linguistics* 4: 225–255.
———. 2011. In the profession: connecting with the public. *Journal of English Linguistics* 0(1): 1–7.
———. 2013. Community commitment and responsibility. In *The Handbook of Language Variation and Change*, 2 edn, eds. J.K. Chambers, and N. Schilling, 557–576. Malden: Wiley-Blackwell.
Wolfram, Walt, and Donna Christian. 1976. *Appalachian Speech*. Washington, DC: Center for Applied Linguistics.
Wolfram, Walt, and Clare Dannenberg. 1999. Dialect identity in a tri-ethnic context: the case of Lumbee American Indian English. *English World Wide* 20: 179–216.
Wolfram, Walt, Clare Dannenberg, Stanley Knick, and Linda Oxendine. 2002. *Fine in the World: Lumbee Language in Time and Place*. Raleigh, NC: North Carolina State University.
Wolfram, Walt, Jaclyn Daugherty, and Danica Cullinan. 2014. On the (in) significance of English language variation: Cherokee and Lumbee English in comparative perspective. *Penn Working Papers in Linguistics* 20(2): 199–208.
Wolfram, Walt, and Jeffrey Reaser. 2014. *Talkin' Tar Heel: How Our Voices Tell the Story of North Carolina*. Chapel Hill, NC: University of North Carolina Press.
Wolfram, Walt, Jeffrey Reaser, and Charlotte Vaughn. 2008. Operationalizing linguistic gratuity: from principle to practice. *Language and Linguistics Compass* 2(6): 1109–1134.

Wolfram, Walt, and Natalie Schilling-Estes. 1995. Moribund dialects and the endangerment canon: the case of the Ocracoke brogue. *Language* 71: 696–721.

———. 1997. *Hoi Toide on the Sound Side: The Story of the Ocracoke Brogue*. Chapel Hill, NC: University of North Carolina Press.

Wolfram, Walt and Erik R. Thomas. 2002. *The Development of African American English.* Oxford, UK/Malden, MA: Blackwell.

Websites, Software and Online Resources

Boersma, Paul and David Weenink. 2015. *Praat*: *doing phonetics by computer*. Version 5.4.15. http://www.praat.org (accessed 9 August 2015).

North Carolina Language and Life Project (NCLLP). http://www.ncsu.edu/linguistics/ncllp (accessed 17 April 2015).

Open Language Archives Community (OLAC) http://www.language-archives.org (accessed 17 April 2015).

Sociolinguistic Archive and Analysis Project (SLAAP). http://slaap.lib.ncsu.edu (accessed 17 April 2015).

Talkin' Tar Heel: *How Our Voices Tell the Story of North Carolina*. http://www.talkintarheel.com (accessed 17 April 2015).

Author Queries

Chapter No.: 5 0002695180

Queries	Details Required	Author's Response
AU1	Please update the publishing year for reference cullinan "forthcoming".	
AU2	Please update Published year, publisher name and location reference "Thomas in progress".	
AU3	please check vol no	