

NEW PERSPECTIVES ON AFRICAN AMERICAN LANGUAGE THROUGH PUBLIC CORPORA: INTRODUCTION TO SPECIAL ISSUE

TYLER KENDALL

University of Oregon

IN JANUARY 2018, the initial release of the Corpus of Regional African American Language (CORAAAL) (Kendall and Farrington 2018b) became available to the public.¹ CORAAAL is the first public spoken corpus of African American Language and one of the few publicly available large-scale sociolinguistic data sets. With this special issue of *American Speech*, we more fully unveil the corpus to the research community and celebrate its release.

During the period of its development, from 2015 to 2018, a number of scholars participated in the building of CORAAAL, several as “beta testers” who worked with early versions of the data and provided feedback to the corpus creation team as the data were being collected and annotated. This issue brings together some (but not all) of the scholars who took part in the process to showcase their early work with the corpus and to provide the research community with some initial perspectives the corpus may offer on African American Language (AAL) and for sociolinguistic research more generally. The authors in this issue bring new perspectives to the study of AAL, but they do this in a novel way in sociolinguistic research traditions, by each examining THE SAME DATA from different perspectives. Each of the articles in this issue takes as its starting place data available in the new public corpus and uses those data to ask a combination of new and old questions.

CORPUS OF REGIONAL AFRICAN AMERICAN LANGUAGE

CORAAAL is a long-term corpus-building project conceived of in terms of several components. The core components of CORAAAL—and the papers in this issue—focus on AAL in Washington, D.C., the nation’s capital, a city with a long-standing African American majority, and the site of much early research on AAL (see Farrington and Schilling 2019 [this issue]). These components, CORAAAL:DC, contain over 100 sociolinguistic interviews with AAL speakers in Washington, D.C., who were born between 1890 and

2005. CORAAL:DC consists of two subcomponents: CORAAL:DCA (Kendall, Fasold, et al. 2018) and CORAAL:DCB (Kendall, Quartey, et al. 2018). DCA includes a set of speakers interviewed between March 1968 and August 1969 as a part of Ralph Fasold's foundational study of AAL in Washington, D.C. (Fasold 1972, 2019 [this issue]). DCB includes speakers interviewed primarily by Minnie Quartey beginning July 2015, specifically for CORAAL.

Both DC subcomponents are developed following common sociolinguistic practice (see Schilling 2013), with a sampling matrix involving 4×3 cells, with four age groups (under 19, 20–29, 30–50, 51 and over) and three social class groups (labeled in the corpus groups 1–3 and meant to capture broad social strata, ranging from working class to upper-middle class). For each cell, the corpus attempts to contain equal numbers of females and males, although as several authors in this issue note, the actual number of speakers available varies to some extent across the demographic matrix. For DCA, this is a function of what recordings were collected during the original fieldwork and of the fact that Fasold's team's sampling goals were not the same as CORAAL's. For DCB, this reflects the complexity of recruitment in "the field" and the fact that CORAAL's DCB component is still being developed.²

All CORAAL recordings have been anonymized and orthographically transcribed with time-alignment at the utterance level (transcripts are available in formats for several common programs). CORAAL can be downloaded, for free under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, from its website at <https://oraal.uoregon.edu/coraal/> and can be accessed online at <http://lingtools.uoregon.edu/coraal/explorer/>. As a corpus developed with a focus on sociolinguistic interests, metadata are included in the corpus covering a wider range of information about the speakers than are typically available in public corpora. Several of the articles in this issue provide more specific discussions of aspects of the corpus. Readers are referred to the *CORAAL User Guide* (Kendall and Farrington 2018a; available from the website) for a more complete description of the corpus, its transcription and metadata, and other aspects of its design and development.

THE VALUE OF PUBLIC CORPORA

Extensive studies of AAL (see, e.g., Lanehart 2015, 2019 [this issue]) over the course of the twentieth century and the first decades of the present century have resulted in great advances in our knowledge about AAL varieties and about many of the sociolinguistic features that differentiate them from European American varieties of American English at both the qualitative level (e.g., the use of invariant *be*, copula absence, etc.) and quantitative level

(e.g., higher rates of features like consonant cluster reduction and velar nasal fronting than matrix European American varieties). However, at the same time that AAL has been so extensively studied, it remains massively under-represented in terms of publicly available data sets and in terms of its use in general linguistic theory building (Green 2002; Kendall, Bresnan, and Van Herk 2011). Almost all of the AAL data collected to date are in private collections and often not collected in a way that allows for wider, public sharing (as discussed further in Cukor-Avila and Balcazar 2019 [this issue]). These circumstances have created a situation in which the variety of American English often described as the most studied (Schneider 1996, 3)—a variety spoken through the history of the United States and currently spoken by millions of Americans—remains inaccessible to larger-scale, computational and corpus-based study. It is also difficult for those who have not played a role in data collection projects themselves, such as educational professionals or graduate students, to obtain primary materials for research and application. This is particularly an issue given how important public outreach and engagement is for AAL, which remains publicly misunderstood and its speakers subject to prejudice (see, e.g., Rickford and King 2016). Linguists have increasingly adopted models of data compilation in recent years that include data sharing and promoting data reuse (see Berez-Kroeker et al. 2018), but thus far almost all AAL data remain unavailable for wider public sharing, due to ethical considerations or limitations from how the data were collected (e.g., participant consent [Warner 2014]).

The availability of a public corpus of AAL is meant to enable new research and new uses. It provides access to primary data for a wider range of scholars and also seeks to support new “open science”-based approaches, where direct testing of competing theories or methodologies or reanalysis (see Rickford et al. 1991; Kendall 2011) can be made on the same data. Of course, a corpus is just a corpus, and ultimately how and why it is used comes down to the goals of its users and the values of the research community. CORAAL’s creators hope that it can become a valuable resource for advancing understanding about AAL, both within linguistics and across society.

THIS ISSUE

This special issue of *American Speech* brings together new work on AAL using the newly available CORAAL. In an unusual but exciting step for sociolinguistic research, each of the studies in this issue takes up different questions but is based on the same publicly available data. The studies all provide some initial insight into what CORAAL can tell us about AAL and sociolinguistic variation more generally, spanning a range of topics from variationist focus

on “core” English/AAL variables, to intonation, style-shifting, discourse, and computational analytic methods. In addition to providing us new perspectives on AAL through their range of research questions, the studies also demonstrate a number of ways that scholars and students can use CORAAL, and other public corpora for that matter, to investigate important questions about language. Online versions of the articles also take advantage of *American Speech*'s support of multimedia and the public nature of the corpus to incorporate audio excerpts for all data examples.³

The first two components of CORAAL, the focus of the studies here, feature AAL data from Washington, D.C., collected from two projects 50 years apart. This geographic focus on Washington, D.C., aligns with a historical focus in sociolinguistics, and in studies on AAL in particular, on Washington, D.C., and allows us to celebrate the contributions that have been made to sociolinguistic knowledge based on work in D.C. Thus, the issue begins with an article by Charlie Farrington and Natalie Schilling (2019) that reviews these important—some well known and some less so—contributions. Their article also provides useful background for future use of the CORAAL:DC components. Following the historical overview by Farrington and Schilling, six articles explore empirical topics.

Patricia Cukor-Avila and Ashley Balcazar (2019) take up questions about the distribution of core grammatical features of AAL, focusing in particular on present tense marking, across a subset of speakers in CORAAL:DC. In addition to providing an overview of how forms are distributed among the lower social strata speakers, they simultaneously use the corpus to raise important questions for empirical research on grammatical variation based on the “classic” sociolinguistic interview (see Becker 2013).

Jessica Grieser (2019) takes advantage of the consistency of the “classic” sociolinguistic interviews in DCA to offer new insights into topic-based style shifting. Her work expands upon the only handful of studies like Rickford and McNair-Knox (1994) to quantitatively probe the ways that morpho-syntactic variation correlates with interview topics, while also providing a powerful illustration of the value of putting to new work older data collected for other purposes.

Jon Forrest and Walt Wolfram (2019) examine the social and linguistic patterns of variable (ING), perhaps the mostly widely studied variable in English but one that has been relatively rarely examined in detail in AAL. Among other findings, they observe differences between the patterns in DCA and DCB, which is unexpected given the wide acknowledgment and evidence for (ING) being the paragon of a stable sociolinguistic variable.

Jason McLarty, Taylor Jones, and Christopher Hall (2019) explore how data sets such as CORAAL can usefully be examined in new, computational

ways. They consider postvocalic *r*-lessness, a classic sociolinguistic variable typically examined through binary (absence versus presence), impressionistic coding, as a more gradient feature using machine learning techniques, providing exciting food for thought for the future of sociolinguistic analysis.

Next, Nicole Holliday (2019) continues taking us in new directions, by examining the patterns of question intonation in the corpus. Her study simultaneously pioneers new questions in sociolinguistic study, the realization of boundary tones in questions of various types, while exploring and suggesting new corpus-based sociophonetic approaches to intonational analysis.

Finally, Minnie Quartey and Natalie Schilling (2019) examine a set of interviews from the DCB component of CORAAL to unpack how interviewees' stances and discourse patterns negotiate identities of connectedness and disconnectedness. Together, as researchers with long ties to Washington, D.C., Quartey, the primary field-worker for DCB, and Schilling offer valuable insights into the discourse patterns of speakers in the DCB component of CORAAL and the people and places of Washington, D.C.

Altogether, these studies provide exciting new perspectives on AAL and corpus-based sociolinguistics and set the stage, we hope, for further new work on AAL using public corpora.

NOTES

1. CORAAL is available for free download from its website at <https://oraal.uoregon.edu/coraal/>, where readers will also find a comprehensive user guide (Kendall and Farrington 2018a), which provides extensive details about the corpus and its creation.

CORAAL was made possible with the help of very many individuals, ranging from the many field-workers who collected data, to the very many interviewees who contributed their valuable voices and time, to the many research assistants at the University of Oregon who helped to collate, transcribe, annotate, and redact the data for the corpus, and, finally, to the many colleagues who participated in this issue or otherwise helped in the preparation of the corpus or by beta-testing or giving feedback on earlier versions of the corpus. This wide range of individuals, as well as the participants who asked not to remain anonymous, are thanked in the user guide.

CORAAL, and its parent project, Online Resources for African American Language (<https://oraal.uoregon.edu/>), have been made possible by support from the U.S. National Science Foundation (Grant No. BCS-1358724) and from the University of Oregon.

2. In addition to CORAAL's two D.C. components, the corpus is scheduled to add several smaller, supplemental components to provide regional breadth. Two such components are already available: CORAAL:PRV (Rowe et al. 2018), which

includes a sample of sociolinguistic interviews from Princeville, a rural African American community in central North Carolina, and CORAAL:ROC (King et al. 2018), which includes interviews from Rochester, New York. These components were not examined by the authors in this issue.

3. These samples can be viewed and heard, in situ, through the CORAAL Explorer website (<http://lingtools.uoregon.edu/coraal/explorer/>).

REFERENCES

- Becker, Kara. 2013. "The Sociolinguistic Interview." In *Data Collection in Sociolinguistics: Methods and Applications*, edited by Christine Mallinson, Becky Childs, and Gerard Van Herk, 107–33. New York: Routledge.
- Berez-Kroeker, Andrea L., Lauren Gawne, Susan Smyth Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, et al. 2018. "Reproducible Research in Linguistics: A Position Statement on Data Citation and Attribution in Our Field." *Linguistics*, 56, no. 1 (Jan.): 1–18. <https://doi.org/10.1515/ling-2017-0032>.
- Cukor-Avila, Patricia, and Ashley Balcazar. 2019. "Exploring Grammatical Variation in the Corpus of Regional African American Language." In Kendall and Farrington 2019, 36–53. <https://doi.org/10.1215/00031283-7321989>.
- Farrington, Charlie, and Natalie Schilling. 2019. "Contextualizing the Corpus of Regional African American Language, D.C.: AAL in the Nation's Capital." In Kendall and Farrington 2019, 21–35. <https://doi.org/10.1215/00031283-7308060>.
- Fasold, Ralph W. 1972. *Tense Marking in Black English: A Linguistic and Social Analysis*. Arlington, Va.: Center for Applied Linguistics. <https://eric.ed.gov/?id=ED129065>.
- . 2019. "Looking Back at 1968 Fieldwork in Washington, D.C." Preface to Kendall and Farrington 2019, 9–12. <https://doi.org/10.1215/00031283-7482420>.
- Forrest, Jon, and Walt Wolfram. 2019. "The Status of (ING) in African American Language: A Quantitative Analysis of Social Factors and Internal Constraints." In Kendall and Farrington 2019, 72–90. <https://doi.org/10.1215/00031283-7308049>.
- Green, Lisa J. 2002. *African American English: A Linguistic Introduction*. Cambridge: Cambridge University Press.
- Grieser, Jessica. 2019. "Investigating Topic-Based Style Shifting in the Classic Sociolinguistic Interview." In Kendall and Farrington 2019, 54–71. <https://doi.org/10.1215/00031283-7322011>.
- Holliday, Nicole. 2019. "Variation in Question Intonation in the Corpus of Regional African American Language." In Kendall and Farrington 2019, 110–30. <https://doi.org/10.1215/00031283-7308038>.
- Kendall, Tyler. 2011. "Corpora from a Sociolinguistic Perspective." In "Corpus Studies: Future Directions," edited by Stefan Th. Gries. Special issue, *Revista Brasileira de Linguística Aplicada* 11, no. 2: 361–89. <https://doi.org/10.1590/S1984-63982011000200005>.

- Kendall, Tyler, and Charlie Farrington, eds. 2018a. *CORAAL User Guide*. Version 2018.10.06. Eugene, Ore.: Online Resources for African American Language Project. <http://lingtools.uoregon.edu/coraal/userguide>.
- . 2018b. The Corpus of Regional African American Language. Version 2018.10.06. Eugene, Ore.: Online Resources for African American Language Project. <https://oraal.uoregon.edu/coraal>.
- , eds. 2019. "Exploring African American Language in the Nation's Capital: Studies with the Corpus of Regional African American Language." Special issue, *American Speech* 94, no. 1 (Feb.). <https://read.dukeupress.edu/american-speech/issue/94/1>.
- Kendall, Tyler, Joan Bresnan, and Gerard Van Herk. 2011. "The Dative Alternation in African American English: Researching Syntactic Variation and Change in a Conglomerated Corpus." *Corpus Linguistics and Linguistic Theory* 7, no. 2 (Oct.): 229–44. <https://doi.org/10.1515/cllt.2011.011>.
- Kendall, Tyler, Ralph Fasold, Charlie Farrington, Jason McLarty, Shelby Arnson, and Brooke Josler. 2018. The Corpus of Regional African American Language: DCA (Washington, D.C., 1968). Version 2018.10.06. Eugene, Ore.: Online Resources for African American Language Project. <https://oraal.uoregon.edu/coraal>.
- Kendall, Tyler, Minnie Quartey, Charlie Farrington, Jason McLarty, Shelby Arnson, and Brooke Josler. 2018. The Corpus of Regional African American Language: DCB (Washington, D.C., 2016). Version 2018.10.06. Eugene, Ore.: Online Resources for African American Language Project. <https://oraal.uoregon.edu/coraal>.
- King, Sharese, Charlie Farrington, Tyler Kendall, Emma Mullen, Shelby Arnson, and Lucas Jenson. 2018. The Corpus of Regional African American Language: ROC (Rochester, N.Y., 2016). Version 2018.10.06. Eugene, Ore.: Online Resources for African American Language Project. <https://oraal.uoregon.edu/coraal>.
- Lanehart, Sonja, ed. 2015. *The Oxford Handbook of African American Language*. New York: Oxford University Press.
- . 2019. "CORAAL Is Only the Beginning." Preface to Kendall and Farrington 2019, 6–8. <https://doi.org/10.1215/00031283-7482413>.
- McLarty, Jason, Taylor Jones, and Christopher Hall. 2019. "Corpus-Based Sociophonetic Approaches to Postvocalic R-lessness in African American Language." In Kendall and Farrington 2019, 91–109. <https://doi.org/10.1215/00031283-7362239>.
- Quartey, Minnie, and Natalie Schilling. 2019. "Shaping 'Connected' versus 'Disconnected' Identities in Narrative Discourse in D.C. African American Language." In Kendall and Farrington 2019, 131–47. <https://doi.org/10.1215/00031283-7322000>.
- Rickford, John R., and Sharese King. 2016. "Language and Linguistics on Trial: Hearing Rachel Jeantel (and Other Vernacular Speakers) in the Courtroom and Beyond." *Language* 92, no. 4 (Dec.): 948–88. <https://doi.org/10.1353/lan.2016.0078>.

- Rickford, John R., and Faye McNair-Knox. 1994. "Addressee- and Topic-Influenced Style Shift: A Quantitative Sociolinguistic Study." In *Sociolinguistic Perspectives on Register*, edited by Douglas Biber and Edward Finegan, 235–76. New York: Oxford University Press.
- Rickford, John R., Arnetha Ball, Renee Blake, Raina Jackson, and Nomi Martin. 1991. "Rappin on the Copula Coffin: Theoretical and Methodological Issues in the Analysis of Copula Variation in African-American Vernacular English." *Language Variation and Change* 3, no. 1 (Mar.): 103–32. <https://doi.org/10.1017/S0954394500000466>.
- Rowe, Ryan, Walt Wolfram, Tyler Kendall, Charlie Farrington, and Brooke Josler. 2018. The Corpus of Regional African American Language: PRV (Princeville, N.C., 2004). Version 2018.10.06. Eugene, Ore.: Online Resources for African American Language. <https://oraal.uoregon.edu/coraal>.
- Schilling, Natalie. 2013. *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.
- Schneider, Edgar W. 1996. "Introduction: Research Trends in the Study of American English." In *Focus on the USA*, edited by Edgar W. Schneider, 3–12. Amsterdam: Benjamins.
- Warner, Natasha. 2014. "Sharing of Data as It Relates to Human Subjects Issues and Data Management Plans." *Language and Linguistics Compass* 8, no. 11 (Nov.): 512–18. <https://doi.org/10.1111/lnc3.12107>.

TYLER KENDALL is associate professor of linguistics at the University of Oregon. Much of his research focuses on variation and change in American English, with emphases on both language production and perception across regional and ethnic varieties of English. He is the author of *Speech Rate, Pause, and Sociolinguistic Variation: Studies in Corpus Sociophonetics* (Palgrave Macmillan, 2013) and coeditor of two recent PADS volumes, *Speech in the Western States*, volumes 1 (*The Coastal States*) and 2 (*The Mountain West*) (Duke Univ. Press, 2016, 2017). In addition to CORAAL, he has developed and currently maintains a number of resources for sociolinguistic research including the Sociolinguistic Archive and Analysis Project and the NORM website and vowels.R package for vowel normalization. Email: tsk@uoregon.edu.