The Oral History Review

ISSN: 0094-0798 (Print) 1533-8592 (Online) Journal homepage: www.tandfonline.com/journals/uohr20

A Meeting of Two Worlds—Oral History and Linguistics: Partnerships, Perplexities, and Potentialities in Researching African American Language

Matthew F. Simmons, Anna Hamilton, Alexis Davis, William Dyer, Christina Jean Philippe, Jessica Heady, Adolfho Romero, Rebekah Anne Cordova, Deborah Hendrix, Paul Ortiz, Pryce Houck, Kevin Tang & Sarah Moeller

To cite this article: Matthew F. Simmons, Anna Hamilton, Alexis Davis, William Dyer, Christina Jean Philippe, Jessica Heady, Adolfho Romero, Rebekah Anne Cordova, Deborah Hendrix, Paul Ortiz, Pryce Houck, Kevin Tang & Sarah Moeller (06 Aug 2025): A Meeting of Two Worlds—Oral History and Linguistics: Partnerships, Perplexities, and Potentialities in Researching African American Language, The Oral History Review, DOI: 10.1080/00940798.2025.2536129

To link to this article: https://doi.org/10.1080/00940798.2025.2536129

THE ORAL HISTORY REVIEW
https://doi.org/10.1080/00940798.2025.2536129

Routledge
Taylor & Francis Group

A Meeting of Two Worlds—Oral History and Linguistics: Partnerships, Perplexities, and Potentialities in Researching African American Language

Matthew F. Simmons, Anna Hamilton, Alexis Davis, William Dyer, Christina Jean Philippe, Jessica Heady, Adolfho Romero, Rebekah Anne Cordova, Deborah Hendrix, Paul Ortiz, Pryce Houck, Kevin Tang, and Sarah Moeller

ABSTRACT

The realms of oral history and linguistics can provide insight into the geographic, social, historical, and linguistic connections of different populations. In this article, we seek to highlight the interactions between the fields as they relate to *African American Language* (AAL), specifically spoken by African Americans in the Gulf South. We present a case study of how the two disciplines, linguistics and oral history, complement each other through an analysis of the Joel Buchanan Archive at the University of Florida Digital Collections. The unique grammatical features of AAL are present in many oral history interviews that we analyzed, and we argue that our findings show that oral history is beneficial for education and linguistics and that linguistics technological endeavors are increasingly important to the oral history field. Throughout our process, we address the challenges of preserving unique grammatical features in language varieties without a written standard and examine possible routes to provide consistent and linguistically accurate transcripts of an oral language variety through oral history work. AAL has a rich history within the United States, and it is an important part of Black culture. We believe our efforts make oral histories conducted in AAL accessible to the general population, provide informed guidelines for transcription and linguistic feature annotation, and streamline the collection of AAL for use in research, education, and technology settings.

KEYWORDS

African American Language; African Americans; Black Americans; Black culture, grammar; Gulf South; language science; linguistics

This article stems from a project titled "Reanimating African American English Oral Histories of the Gulf South." This collaboration is primarily between the Samuel Proctor Oral History Program (SPOHP) and the Department of Linguistics at the University of Florida (UF), but it also involves statisticians, computer scientists, and K–12 education experts, as well as the African American community in Putnam County, Florida, where the state's first Black high school stands. The linguistics team has two computational linguist faculty and several student assistants who are mainly from the departments of Linguistics and Computer and Information Science and Engineering at UF and the Department of English Language and Linguistics

at Heinrich Heine University in Düsseldorf, Germany. The oral history team consists of SPOHP oral history practitioners and digital humanities production specialists. Our archival experts and website manager work at UF's George A. Smathers Libraries. Additional team members include a subject matter expert whose role is to identify important historical themes and build a K–12 curriculum about African American history and language.

Through this collaboration, we seek to produce, test, and disseminate a new online interface with accompanying annotations that enrich the narratives beyond the transcripts and to prepare a training data set for speech technology to better recognize African American speakers. This work utilizes the content of interviews with African Americans in the Gulf South and is geared toward application in both educational and research settings. To prepare the oral history narratives, we developed the following workflow:

1. Review transcripts, checking that they adhere to SPOHP transcription guidelines, making them more verbatim where they cause issues for automatic time-alignment (forced alignment), and editing them to fit our new African American Language (AAL) transcription guidelines.
2. Analyze and annotate sentences for occurrences of seven AAL grammatical features to accommodate in-depth analysis of AAL and better recognition of AAL by natural language processing (NLP)/artificial intelligence (AI), simultaneously aligning the transcribed sentences to their timestamps in the audio files.
3. Create K–12 history curriculum based on a subset of these interviews.
4. Design a website and customize an interactive software that enhances public engagement with the Joel Buchanan African American Oral History Archive (JBA) at UF through the material we produced in the previous steps.

This article describes the lessons we learned from our collaborative work, particularly in steps 1 and 2.

Oral History Collections

Oral history archives house an abundance of information for language science, which examines not only narrative content and conversation but also structural properties of language (such as phonetics, phonology, morphology, syntax, semantics, and pragmatics) and their usage. For example, University of Southern California's Shoah Foundation Virtual History Archive Online and the United States Holocaust Memorial Museum's Oral History Collection include interviews in more than thirty different languages.[1] Oral history interviews capture spontaneous and authentic language behavior in real-life situations, and their contents are examples of sociolinguistic (sociology of language

use) information. Oral history interviews, with their emphasis on individual life stories, mirror how linguistic researchers begin the study of a language, particularly one that has not been documented yet.

In this article, we describe collaborative efforts between the Department of Linguistics and SPOHP at UF that were funded through the National Endowment for the Humanities and American Council for Learned Societies. The project—Reanimating African American English Oral Histories of the Gulf South—shows how collaboration between oral history and linguistics supports the advancement of both fields and facilitates the development of AI for underserved communities, demonstrating a shared power between oral historians and linguists. We show how oral histories highlight the uniqueness of AAL and how oral historians can facilitate new routes of linguistics research. We also demonstrate how linguistic insight can help oral historians gather naturalistic conversation, address current struggles in work for AAL, and track vital information about the speakers and recordings. Collaboration with linguists provides potential benefits to oral history, such as increased collection from minority communities, shared best practices vis-à-vis transcription and metadata, and the application and development of new technologies (such as automatic speech recognition). We found that an awareness of oral history as a source of linguistic data could open up additional funding and support for both fields.

We also explore this collaboration because extensive labor and time is necessary to accrue resources for linguistics, oral history, and building AI that is accessible to all communities—work that is reduced and enriched by interdisciplinary collaboration between linguists and oral historians. Marriages of linguistics and oral history highlight the potential we emphasize. For example, the Oral History of Empires by Elders in the Arctic project, based out of the Arctic Centre at the University of Lapland, was implemented by a team of anthropologists to document the lives and culture of four Arctic Indigenous groups.[2] They connected the two disciplines by documenting endangered languages in the form of oral histories. They showed how capturing unplanned speech on topics that native speakers personally relate to can align well with expectations of both fields. Crucially, this approach preserves not only history but also linguistic structures and sociocultural knowledge.

Our collaboration focuses on the language variety known as African American Vernacular English (AAVE) or African American Language (AAL).[3] AAL is unique not only in pronunciation and vocabulary, which most English speakers can identify, but it also has distinctive grammatical structures that communicate hues of meaning not as readily available to general or Mainstream American English (MAE). AAL is a vibrant thread in the fabric of Black culture and an important part of the American linguistic landscape. It is estimated to have more than thirty million native speakers, whose history has been lived and their lives expressed primarily through that

linguistic system. Sadly, our understanding of AAL is limited primarily because naturalistic examples of AAL speech are limited. For example, although we know AAL is not monolithic, little is known about the regional variations of AAL.

The valuable range of regions, educational backgrounds, and generations that are found in oral history collections is something that oral history can deliver to linguistics research, particularly through projects that focus on specific topics, regions, and populations. If this same work were to be conducted solely under a linguistic study, the focus would likely be narrower in accordance with the more limited resources, the capabilities of the limited number of researchers, and the time available to work on one project. Oral historians can benefit from a deeper understanding of AAL and its regional variations since language and culture are linked, providing an opportunity for oral historians to better understand the deeper meanings behind the words collected through the interview process. The beauty of the language is reflected in the SPOHP archives by voices of people like Benny Goodman: "Then the Cuban Crisis broke out, and they put me on an airplane and sent me to Fort Stewart, Georgia, to go down to Guantanamo to fight Castro . . . . And the only thing I could think about was that somebody was going to die, because I could not talk on the microphone. My job was Morse code; and nobody knew it but me, so I prayed, *Lord please don't let this war slide*. It happened because of the dude; somebody is going to die 'cause I'm not going to be able to communicate with nobody with a microphone. And I'm trying to say, 'Do something about some fire power or something'; they weren't going to get it 'cause nobody knew Morse code but me. So, it didn't happen. Kennedy came down and told us that Khrushchev turned the ships back."[4]

Linguistic Treasures in Oral History Collections

Amid pressures of globalization and monolingual education, linguists estimate 30 to 90 percent of the world's 7,100-plus languages, along with their distinct varieties, may disappear by the end of this century.[5] Transcribed collections of spoken language not only counteract this loss to science but treat each community's unique way of speaking as an intangible heritage worthy of preservation. Within the scope of a language, there may exist unique but mutually intelligible varieties that remain understudied even as they may be evolving into new and distinct languages. Preserving examples of languages and their varieties reveals how our brains convert human experiences into linguistic systems, but serious attempts to catalog the world's languages did not begin until the 1950s. Up until the 1990s, when the crisis of language endangerment came into focus, linguists rarely archived their recordings. This dearth of historical documentation hinders our understanding of human

communication and the identification of distinctive structures, pronunciation, and variations in and across communities.[6]

The linguistic value of oral history collections increases when a collection focuses on socially or geographically distinct communities. The factors that grant a community their unique history are the same that produce a unique language. Collections like the JBA at UF or the Yiddish Book Center's Wexler Oral History Project preserve personal narratives and preserve unique phonetic and grammatical intricacies of their language varieties.[7] Since our case study addresses a variety of English spoken by many African Americans, we note here that the term "language variety" is preferred by many linguists over "dialect." The key distinction between variety and dialect lies in flexibility. Variety is a broad, neutral term that encompasses languages, dialects, jargons, written forms, registers, and even individual speech styles. This term is especially useful in fields like sociolinguistics, dialectology, and historical linguistics, as it helps categorize different forms of mutually intelligible language without implicating politically or emotionally charged connotations that can arise with a less precise use of "dialect." When used, dialect typically refers specifically to a regional form of a language that differs from other forms in pronunciation, grammar, and vocabulary but has some level of mutual intelligibility with those forms.

It is also worth noting that linguists tend to view documented artifacts of language somewhat differently than oral historians and that this may cause confusion in terminology. Linguists recognize that languages are systems of structures, not collections of words. Linguists shift back and forth between examining language as abstract structures and as a mode of communicating information. When writing mainly from a linguist's perspective, we use "database," "(language) data," or "discourse" or "text"; otherwise, we use "collection," "oral history," "narrative," or the more generic "recording," "transcript," or "transcription." By referring to documented speech as "data," we mean that the recorded language can be studied for underlying structural information (e.g., grammar and phonology) as much as for the content it communicates.

Given that language data takes immense time to simply collect, let alone categorize and process, oral history collections are an asset for language sciences. This is particularly relevant for endeavors related to AAL, because one of the most prominent obstacles to solid understanding of AAL is the lack of publicly available speech recordings (oftentimes, examples of the language are restricted and require special permissions for access). In publicly available oral history projects, narrators grant consent, and their identities are stored (unless they have stipulated otherwise). Personal narratives often include demographic information such as where the speaker grew up or where their parents are from (possible factors in speech varieties). All this reduces the burden for linguistics researchers. Additionally, the organization of oral

history collections is a great model for linguists to use in making their recordings available for the general population.

Another benefit from oral history lies in its contribution to NLP, an intersection of linguistics and computer science that uses AI to help computers "understand" human language. While what constitutes fairness and equity in AI is currently an open question, much research has focused on the data that an AI system is trained on and the "unfair" or "biased" decisions it makes.[8] A fair AI requires many examples encompassing varied expressions. Leveraging audio that is time-stamped to the accompanying transcripts can aid in developing speech technology that accurately reflects underrepresented dialects and accents in marginalized communities most affected by biased AI.[9] Speech recognition models can be either trained entirely or fine-tuned on oral history interviews, which could enable them to better replicate diverse speech patterns and storytelling in underrepresented cultures and accents.[10] Relatedly, by annotating the transcripts with linguistic information, NLP models learn to process meaning and retrieve information across various linguistic groups. Furthermore, oral history interviews can inform conversational AI systems, ensuring that they better understand diverse ways of dialogue, question-asking styles, and emotional context that may be misinterpreted due to linguistic or cultural variation.

Oral Historians Benefit from Collaborating with Linguists

Linguistics also offers value for oral history in the collection of narrative texts.[11] Linguists who document endangered languages and language varieties start by visiting the community of speakers to record samples of their language and transcribe those recordings. Documentary linguists understand that a study of a language entails the speakers' history and culture that are contained in personal narratives, conversation, and oral literature (poems, songs, sermons, etc.). As documentary linguists Michael Rießler and Joshua Wilbur note, their research on endangered Uralic languages spoken in the Artic "can be considered an additional source for future oral history studies."[12] Other subfields such as sociolinguistics that look closely at the content of communication also gather information through personal interviews, sometimes aiming to gather a person's linguistic biography or full life history. This work overlaps directly with oral history and its emphasis on capturing "history from below": stories from marginalized people who might otherwise be lost to the historical record. Consequently, linguistic fieldwork presents a resource that oral historians may not be aware of.

The field of oral history also stands to benefit from new transcription technologies being developed by linguists working in NLP. Speech recognition models that are trained to transcribe speech of underrepresented communities can expedite the process for oral historians transcribing interviews of speakers

in those communities. Traditionally, writing the first draft of transcripts has been incredibly time-consuming. Recent innovations in speech recognition technology dramatically speed up this step. Linguists are further refining recent technologies by utilizing language data to more completely capture speech variations. By consulting with linguists, oral historians on our team learned that one might use an automatic speech recognition (ASR) tool, such as CrisperWhisper,[13] to transcribe verbatim as a first draft instead of OpenAI's Whisper, which often omits disfluencies (interruptions or disruptions in the flow of speech).[14] CrisperWhisper transcribes every sound exactly as it is, including fillers (like "um" and "uh"), pauses, stutters, and false starts. Although capturing these details may not fit the oral historians' goals, they happen to be particularly valuable for automatically aligning transcripts to the recordings' time stamps. This time-alignment easily accommodates creating an auto-scroll feature for presenting transcripts while listening to a narrative. Such interactive displays are ways for linguists to "give back" to the community, and oral historians may adopt similar methods that make the narratives more accessible.

Oral historians may also borrow methodological approaches from linguistics for their own studies. For example, sociolinguists have long established that the attitude of speakers toward their culture and community is reflected in speech subtleties. Linguists, using NLP or manual analysis, can detect a narrator's sentiment toward entities or historical events. For instance, an interviewee who has a negative attitude toward their African American identity might use fewer linguistic features associated with AAL, or an emotionally fraught episode in one's life might be revealed by the appearance of features from the narrator's native speech variety rather than through the use of emotional vocabulary. This type of insight can enrich oral history interviews, allowing oral historians to place spoken words within a broader sociological context surrounding the narrator's belonging and identity. Still, the responsible use of AI tools like sentiment analysis by oral historians necessitates that this be just one tool out of many in their toolbox—one that should be used in conjunction with other means of determining the relationship between speakers and their community.[10]

Moreover, a recent surge in connecting oral history with computational methods shows that computational linguistics has the potential to enhance the study of oral history interviews. Previous work by Francisca Pessanha and Almila Akdag Salah demonstrated how oral historians can develop a deeper understanding of interviews through *forced alignment*.[15] This method "aligns" the acoustic signal to its corresponding orthographic (i.e., written) representation, creating a time-stamped transcription. Crucially, it can be done automatically with minimal manual input.[16] Once aligned, the interview can then be subjected to a line-by-line analysis with annotation of the corresponding transcript to show how each part reflects the identity of the interviewees or

how the information structure of their discourse reveals the salience of entities and events in their narrative. With NLP, oral historians can now more easily examine narratives and gain new insights into the expressions of interviewees at a granular level, examining not only what is expressed but how it is expressed.

Case Study: African American Language

AAL differs from other English varieties in pronunciation, grammar, and vocabulary. Our project is currently focusing on seven distinctive grammatical structures used by African American speakers:[17]

- Person/number disagreement ("They was just across the park from me.")
- Habitual *be* ("We always be hungry when we smell that food.")
- Multiple negators ("She don't never watch where she's driving.")
- Remote past *been* ("I been cleaned up the kitchen.")
- Existential *they* ("Dey was over here yesterday.")
- Perfect *done* ("They done asked too much of me.")
- Null copula ("We so old now.")

We chose these seven features for their frequency, their significance in AAL speech, and their semantic relevance to mutual intelligibility with other English varieties. For example, the habitual *be* is a trademark feature in AAL: "I be in my office by 7:30" versus "I am (usually) in my office by 7:30" in standard English. These sentences are representative of possible misunderstandings, as the AAL sentence communicates the individual is present in their office on a consistent basis by a certain time, but the second sentence is more ambiguous about the person's consistent presence.[18]

A few of the other features require some explanation. Multiple negation is when multiple negator words within a single grammatical clause confirm an overall negative meaning. For example, the negative meaning in "I ain't step on no dog" is communicated first in "ain't" and then confirmed in the quantifier "no." This clausal negation strategy is common across many languages and in several varieties of English. It is the most prominent AAL grammatical feature in our oral histories, and it seems to be the most identifiable feature for our research assistants. A 2010 interview with Christine Holmes contains as many as thirty instances of multiple negators, such as in the following example: "They won't let me walk nowhere."[19]

Remote past *been* (or *bin*) operates with a similar function as the habitual *be*, but its usage changes the time (or tense) of the event named by the main verb, specifying the event was in the distant past. For example, if a mother states, "We been adding cinnamon to the cookies," she is not discussing a recent change made to the recipe but explaining the tradition has long

been established. A 2010 interview with Carnell Henderson includes several instances of remote past *bin*, such as, "But we been enjoying it very much from way back up until now."[20] Transcribers who lack a background in AAL, or those who follow a prescriptive approach, may incorrectly apply Standard or General American English (SAE or GAE) logic and add "has/have/had" before the "been," not realizing that this addition completely changes the meaning.

Existential *it*/*they* (or *it*/*dey*) represents the pronunciation of the existential construction. It is included because this pronunciation distinguishes the existential meaning (for example, "There are red notebooks in every store") from the locative ("There are my red notebooks!") and because it is often used in conjunction with the null copula (e.g., "They some coffee in the kitchen"). A 2010 interview with Alvin Butler uses a comparatively large number of utterances containing existential *it*/*dey*, such as the following: "When I went to high school, it wasn't but four students had automobiles."[21] Existential *it*/*dey* occurs in our data set significantly less frequently than other morphosyntactic features of AAL.

The null copula (also called "zero copula" and "no copula") is the absence of the copula (linking) verb "to be." The null copula assumes a default present tense meaning. For example, "She likin' me . . . she likin' George too" is assumed to be the equivalent of "She is liking me, and she is liking George too." The context in which a speaker omits the copular verb may vary across many sentence constructions with various subjects and predicates. Christine Holmes's interview produced fifty-two utterances containing null copula, including the following question: "Check and see, what year that—what year Frankie born?"[22]

African American Speech Collections

Linguistic study of AAL is limited primarily because naturalistic examples of AAL speech specifically collected and designed for linguistic research are limited. The lack of data leaves linguists to rely on "web crawls" and data mining of social media websites (as opposed to traditional sociolinguistic interviews), which poses issues because the identities and geographical origin of AAL users are not always verifiable online. Non–African American speakers of AAL may also use vocabulary and grammatical structures similar to but still different in ways from African Americans, and this skews the linguistic analyses. Also, web crawls and data mining pose privacy concerns.

The first linguistic AAL speech database of significant size that is also publicly and freely available is the Corpus of Oral Regional African American Language (CORAAL).[23] CORAAL is part of the University of Oregon's Online Resources for African American Language Program (ORAAL). CORAAL is one of the only, and largest, online collections of AAL speech, mostly from projects conducted in the Southeast (such as in Washington, DC). Speech is not monolithic,

and endeavors such as CORAAL highlight the importance of authentic and varied speech samples from Black communities. These recordings came from different research projects such as Ralph Fasold's fieldwork in Washington, DC, and Sharese King's dissertation on AAL in Rochester, New York.[24] They are categorized into four age groups (12–19, 20–29, 30–50, 51+) and three social classes ranging from lower working class to upper middle class. The corpus is aimed to be gender-balanced with two male and two female speakers for each demographic combination of age groups and social classes.

The JBA—housed at SPOHP and online at UF Digital Collections—was founded in 2019 and is one of the largest public access oral history collections in the United States.[25] This growing digital and physical archive contains more than seven hundred interviews with Black elders in Florida and the Gulf South and aggregates a number of distinct SPOHP projects from as early as the 1970s. Many narrators in this collection formed the leading cadres for the civil rights movement and recall connections to historical luminaries like Martin Luther King, Jr., Mary McLeod Bethune, and Zora Neale Hurston. The JBA's namesake, Joel Buchanan, was a longtime resident of Gainesville, Florida (where UF is located) and a historian of African American history in Alachua County (where Gainesville is located). Buchanan was active in local civil rights efforts and has the distinction of being one of the first Black students to integrate Gainesville High School in 1964. Buchanan received his bachelor's and master's degrees from UF, worked as a teacher in area schools, and joined UF's Department of Special and Area Studies Collections in 2004 to help document and preserve the history of Alachua County's African American residents.

The JBA includes the Mississippi Freedom Project, which derives from fifteen years of SPOHP's annual trips to the Mississippi delta to interview civil rights movement veterans; the Oscar Mack Project, detailing the remarkable story and legacy of Oscar Mack, a World War I veteran who was lynched in Osceola County, Florida, in 1922 but survived, along with his family; the Underground Railroad collection, which includes interviews with Black Seminoles and Gullah-Geechee elders and leaders; the Civil Rights in St. Augustine collection begun by David Colburn in the late 1970s; the St. Augustine African American History collection; the Bo Diddley Oral History Project; Black Faculty Retention at UF Oral History Project; Black Students Sharing Stories; Coalition of Immokalee Workers; Farmworker Association of Florida; Florida Blacks; Martin County Black Heritage; Melrose, Florida; Pandemic Oral History Project; and Vietnam War Veterans; among others. Many collections in the JBA are accessible through the UF's Digital Collections, while the remaining non-digitized materials may be accessed at SPOHP.[26]

The participants in these interviews were aged anywhere from their teens to late seventies. For example, Alexis Cooper, who participated in the Mississippi

Freedom Project, was a teen in middle school when her interview was conducted, and John Booth, who participated in the African American History Project, was sixty-two years old when he was interviewed in 2009. Moreover, participants' socioeconomic statuses in general are quite varied.

Challenges for Linguists Leveraging Oral Histories

Before oral history collections can be used for linguistic research or AI training, some preprocessing (preparation and formatting of digital data) is required. Unfortunately, our experience preparing the recordings and transcripts did not run smoothly. Small but significant differences in common workflows and goals between the two fields left gaps between our expectations and reality. Here, we present a few major challenges when preprocessing oral histories for linguistic research. We then discuss solutions we implemented and propose others for consideration. We wish to note that the preprocessing steps described here do not remove historical content. The goal is to prepare the collections so that key scientific information found in the narratives can be annotated for further study.

The first challenge arose from differences in metadata expectations. Simple pieces of information convey invaluable sociolinguistic and situational insights that may have impacted the historical content communicated in the narrative as well as how and why that content was communicated. According to communication accommodation theory, humans navigate identity and solidarity with interlocutors by adjusting their language.[27] People implicitly and explicitly identify themselves and their relation or attitudes during social interactions by modifying various aspects of communication. For example, a narrator may unconsciously choose a more formal linguistic *register* (variety of speech used in a specific context) when speaking to an older person or with someone who is "dressed up" compared to how they might speak with a casually dressed student. An African American speaking to a White person may tend their speech toward a more "standard" English, particularly in a perceived formal situation (e.g., being recorded for history). Other factors such as the presence of one's grandchildren may influence the choice of stories as well as the specific genre of storytelling. It can also provide additional and richer contextualization, such as in the case of interviewee Charles Bryant's grandchild: "Hey Granddad, now, earlier you said that the White and the Black people didn't like to see you have, or grow in what you had. Which one was worse? Was it worse that the Black people did you wrong, or was it worse that the White people did you wrong?"[28]

To understand how these factors impacted narratives, metadata about the participants and interviewers is key, just as they are for oral historians seeking to understand interview dynamics. Were the interlocutors of the same age, ethnicity, or region of origin? Were the interviewers primarily students and the narrators primarily older adults? Similarly, the linguistic background of

the transcribers can help us understand what factors might have guided transcription decisions when they encountered language varieties without a standardized orthography (writing system), such as AAL.

When the linguists on our team first approached the JBA, we were delighted with the wealth of language data. As a first step, we needed to filter out the JBA interviews with White narrators who speak about their experience during the 1960s civil rights demonstrations to ensure we were not characterizing African American speech based on transcripts of non-AAL speakers. The interview with Allen Cooper is one such example from the archives. Cooper was a civil rights activist and member of CORE (Congress of Racial Equality) who was interviewed as part of the Mississippi Freedom Project in 2004. Fortunately for researchers, Cooper identifies himself as "an ordinary little White guy in Albuquerque" very early on in the interview.[29] We began requesting metadata records from SPOHP, expecting to receive what we assumed were examples of essential information to be documented in any speech recording project. To our surprise, SPOHP's oral historians do not keep extensive structured records about the narrators; information about transcribers and interviewers is even more scarce. We were surprised that the oral history metadata records in an African American history project did not indicate at least race/ethnicity. Upon further research we realized that this was not an issue unique to SPOHP but rather a long-standing practice within the field. Oral historians have only recently begun to examine the practical implications of "thin" metadata and its impact on present-day research. For example, recent scholarship on the McKeldin-Jackson Project, a civil rights movement project from 1975, was particularly hamstrung by an inability to identify the race of interviewers and narrators and had to rely on additional means such as obituaries from the *Baltimore Sun* to determine the race of speakers.[30]

A second major challenge our linguists encountered is a lack of policy for orthographic representation of speech, particularly for nonstandard varieties. At first, due to our own training, the linguists assumed that someone trained to transcribe oral history would never consciously change "it was so many of us today" to "there were so many of us today," but we found this and other examples in JBA transcripts. (Without demographic metadata about the transcriber, it is difficult to say why this happened, as one student research assistant discovered when they undertook an honors thesis to explore the reasons.) We surveyed (to a limited extent) oral history transcription guidelines to better understand how oral historians approach transcription of recorded speech.[31] We also wanted to devise transcription guidelines for AAL that would satisfy linguistic precision but match oral historian common practices for handling varieties without standardized orthographic representations. We discovered a range of instructions for transcription across the oral history projects we surveyed and very little instruction about handling nonstandard English varieties. We explore this in more detail in a later section.

A third challenge presented to the linguists is that oral history transcripts are often not time-aligned to the speech (especially archival, nondigital transcripts). Writing serves only as a reference to the real language that is reflected in the recordings. A key step for linguistic investigation is aligning transcribed phones (individual sounds), words, phrases, and utterances (e.g., sentences) to time stamps in the audio file. For linguists, a traditional pre-step before transcribing is segmenting the audio file into larger units such as utterances, although ASR, when available for a given language variety, makes this possible as a postprocessing step. Voice technology to correctly recognize nondominant speech varieties absolutely requires the creation of time-aligned transcripts. Therefore, by applying time-alignment, oral history collections become usable for both broader research and training speech technology that is broadly accessible. Incidentally, prioritizing standard written rules over more precise representation of speech makes it difficult to train and employ AI assistance to automatically time-align. The mismatch between audio and transcription can also inhibit attempts to reduce AI bias. Finally, time-alignment can be directly useful to oral historians who may, for example, want to search for a keyword in a transcription and then jump immediately to that audio excerpt.

Avenues of Collaboration: Metadata That Benefit Both Fields

We found the most challenging barrier to collaboration between linguists and oral historians arose from "thin" metadata in the oral history collection. This primarily applies to information about the people involved in gathering and processing the oral histories. The lack of structured metadata forced us to abandon some fascinating questions about AAL regional differences and how social settings impact speakers' use of AAL. This section describes why metadata is important and recommends some (hopefully minimally burdensome) ideas for oral historians who want to make their work more accessible to other research areas, as well as other historians.

Fortunately, oral history interviewers usually gather basic participant metadata at the beginning of interviews. However, this information is not always extracted, recorded, or structured as metadata. Such information may be automatically extracted with an information extraction AI tool, assuming one has been pretrained on that language or language variety (another reason to prepare at least some separate metadata records is that structured metadata would allow such a tool to be trained or fine-tuned to a particular collection). We are currently exploring AI information extraction to rediscover key items in the narratives. In the meantime, since many of the narrators are public figures in their communities, we were able to recover some information on the internet. When it came to recovering information about the interviewers and transcriber, in many cases, we only had the oral historians' memories. Often,

we only found the interviewer's name, and it was not clear which transcribers worked on which narratives.

Beyond their basic initial questions (name, year of birth, place of residence, etc.), we encourage oral historians to collect detailed demographic information (we are aware that the metadata varies by organization—the Veterans History Project at the Library of Congress is one example of thorough metadata collection)[32]. We strongly encourage oral historians to keep separate, structured demographic records not only of the narrators but also of the interviewers and transcribers. Not all essential sociolinguistic metadata can be captured every time, but every bit enhances the value of oral histories for scholars in social sciences and allows for increased utilization of oral history collections. As an example, we draw attention to the metadata fields in the image in figure 1. This is a screenshot of LaMeta, a language documentation software to record metadata and organize audio and text files.[33]

Best practices in linguistics demand keeping metadata records about the audio recorder, microphones, and their settings. Our linguists noted that oral history recording methods are similar to linguistics field recording. However, the expected technical and situational metadata was missing from the interviews we studied. Fortunately, basic technical metadata such as stereo versus mono tracks and sampling rate can often be extracted from original digital audio files. Other technical information is key to understanding how the act of recording affects documented speech, including the microphone's polar patterns and frequency range. This information may be recoverable if the microphone's brand and make has been documented.

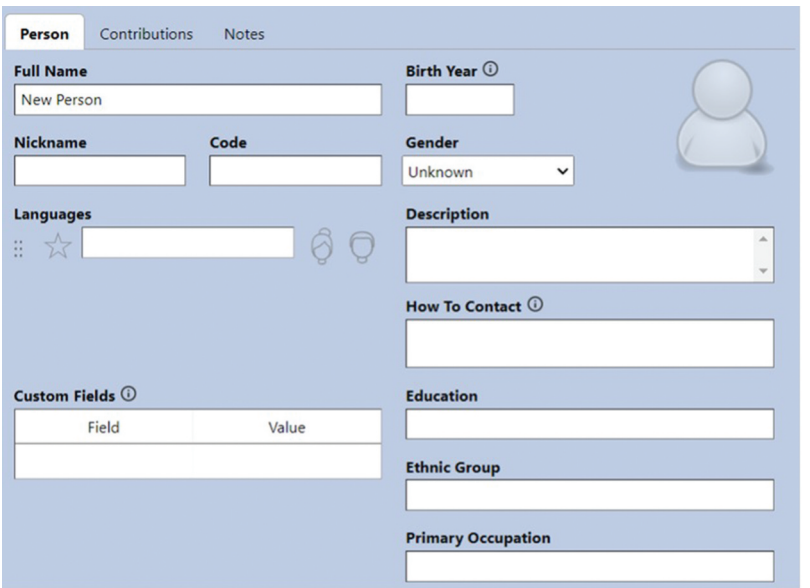

Figure 1. Screenshot of LaMeta in the process of adding a new participant to a corpus.

Much useful information can be conveyed by taking a photo of the placement of recording equipment. We are aware that oral historians sometimes take photos of interviewees, but these images are generally focused on individuals. Although that focus may address some of the sociolinguistic questions discussed above, such as the relative ages of the narrator and interviewer and possible ethno-linguistic differences, it does not reveal information about the recording event itself. One or two pictures would ideally include the microphone placement relative to the interviewer and narrator. Also, a photo that shows the other people who were present during the narration reveals factors that may influence the narrator's unconscious linguistic choices.

Avenues of Collaboration: Tackling Transcription of AAL

Careful, consistent transcription simplifies the adaptation of archival material for other researchers. Clearly documented guidelines for transcription could improve efficiency on two fronts. First, transcription standards allow other researchers to develop uniformity in representations across and within texts. Second, it allows linguists to pinpoint features in the speech that contribute most to distinguishing between language varieties without requiring them to relisten and retranscribe. Transcription improves the accessibility of oral histories by allowing us to read and search textual representation of speech. Linguists also rely on textual representation of speech to conduct qualitative and quantitative analyses. Despite the overlap, the first step our linguists had to undertake was checking and sometimes completing the transcripts. In our case, this also meant devising a way to handle AAL consistently and accurately in the transcription process.

Brief Overview of Approaches to Transcription

Broadly speaking, orthographic transcription (i.e., transcription that uses a standardized writing system) can be divided into three categories: verbatim, edited, and intelligent. Verbatim transcription is quite meticulous and includes all utterances, such as filler words (like "um," "yeah," and "huh") and non-verbal communication (laughter, coughing, sighing). This requires great concentration, but it can be helpful for capturing paralinguistic aspects in conversations—such as interruptions, pauses, pitch, or speech modulation —that might be specific to a language variety like AAL. For instance, CORAAL specifically annotates the teeth-sucking sound with "ts." Verbatim transcription seems rare in oral history, possibly due to the great amount of time it takes to capture every nuance, and because the "likes" and "uhs" are distracting.

Edited and intelligent transcriptions keep the narrative content as the main focus. Edited transcription allows some freedom to ignore false starts, stutters, repeated words, or obvious unintentional misspeaks, which can make reading the interview flow more smoothly. Intelligent

transcription allows greater freedom to refactor the speech, including to correct the narrator's grammar (however they interpret "correct grammar"), which may be seen as a way of respecting the narrator by presenting their speech as something more akin to polished written language. For example, one oral history project's guidelines states, "The transcript should not embarrass the narrator or make him or her look foolish . . . verb tenses should be made consistent, and *subjects and verbs should agree*" (emphasis added).[34] Another guide prohibits reduced forms and slang, stating that "slang such as 'y'all' is acceptable—very occasionally—if that's what was spoken, although it should not be used extensively for regional approximations à la Mark Twain."[35] Such freedom to edit "intelligently" may result in removing notable regional words (e.g., y'all/you guys/youse) or expressive dialectal grammar (e.g., AAL person-number disagreement). With perceptions of AAL being "broken" English rather than a complete and complex linguistic system, this freedom can lead to the erasure of people's lived experiences as well as linguistically significant information. Note that such mismatches between transcript and speech are not what might result from poor audio quality.[36] All in all, our linguists found that edited transcription seems most common in oral history guidelines. However, even when an oral history program discourages significant changes from the spoken word, such as SPOHP does, the oral historians on our team report that they acquiesce to requests from narrators to make their speech more polished on the written page.

Linguists aim for an accurate snapshot of each speaker's language and employ fine-grained transcription to do this. Phonetic or phonemic transcription is the most careful, aiming for an exact replication of individual speech sounds or how sounds change in context. Both use the International Phonetic Alphabet (IPA), which contains symbols for every sound known to exist in any language. Coarser-grained linguistic transcription is orthographic, meaning the language's standardized writing system is used, if one exists. A wide range of details may be included that roughly mirror the differences between verbatim and edited/intelligent transcripts. However, even linguistic transcripts tend to prioritize accurately representing dialectal variations over conscientiously following a standardization, which is usually based on a dominant spoken variety that may no longer exist and typically fail to reflect how anyone actually speaks.

Issues with Transcription of AAL

Transcription practices cannot be generalized and then applied without consideration of the unique traits of a language. AAL does not have an established orthographic standardization, which presents additional problems for

transcribing not just AAL pronunciation but its distinctive and expressive grammatical system.[37] Written representation of AAL has often proved to be a challenge for legal transcribers, linguists, and oral historians alike.[38] Differing representations pose issues for the recognition of AAL. For example, a transcriber who is not versed in AAL grammar may effectively remove indications that a narrator used AAL, such as transcribing the existential *it*/*dey* construction "*it was* so many of us today" as "*there were* so many of us today." Differing representations also leave room for stereotypical representation of the language. Although not specifically AAL (but related to Black language more widely), Sojourner Truth's famous speech was altered in writing by Frances Gage and published with embellishments that were thought to be reflective of Truth's African American dialect, even though Truth was in fact a native Dutch speaker.[39] These same challenges are not necessarily faced with General American English, so it is important that transcription guidelines consider differences between standardized written varieties and less dominant spoken varieties.

Interestingly, as the language continues to develop in online spaces, users are increasingly more creative in making realistic representations of verbal communication. The use of AAL is widespread across the digital landscape. This can be seen on social media sites such as X (formerly known as Twitter) and within podcasts and vlogs.[40] In an examination of 250 AAL tweets, we found various phonological and syntactic features of AAL. Alternate spellings of common words reflect AAL phonological patterns: for example, "da" for "the," "dat" or "dhat" for "that," "dis" or "dhis" for "this," "ion" or "iont" for "I don't," and "ova" for "over." These tweets also demonstrate syntactic features of AAL. For example, the null copula is apparent in "If u wit me den u pose to RESPECT ME" (which corresponds to "If you (are) with me, then you (are) supposed to respect me") because of the absence of the linking verb "are."[41]

Despite the prevalence of AAL in the digital space, computer tools remain quite intolerant of inconsistencies. Thus, we feel that the lack of standardization in written AAL should be corrected, at least by academics handling examples of the language. AAL has distinctive features, and one role of linguists—hopefully in tandem with oral historians—is to craft methods of representing these features consistently in text so they can be identified and studied. Even so, with the vibrancy coming from the language's diverse representations, it can be difficult to strike a balance between authenticity and accessibility.

Our Approach to AAL Transcription

An ideal representation of AAL ensures that its unique aspects of speech are preserved for historical, technological, and educational purposes.[42] If the nuances of AAL's grammatical structures and vocabulary are not documented accurately, then AAL could be at a disadvantage for receiving better

understanding in language science and stronger support in language technology. Though AAL presents challenges toward the goal of accurate transcriptions, efforts are being made to rectify them. The Oxford Dictionary of African American English team, for example, is attempting the large task of providing some regularity for written AAL.[43] The creators of CORAAL wrote a transcription guideline that details its standardizations, adapted from the Sociolinguistic Archive and Analysis Project (SLAAP).[44]

Our team of linguists and oral historians created a transcription guide, shown in table 1. Our standard avoids specialized spelling to reflect AAL pronunciation as used in CORAAL (e.g., *bin* instead of *been* for the remote past tense auxiliary) because we want to minimize the learning curve. Our aim is to provide transcriptions that are accurate as to the distinctive AAL grammatical features, but we felt that requiring transcribers to remember unique spelling would be overly burdensome to the average SPOHP transcribers and readers.

Table 1. AAL Feature Examples and Transcription Guidelines

Feature	SPOHP Transcription	Transcription Notes	Linguistic Notes (Recognition of Features)
Existential *it*/*dey*	"It's some coffee in the kitchen" "They got some coffee in the kitchen"	Do not translate to: "There is . . . "	
No copula (is/are)	"They walking too fast"[1] "Bob here?"[2]	Do not translate to: "They are/Is X . . . "	
Absence of third-person singular *-s*	"She walk to the store" "He ask for their number"	Do not translate to: "She walks/He asks . . . "	In AAL, the third-person present-tense marker may not be pronounced. Transcribe as heard.
Remote past *bin* (been)	"She been running"[3] "I been knowing he died"[4]	Do not translate to: "She has been/I already knew . . . "	*Been* and *bin* do not have an orthographic distinction necessarily, but stress on "been" in speech indicates the use of *bin*.
Habitual *be*	"I always be looking for somewhere to waste time"[5] "They probably be up there laughing"[6]	Do not translate to: "I am always looking/They are probably . . . "	
Perfect *done*	"I told him you done changed"[7] "I done pushed it"[8]	Do not translate to: "I already told him/I pushed it . . . "	"Done" in these sentences indicates that the situation/event has been completed or ended.

[1]Lisa J. Green, *African American English: A Linguistic Introduction* (Cambridge University Press, 2002), 40.
[2]Green, African American English, 42.
[3]Green, African American English, 55.
[4]Green, African American English, 56.
[5]Green, African American English, 51.
[6]Green, African American English, 51.
[7]Green, African American English, 60.
[8]Green, African American English, 60.

Outcomes of Collaboration: Analysis of AAL Grammatical Features

The interviews within the JBA provide a relatively sharp snapshot of what AAL sounds like in the Gulf South. We identified AAL features in the interviews as groundwork for further study. We wanted to learn how AAL grammatical features are distributed in semiformal speech. A primary analysis that included some annotation quality control is presented here. This serves as an example of how oral histories provide insight into human language.

Student assistants, who were trained and tested for at least 90 percent proficiency at recognizing the AAL features, have annotated 220 interviews. Looking at table 2, we can see the relative distribution of AAL grammatical features that appear in the JBA interviews. Person and number disagreement (lack of *-s* on third-person singular verbs) is the most frequent feature, comprising a third of the total features. Null copula is also frequent, followed by multiple negation. Based on this preliminary analysis, we posit that these three constructions are the most prominent grammatical features of AAL. Four of the features that are arguably unique to AAL (existential *it*/*dey*, perfect *done*, habitual *be*, and remote past *bin*) are the least frequent, comprising about 10 percent of the total features found. It may be because these features are the most semantically constrained, being limited to specific contexts. For instance, the habitual *be* expresses actions that occur regularly or habitually. It is not about a specific moment but rather a recurrent pattern. Similarly, the use of *bin* in AAL refers to the remote past—something that happened a long time ago—rather than simple past tense. Both features are used to describe a more specific time of an event. Person and number disagreement, null copula, and multiple negation are not tied to specific, context-dependent meanings or limited to particular tenses or actions. Instead, these features are more general in their usage and can be applied across a wider variety of sentences and contexts.

Remote past *bin* (been) is the rarest unique AAL feature used by the narrators, and early on, we noticed that annotators had difficulty identifying when this feature appears. Moreover, in an introductory computational linguistics course at UF, undergraduate and graduate students listened to a one-hour lecture about the AAL grammatical features and then were each assigned to annotate a text for these features. Surprisingly, the students who were primarily

Table 2. AAL Grammatical Feature Counts

AAL Grammatical Feature	Count in 220 Interviews
Person/number agreement	3,864
Null copula	2,674
Multiple negators	1,289
Existential *it*/*dey*	560
Perfect *done*	176
Habitual *be*	118
Remote past *bin*	82
Total	**8,763**

native speakers of General or Southern American English often missed the null copula. This suggests that, although the null copula is very frequent in the data, it may not be as cognitively prominent to speakers of other English varieties. We suspect therefore that null copula is one of the more complex AAL features and warrants deeper study.

Our analysis shows that AAL features are not used indiscriminately. The distribution of these features helps us better understand the ways in which many African Americans use language. Current team members are using this within-individual breakdown to determine whether any correlations exist between features or whether the presence of the most frequent or unique AAL features influence the presence of other features. This opportunity to study the language in detail would not exist without the work of oral historians.

Conclusion

Discovering information about AAL as we did from just a modest number of oral histories reveals the possibilities for oral history archives to boost research in linguistics. No longer limited to self-collected data in small amounts or social media sites that contain unverifiable sources, linguists can address questions about different populations of speakers. Using oral histories has enabled us to examine speech features of AAL speakers across the Gulf South for which no linguistic corpus exists, certainly none of this magnitude and diversity.[45]

For the oral historians on this project, this collaboration and resulting feedback on transcription processes and representation of nonstandard language varieties has been valuable. As public institutions and private industry race toward investment in AI, oral historians find themselves compelled to determine new rules of acceptability regarding ethics, process, and research utility. In July 2024, Baylor University and the Oral History Association convened a conference on AI and oral history, the first such gathering of its kind, to grapple with these emerging challenges. Panelists from across the globe presented on the tensions and applicability of AI in oral history, noting the benefits and risks of AI in a series of project case studies. What became evident is that, while many (if not most) oral history programs currently incorporate AI in their workflow, the field has not yet developed a consensus on specific best practices for AI usage.[46]

Our linguistics and oral history collaboration brings into focus a few AI-inflected points for our program and practice to consider on a broader scale. For SPOHP, AI is integral in processing interviews. The program uses AI software to generate transcript drafts, thus cutting down on transcription time and making the process more efficient. The linguists' work leads to new language technology, fine-tuned to a particular language variety or collection.

Moreover, oral historians make privacy commitments to narrators, and, as our linguistics team has shown, any other data embedded in the interview may yet be revealed. JBA represents a deep well that is ripe for automated methods of information

extraction. Many oral historians rightfully worry that treating archives as data farms presents serious ethical dilemmas that violate the trust and stewardship mandate inherent in our practice and programs. For SPOHP, this collaboration has prompted a renewed interest in revisiting our consent forms and exploring new legal agreements that might preserve our obligations to our narrators while leaving room for AI innovations. This collaboration between SPOHP and the linguistics team offers an opportunity to calculate some of the emerging risks of engaging AI for oral history while helping us determine what guardrails we might develop and put in place to retain the human voice in the practice, process, and product of oral history.

In conclusion, for collaborations between oral historians and linguistic researchers to succeed, particular attention must be paid to different aspects of the collection process in oral history and the research process in linguistics. Initially, the goals of both fields should be outlined, such as whether a project is education-oriented or geared toward NLP endeavors, which could change the study design. Conducting any necessary preprocessing will take labor that can be coordinated to the benefit of both oral historians and linguists. Team members must decide how to standardize the language of the transcripts they work with and determine what types of demographic metadata will be analyzed. Oral history is a crucial component of increasing public knowledge of AAL and of its speakers. Finally, we believe that our project provides insight into the particular features and patterns of AAL speakers in the Gulf South and that our work can therefore serve as a blueprint for further research in other regional variants of AAL. Any efforts that champion linguistic authenticity by staying true to the stories shared in oral history collections and to the language in which they are shared will surely contribute comprehensive data and a wealth of education for generations to come.

Acknowledgments

This work was supported by the National Endowment for the Humanities under grant PW-277433-21 and by the American Council of Learned Societies under the "Reanimating African American Oral Histories of the Gulf South: Enhancing Public Engagement" grant.

CRediT Statement

We follow the Contributor Roles Taxonomy (CRediT) (https://credit.niso.org/).
Conceptualization: MS, AH, AD, RAC, SM, KT, WD
Data curation: AD, CJP, PH
Funding acquisition: PO, SM, KT
Investigation: AD, CJP, PH
Project administration: AD, JH, RAC, WD
Software: PH, CJP, WD
Resources: PO, DH, SM
Supervision: MS, AH, AD, JH, RAC, KT, SM
Validation: PH

Visualization: AD, PH, SM
Writing—original draft: AD, AR, RAC, DH, PO, SM, KT, WD
Writing—review and editing: MS, AH, SM, KT

Disclosure Statement

No potential conflicts of interest are reported by the authors(s).

Funding

This work was supported by the National Endowment for the Humanities [PW-277433-21].

ORCID

Matthew F. Simmons http://orcid.org/0009-0006-0235-5253
Alexis Davis http://orcid.org/0000-0001-6096-3229
William Dyer http://orcid.org/0009-0005-7470-9527
Adolfho Romero http://orcid.org/0009-0000-2505-0192
Kevin Tang http://orcid.org/0000-0001-7382-9344
Sarah Moeller http://orcid.org/0000-0003-3612-2811

Notes

1. "Visual History Archive," USC Shoah Foundation, https://vha.usc.edu/home; "Oral History Collection," United States Holocaust Memorial Museum, https://www.ushmm.org/.
2. Michael Rießler and Joshua Wilbur, "Documenting Endangered Oral Histories of the Arctic: A Proposed Symbiosis for Documentary Linguistics and Oral History Research, Illustrated by Saami and Komi Examples," in *Oral History Meets Linguistics*, ed. Erich Kasten et al. (Verlag der Kulturstiftung Sibirien, 2017), 31–64.
3. We use the term African American Language (AAL) rather than African American Vernacular English (AAVE) throughout this article. While AAVE remains widely used in sociolinguistics, some scholars prefer to reflect the broader and more dynamic nature of this linguistic system. This shift emphasizes AAL's role as a fully developed linguistic system shaped by a distinct cultural and historical experience, rather than defining it solely in contrast to "standard" English. We also refer to Mainstream American English (MAE) as a descriptor for the socially dominant variety of English in the US, also called Standard American English (SAE). It is often treated (though problematically) as the linguistic norm and is contrasted with AAVE/AAL. For further discussion, see Sharese King, "From African American Vernacular English to African American Language: Rethinking the Study of Race and Language in African Americans' Speech," *Annual Review of Linguistics* 6 (2020), https://www.annualreviews.org/content/journals/10.1146/annurev-linguistics-011619-030556.
4. Benny Goodman, interview by Cindy Bobadilla, July 14, 2012, transcript, p. 11, Samuel Proctor Oral History Program Collection, P. K. Yonge Library of Florida, University of Florida.

5. Frank Seifart et al., "Language Documentation Twenty-Five Years On," *Language* 94, no. 4 (2018): e324–e345, https://doi.org/10.1353/lan.2018.0070.
6. "MUNFLA Brochure," Memorial University Folklore and Language Archive (MUNFLA), Memorial University, https://www.mun.ca/munfla/media/production/memorial/academic/faculty-of-humanities-and-social-sciences/munfla/documents/MUNFLA%20Brochure%20MAR%2031%20(print).pdf.
7. "Joel Buchanan Archive of African American Oral History," Samuel Proctor Oral History Program, George A. Smathers Libraries, University of Florida, https://ufdc.ufl.edu/collections/ohfb; "Oral Histories," Digital Library and Collections, Yiddish Book Center, https://www.yiddishbookcenter.org/collections/oral-histories.
8. Emily M. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2021), 610–23, https://doi.org/10.1145/3442188.3445922.
9. Andrew D. Selbst et al., "Fairness and Abstraction in Sociotechnical Systems," in *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Association for Computing Machinery, 2019), 59–68.
10. Christine Holmes, interview by Marna Weston, February 19, 2010, p. 29, Samuel Proctor Oral History Program Collection, P. K. Yonge Library of Florida, University of Florida.
11. Erich Kasten et al., eds., *Oral History Meets Linguistics* (Verlag der Kulturstiftung Sibirien, 2017).
12. Rießler and Wilbur, "Documenting Endangered Oral Histories of the Arctic," 31–64.
13. Laurin Wagner et al., "CrisperWhisper: Accurate Timestamps on Verbatim Speech Transcriptions," in *Proceedings of Interspeech 2024* (2024), 1265–69.
14. Alec Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," preprint, arXiv, December 6, 2022, https://doi.org/10.48550/arXiv.2212.04356.
15. William Labov, *Language in the Inner City: Studies in the Black English Vernacular* (University of Pennsylvania Press, 1972).
16. Francisca Pessanha and Almila Akdag Salah, "A Computational Look at Oral History Archives," *ACM Journal on Computing and Cultural Heritage* 15, no. 1 (2021): 1–16.
17. Lisa J. Green, *African American English: A Linguistic Introduction* (Cambridge University Press, 2002).
18. Santiago Harrison et al., "Disambiguation of Morpho-Syntactic Features of African American English—the Case of Habitual Be," in *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion* (2022), 70–75.
19. Holmes interview, p. 29.
20. Carnell Henderson, Henry W. Jones, and Jettie Henderson, interview by Anna Brodrecht and Memree Stuart, June 8, 2010, Samuel Proctor Oral History Program Collection, P. K. Yonge Library of Florida, University of Florida.
21. Alvin Butler, interview by Bria Shea, June 2, 2010, Samuel Proctor Oral History Program Collection, P. K. Yonge Library of Florida, University of Florida.
22. Holmes interview, p. 23.
23. Sonja L. Lanehart, "CORAAL Is Only the Beginning," *American Speech* 94, no. 1 (2019): 6–8.
24. Ralph W. Fasold, *Tense Marking in Black English: A Linguistic and Social Analysis* (Center for Applied Linguistics, 1972).
25. "Joel Buchanan Archive," Samuel Proctor Oral History Program, University of Florida, https://oral.history.ufl.edu/joel-buchanan-archive/; "Joel Buchanan Archive of African American Oral History," https://ufdc.ufl.edu/collections/ohfb.

26. "Samuel Proctor Oral History Program," University of Florida, https://oral.history.ufl.edu/.
27. Howard Giles et al., "Communication Accommodation Theory: Past Accomplishments, Current Trends, and Future Prospects," *Language Sciences* 99 (2023): 101571, https://doi.org/10.1016/j.langsci.2023.101571.
28. Charles Bryant, interview by Justin Dunnavant and Justin Hosbey, April 25, 2014, transcript, p. 26, Samuel Proctor Oral History Program Collection, P. K. Yonge Library of Florida, University of Florida. In this case, not only did the family member elicit a deeper response from the interviewee but Bryant's grandchild also encouraged the interviewees to interview Mary Bryant (Charles Bryant's spouse) as well, which they did.
29. Allen Cooper, interview by Paul Ortiz, August 6, 2004, transcript, p. 1, Samuel Proctor Oral History Program Collection, P. K. Yonge Library of Florida, University of Florida.
30. Holly Werner-Thomas, "Is Oral History White? The Civil Rights Movement in Baltimore, an Oral History Project from 1976, and Best Practices Today," *Oral History Review* 49, no. 2 (2022): 377–98, https://doi.org/10.1080/00940798.2022.2096474.
31. Sarah Moeller et al., "Compiling a Corpus of African American Language from Oral Histories," *Research in Corpus Linguistics* 12, no. 2 (2024): 45–79, https://doi.org/10.32714/ricl.12.02.04.
32. "About," Veterans History Project, Library of Congress, https://www.loc.gov/programs/veterans-history-project/about-this-program/.
33. John Hatton et al., "Lameta," https://sites.google.com/site/metadatatooldiscussion/home.
34. Minnesota Historical Society, *Editing an Oral History Transcript* (n.d.), https://www.mnhs.org/sites/default/files/library/research/oral-history/editing_oral_history_transcript.pdf.
35. Liz H. Strong, *Oral History Transcription Style Guide*, ed. Mary Marshall Clark and Caitlin Bertin-Mahieux (Columbia University Center for Oral History Research, 2018), https://static1.squarespace.com/static/575a10ba27d4bd5d7300a207/t/5a5f95a2652dea3cec32aacc/1516213672572/CCOHR+Transcript+Style+Guide+2018.pdf.
36. Sharese King and Katherine D. Kinzler, "Op-Ed: Bias Against African American English Speakers Is a Pillar of Systemic Racism," *Los Angeles Times*, July 14, 2020, https://www.latimes.com/opinion/story/2020-07-14/african-american-english-racism-discrimination-speech.
37. Green, *African American English*; Taylor Jones, "Toward a Description of African American Vernacular English Dialect Regions Using 'Black Twitter,' " *American Speech* 90, no. 4 (2015): 403–40, https://doi.org/10.1215/00031283-3442117; Taylor Jones, "Tweets as Graffiti: What the Reconstruction of Vulgar Latin Can Tell Us About Black Twitter," in *English in Computer-Mediated Communication—Variation, Representation, and Change*, ed. Lauren Squires (De Gruyter Mouton, 2016), 43–68, https://doi.org/doi:10.1515/9783110490817-004; Taylor Jones, "African American English Intensifier *Dennamug*: Using Twitter to Investigate Syntactic Change in Low-Frequency Forms," in "Language and Computation," special section, *Frontiers in Artificial Intelligence* 5 (2022): 683104, https://doi.org/10.3389/frai.2022.683104.
38. Patricia Fulton, "The Memorial University of Newfoundland Folklore and Language Archive," *CAML Review* 30, no. 1 (2002), https://doi.org/10.25071/1708-6701.2932.
39. "Sojourner Truth," Biography.com, Hearst Digital Media, updated January 6, 2021, https://www.biography.com/activists/sojourner-truth.
40. Su Lin Blodgett et al., "Twitter Universal Dependency Parsing for African-American and Mainstream American English," in *Proceedings of the 56th Annual Meeting of the*

Association for Computational Linguistics, ed. Iryna Gurevych and Yusuke Miyao (Association for Computational Linguistics, 2018), 1415–25, https://doi.org/10.18653/v1/P18-1131.

41. Blodgett et al., "Twitter Universal Dependency Parsing for African-American and Mainstream American English."
42. Charlie Farrington and Tyler Kendall, "CORAAL," Online Resources for African American Language, 2023, https://doi.org/10.7264/1AD5-6T35.
43. "Oxford Dictionary of African American English (ODAAE)," Hutchins Center for African and African American Research, Harvard University, https://hutchinscenter.fas.harvard.edu/odaae.
44. Tyler Kendall and Charlie Farrington, eds., *CORAAL User Guide*, June 2023, http://lingtools.uoregon.edu/coraal/userguide/CORAALUserGuide_current.pdf.
45. Sophie Groenwold et al., "Investigating African-American Vernacular English in Transformer-Based Text Generation," in *Proceedings of EMNLP* (2020), 5877–83, https://doi.org/10.48550/arXiv.2010.02510.
46. In July 2024, the Oral History Association held a virtual symposium about the effects of generative AI on the practice of oral history. For more information, see "AI in OH: How New and Evolving Technologies Will Impact the Profession," Symposia, Oral History Association, July 2025, https://oralhistory.org/ai/.

Matthew F. Simmons is an assistant professor of history at Emmanuel University. He is a former graduate coordinator for the Samuel Proctor Oral History Program. He studies the intersection of labor and race in the twentieth-century rural American South and its implications for social movement building. His research has been published in *Tampa Bay History* and *The Chronicles of Oklahoma*. He is also a podcast host of the American South channel on the *New Books Network*.

Anna Hamilton is the assistant director of the University of Florida's Samuel Proctor Oral History Program and a PhD candidate in American studies at the University of North Carolina at Chapel Hill. Her research in foodways and environmental humanities engages oral history to better understand place-making and migration associated with the climate crisis. She is also a journalist and cofounder of The Marjorie, a Florida-based reporting outlet covering social justice and environmental issues.

Alexis Davis is a PhD candidate in linguistics at the University of Florida. Centering his research on topics related to African American Language, his expertise lies in improving sociolinguistic methodologies and exploring the connections between artificial intelligence, language technology, Black speech, and Black communities. His most recent work is found in *Research in Corpus Linguistics*.

William Dyer holds a PhD in linguistics from the University of Florida. His dissertation was on the syntax of relative clauses in Sereer. His most recent work has been published in *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*.

Christina Jean Philippe holds a BA from the undergraduate linguistics program at the University of Florida. Her interests lie in artificial intelligence, machine learning, and African American Language.

Jessica Heady is a PhD student in linguistics at the University of Florida with research interests in syntax, semantics, African American Language, and African American Sign Language. She is Project Coordinator for the Reanimating African American Oral Histories of the Gulf South

Project and is a Research Assistant for the Machine Learning for Endangered Language Documentation (MELD) Lab at UF.

Adolfho Romero is pursuing a PhD under the Industrial Labor Relations program at Cornell University. His scholarly pursuits are anchored in Latinx studies, migration and transnational studies, oral history, and ethnography, reflecting his interest in working-class issues and cultural narratives. Adolfho's previous research centered on the efficacy of grassroots movements, explored through his MA research titled *¡Si se puede! Yes, We Can! An Analysis of a Non-Profit Organization and Its Effectiveness: The Farmworkers Association of Florida*.

Rebekah Anne Cordova is a teacher/educator based in Florida and collaborates with both the Samuel Proctor Oral History Program and the linguistics department at the University of Florida. She leads the All Y'all Social Justice Collective, which examines social justice issues in the American South.

Deborah Hendrix is the digital humanities production coordinator for the Samuel Proctor Oral History Program. She specializes in organizing quality interview recording, archive management, and assisting students and faculty with oral history matters at the University of Florida.

Paul Ortiz is professor of labor history at Cornell University. Between 2008 and 2024, Ortiz was a history professor and director of the Samuel Proctor Oral History Program (SPOHP) at the University of Florida. Under his leadership, SPOHP created the Joel Buchanan African American Oral History Archive housed at the University of Florida Digital Collections, which contains most of the interviews referenced in this article. Ortiz is the author of *Emancipation Betrayed: The Hidden History of Black Organizing and White Violence in Florida from Reconstruction to the Bloody Election of 1920* (University of California Press, 2005).

Pryce Houck is pursuing a BS in computer science at the Herbert Wertheim College of Engineering, University of Florida, with a minor in linguistics. He is an undergraduate research assistant in Dr. Moeller's Machine Learning for Endangered Language Documentation (MELD) Lab, where he focuses primarily on maintaining data sets and deploying human-in-the-loop classifier models to annotate grammatical features of African American English. Houck's recent research has been published in the *2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics (NAACL 2025)*.

Kevin Tang is a university professor in English linguistics with specializations in phonetics/phonology/morphology at the Faculty of Arts and Humanities, Heinrich Heine University Düsseldorf and a courtesy assistant professor in computational language science at the University of Florida. His more recent publications have focused on African American Language and its intersection with language technology. He has authored articles in journals including *Association for Computational Linguistics, Language Resources and Evaluation (LREC), Computational Linguistics (COLING), Cognition, Laboratory Phonology, Language and Speech*, and the *Journal of the Acoustical Society of America*.

Sarah Moeller is an assistant professor of linguistics at the University of Florida. Her research focuses on applying artificial intelligence to the study of underdocumented languages and improving natural language processing (NLP) performance on low-resource languages and language varieties. She serves as president of the Association for Computational Linguistics Special Interest Group for Endangered Languages (ACL-SIGEL). Moeller's recent relevant research has been published in *Research in Corpus Linguistics, Language Resources and Evaluation (LREC)*, and *Computational Linguistics (COLING)*.