Data is patrimony: on developing a decolonial model for access and repatriation of sociolinguistic data

Nicté Fuller Medina

This paper discusses the first dataset installment of the project, Language, Culture and History: Belize in a Digital Age, which focusses on the recovery, preservation and repatriation of legacy sociolinguistic data to Belize. This first dataset, the Older Recordings of Belizean varieties of Spanish, are a collection of sociolinguistic interviews carried out in the late 70s in Mestizo-Maya communities. These interviews are a record of language as it was used at an earlier time as well as narratives of community folklore and cultural beliefs. Re-imagining these data as cultural patrimony and adopting a decolonial framework, this paper describes the intentional steps taken to disrupt the extractivism typical of both memory institutions and linguistics research. These steps include restitution and repatriation, local access, capacity building and challenging Western views of ownership and knowledge production. In sum, it centers communities at both the local and regional level which includes the choice to archive (and eventually make public) the collection via dLOC, a known entity in Belize and the region, allowing for a broader virtual and symbolic repatriation to the region.

Ancestral acknowledgement: Various interviewees whose voices are contained in the collection Older Recordings of Belizean varieties of Spanish are now with the ancestors. I recognize that I am not entitled to work with their voices but rather that, if I am able to work with these recordings, it is because I am being permitted to do so.

This essay discusses the current progress of the first dataset installment of the project *"Language, Culture, and History: Belize in a Digital Age."*[1] The project itself was inspired by interviews collected by Timothy Hagerty in the late 1970s for his linguistic analysis of the Spanish spoken in Belize.[2] These are now compiled in the collection Older Recordings of Belizean varieties of Spanish.[3] The interviews had been in his personal research collection for close to forty years before he transferred them to me in 2016. Initially my interest in the collection was data-driven and primarily sociolinguistic. As a linguist, I am interested in telling the human story of language—how language travels and changes as a result of human migration, how communities use language as a social mechanism, and the notion of language itself as an archive.[4] More specifically, as a Belizeanist sociolinguist, my research focuses on the particularities of the multilingualism found in Belize.[5] However, it quickly became evident that the types of narratives in the interviews would be of interest to a wider range of scholars with interests in culture, oral literary traditions, and history. Moreover, it was clear that the cultural knowledge contained in the interviews should be accessible to communities in Belize and the diaspora. As a consequence, this project reimagines sociolinguistic data as cultural patrimony. In contrast to traditional sociolinguistic corpus creation, where sociolinguists have been foregrounded in data organization and access,[6] the development of this resource decenters this user group and instead gives primary consideration to communities. The goal of creating a public resource that is usable for specialists such as linguists is balanced with the goals of creating a resource that is meaningful and accessible (through repatriation) to communities. In order to leverage existing infrastructure and the discoverability gained from the visibility of an established digital library, the Older Recordings of Belizean varieties of Spanish will be hosted online by the Digital Libraries of the Caribbean (dLOC). Since an online open access resource cannot be equated with repatriation, however, the goal is to return interviews to interviewee communities in Belize while arranging for local memory institutions to facilitate

---

[1] Nicté Fuller Medina, "Sociolinguistic Data as Cultural Patrimony: Challenges, Solutions and Lessons Learned in the Preservation of Legacy Data," paper presented at the *Caribbean Studies Forum, U*niversity of Belize, 4–11 March 2018; Fuller Medina, *"Cultural Safeguarding through Digital Repatriation and Community Language Mapping,"* paper presented at Belize National Research Conference*, University of Belize, Belmopan, 3–4 April 2019;* https://www.clir.org/2020/06/repatriating-sociolinguistic-data-to-belize-can-a-decolonial-model-still-work-in-times-of-covid-19/.

[2] Timothy W. Hagerty, "A Phonological Analysis of the Spanish of Belize" (PhD diss., University of California, Los Angeles, 1979).

[3] Fuller Medina, "Sociolinguistic Data as Cultural Patrimony"; Fuller Medina, "Cultural Safeguarding through Digital Repatriation and Community Language Mapping."

[4] Chris Ehret, *History and the Testimony of Language* (Berkeley: University of California Press, 2011).

[5] I use the term *Belizeanist* here to refer not only to my specialization but also to my stance in centering minoritized Belizean voices and my Belizean heritage, which affords me an insider perspective.

[6] David Nathan, "Access and Accessibility at ELAR: A Social Networking Archive for Endangered Languages Documentation," in Mark Turin, Claire Wheeler, and Eleanor Wilkinson, eds., *Oral Literature in the Digital Age: Archiving Orality and Connecting with Communities* (Cambridge, UK: Open Book, 2013), 21–40; Wesley Y. Leonard, "Producing Language Reclamation by Decolonising 'Language,'" in Wesley Y. Leonard and Haley De Korne*, eds., Language Documentation and Description*, vol. 14 (London: EL, 2017), 15–36, http://www.elpublishing.org/PID/150.

local on-site access.[7] Repatriation in this context includes both restitution—the return of cultural heritage or artifacts to communities or individuals—and repatriation, where the return is to a government or state agency.

This first installment, currently in progress, serves as a test case for a decolonial approach to the curation and repatriation of, as well as access to, legacy audio recordings from Belize. The work described here is also a possible path (and call) for community-centered corpus creation in sociolinguistics, at least with respect to legacy data. Taking a decolonial approach, this project sits at the intersection of critical archival studies, best practices in sociolinguistic corpus creation while drawing on decolonial theory, postcustodialism, the *Protocols for Native American Archival Materials*, and the principles of UNESCO guidelines on intangible cultural heritage.[8]

A Decolonial Approach

In the context of this essay, *decolonial* should be understood as an ideological stance and set of practices that counter coloniality, defined as the prevailing economic, political, and social structures that persist despite the end of the political system of colonialism.[9] One outcome of the colonial project is the practice of extractivism with a concomitant unidirectional flow of resources, along with the privileging of Western ontologies and epistemologies, subordinating or displacing Indigenous (and other non-Western) ones. *Extractivism*, in its most basic form, refers to the process where raw materials are removed from one place for enrichment and benefit of an outside, more privileged and powerful set of individuals, communities, or projects. This was typical during colonialism, for example, with the theft of precious metals and artifacts for the benefit of the European colonial centers. Similarly, both linguists and memory institutions (libraries, archives, and museums) participate in the removal of language data and records to the benefit of a small number of specialists, where both the institution and the specialists are usually located in colonial or imperial (Canada/US) centers.[10] This flow also takes place between marginalized communities and more privileged ones. Sociolinguistics, for example, has benefited greatly from data collected in African American communities, reflecting what John Russell Rickford describes as an "unequal partnership," since the benefit to African Americans

[7] In early conversations with the Institute for Social and Cultural Research for, example, they were open to hosting a digital copy of the collection. On the *Protocols*, see https://www2.nau.edu/libnap-p/index.html.

[8] UN Educational, Scientific, and Cultural Organization (UNESCO), "Ethics and Intangible Cultural Heritage," accessed 19 September 2020, https://ich.unesco.org/en/ethics-and-ich-00866. The Protocols for Native American Archival Materials can be accessed here: https://www2.nau.edu/libnap-p/protocols.html

[9] Aníbal Quijano, "Coloniality of Power, Eurocentrism, and Latin America," *Nepantla: Views from South* 1, no. 3 (2000): 533–80.

[10] J. Bastian, "Taking Custody, Giving Access: A Postcustodial Role for a New Century," *Archivaria: The Journal of the Association of Canadian Archivists* 53 (2002): 76–93; Hannah Alpert-Abrams, David A. Bliss, Itza Carbajal, Marika Cifor, and Jamie A. Lee, "Post-custodialism for the Collective Good: Examining Neoliberalism in US–Latin American Archival Partnerships," in Marika Cifor and Jamie A. Lee, eds., "Evidences, Implications, and Critical Interrogations of Neoliberalism in Information Studies," special issue of *Journal of Critical Library and Information Studies* 2, no. 1 (2019): 1–24.

has not been substantive enough.[11] A second element of persisting colonial methods is the deterministic nature of the ways memory institutions select what is archived and how it is archived. Likewise, linguists, in documenting language or creating corpora, determine (often from a Eurocentric perspective) how that is done and what aspects of language are most important.[12] The work of memory institutions as well as linguists is a highly political enterprise that has historically reinforced colonial hierarchies in privileging the documentation of particular voices, replicating Eurocentric interpretations of these voices and therefore shaping not only what stories are told but how they are told.

Since the practices described above are normalized, they become the default, such that extractivism often takes place whether it is intended or not and, at the same time, alternative approaches are not well supported. While many cultural heritage collections were stolen, appropriated under the violence of colonialism, or under the guise of missionary work (an important arm of the colonial project), many such items were simply obtained through default extractivist practices. This gives the general context of research and of the Older Recordings of Belizean varieties of Spanish. Data was collected through fairly standard practices of the time when Hagerty carried out his interviews. In fact, Hagerty took a number of steps that can be seen as countering the extractivist nature of research in the Global South, such as transferring his data to me, a Belizean and Belizeanist scholar, and publishing some of his work in Belize. Accordingly, repatriation as discussed here is not remedying a theft or an abuse as in cases where artifacts were stolen or remains are kept away from their communities of origin but rather addresses the outcomes of practices that may on the surface appear neutral.

The project spans several years and three different institutions, illustrating the multiplicity of challenges that emerge due to obsolete formats, limited funding, the impoverishment of the Global South, barriers to countering colonial frameworks, and the desire for participation in a more democratic open science. This *Language, Culture, and History* project is decolonial in that it intentionally subverts extractivism by disrupting the unidirectional flow of resources through repatriation and by shifting enrichment to local or related communities rather than individuals or communities in the Global North (here I include academic communities). Enrichment takes the form of community-focused capacity building, a future vision for community access, future teaching resources, and general collaboration (since we expect mutual learning to take place in this process). In addition, the goal is to present alternatives to default conceptualizations such as Western definitions of ownership and the view of the archive as static rather than agentive. Finally, by making these interviews available, we hope to challenge Eurocentric means of knowledge production, the base on which it rests, and its hegemony on voices deemed to be valid.

This essay is organized as follows: I first present a brief background on Belize, followed by a sketch of the decolonial approach taken here, after which I describe the work completed

---

[11] John Russell Rickford, "Unequal Partnership: Sociolinguistics and the African American Speech Community," *Language in Society* (February 1997): 161–97.

[12] Mary S. Linn, "Living Archives: A Community-Based Language Archive Model," *Language Documentation and Description* 12 (July 2014): 53–67; Wesley Y. Leonard, "Producing Language Reclamation by Decolonising 'Language,'" in Wesley Y. Leonard and Haley De Korne*, eds., Language Documentation and Description*, vol. 14 (London: EL, 2017), 15–36, http://www .elpublishing.org/PID/150.

to date. I then discuss the limits of Western concepts of ownership; capacity building as enrichment of local, diaspora, and related community members; and the multiple ways the collection creates space for voice and participation in the production of knowledge. In the final section I discuss the process to date as a dynamic one that emerged from the collection itself as a sentient, agentive archive in interaction with me as the curator and researcher.

Belize: In between Central America and the Caribbean

Belize is located on the Atlantic coast of Central America and shares borders with Mexico to the north and Guatemala to the west and south. Its particular history as a former British colony and its geographic location on the isthmus makes this nation-state both Central American and Caribbean, thus troubling definitions of both. Belize has English as its official language in practice and an English-lexified Creole, Kriol, as a major national language. Together, English and Kriol represent strong historical, social, and linguistic ties to the anglophone Caribbean, particularly to Jamaica, where the administrative center of the colony was located.[13] At the same time, more than half the country speaks Spanish,[14] and Maya languages spoken in Belize are also spoken in Guatemala (Mopan and Ketchi) as well as Mexico (Yucatec), while Garifuna, spoken primarily in southern Belize, is also spoken in Honduras, Guatemala, and Nicaragua, aligning Belize, culturally and linguistically, with other Central American nations as well as Mexico. Additional commonalities are found in the English and English-lexified Creoles spoken along the Caribbean coast Central America in Honduras, Costa Rica, Nicaragua, Panama, and Guatemala.[15] Nonetheless, as a former British colony, and as a result of the nation-making process where English, for example, was adopted as the official language of the constitution, Belize is constructed as anglophone. This effectively contributes to the erasure of Spanish-speaking Belizeans as well as the Maya and Garifuna, since this disregard for other languages in Belize represents a rhetorical construction of a (constitutional) identity rooted in colonial patterns of exclusion.[16] Since the historically large Black population in Belize unsettles the conception of Central Americans as Mestizx subjects, this erasure is also a racialized one. Thus, the construction of the Mestizx subject through national narratives renders Black Central American hispanophones and creolophones along the Caribbean coast of Central America invisible. These multilayered erasures are rooted in raciolinguistic ideologies resulting in the exclusion of these voices in narratives of the region and in various subfields of linguistics (language contact, creolistics, Hispanic linguistics), though this is slowly changing.

---

[13] Colville N. Young, *Language and Education in Belize* (Belize City: National Printers, 1995).

[14] Statistical Institute of Belize, *Belize Population and Housing: Census Country Report 2010* (Belmopan: Statistical Institute of Belize, 2013).

[15] John Holm, *Central American English* (Heidelberg, Germany: Julius Groos, 1983); Glenda A. Leung and Miki Loschky, eds., *When Creole and Spanish Collide: Language and Cultural Contact in the Caribbean* (Leiden, the Netherlands: Brill, 2021).

[16] Randy B. Marfield, "A Rhetorical Analysis of the 1981 Belizean Constitution: National Identity in the Context of History, Race, and Language" (PhD diss., East Carolina University, 2016).

Belizean varieties of Spanish,[17] for example, are largely undescribed in Hispanic linguistics and the language and cultural contact between Spanish and English-lexified Creoles in Central America is only recently gaining substantive attention.[18]

The dual status of Spanish as a form of invisibility is important to note here. While a colonial language, Belizean varieties of Spanish are stigmatized both within Belize and the region[19] and have been prohibited at different times along with Indigenous languages in favor of English.[20] In addition, Spanish is often codified as a foreign language and varieties perceived as standard are privileged.[21] Furthermore, that the language of the interviews was primarily Spanish may mask multiple identities as some interviewees identified themselves as multilingual speakers of Yucatec Maya, Spanish, English, and/or Kriol; as Yucatec Maya–dominant speakers; or as Spanish-English bilinguals. These interviews archive an ethnolinguistic complexity rooted in the colonial history and resistance of Belize, which have yet to be extensively problematized. Nevertheless, highlighting the exclusions above is not a call for communities to be further studied and certainly not from the perspective of the extractivist colonial gaze. Rather, it signifies that this region is largely overlooked in conversations where its voices should be present, particularly minoritized voices. The project described here is an attempt to bring some of these voices into conversation through the creation of an online repository of narratives from Belizean Mestizo-Maya communities.

The Language, Culture, and History Project

According to the United Nations, audiovisual materials "are extremely vulnerable . . . [and] can be irretrievably lost as a result of neglect, natural decay and technological obsolescence" yet "contain the primary records of the history of the 20th and 21st centuries."[22] This includes a record of language as it was spoken at an earlier point in time. Historical language data is invaluable for the scientific study of language change and variation, yet linguists must

---

[17] I define Belizean varieties of Spanish as the two varieties that were first established in the mid-nineteenth century. They include Western Belizean Spanish and Northern Belizean Spanish. Nicté Fuller Medina, "'Lo Que Hacen Mix es el Kriol y el English': How Spanish Speakers Reconcile Linguistic Encounters with English and Kriol in Belize," in Leung and Miki Loschky, *When Creole and Spanish Collide*, 126–54. Hagerty, "A Phonological Analysis"

[18] Leung and Loschky, When Creole and Spanish Collide.

[19] Fuller Medina, "'Lo Que Hacen Mix es el Kriol y el English'"; Nicté Fuller Medina, "Belizean Varieties of Spanish: Language Contact and Plurilingual Practices." Anuario de Estudios Centroamericanos: Dossier on Belize 46, no.1: (2020) 1-29. DOI:10.15517/AECA.V46I0.42202

[20] During interviews I conducted in 2013–14, more than one educator in Spanish-speaking communities indicated that they implemented a no-Spanish rule at their schools due to concerns that students would not learn English. Heriberto Cocom recounts in a 2016 interview with CaribNation TV that under British rule neither Spanish nor Maya languages could be spoken or taught in schools; https://www.youtube.com/watch?v=r0zvC31FYwY.

[21] Britta Schneider, "The Multiplex Symbolic Functions of Spanish in Multilingual Belize," in Leung and Loschky, *When Creole and Spanish Collide*, 253–76.

[22] United Nations, "World Day of Audiovisual Heritage," accessed 19 September 2020, https://www.un.org/en/events/audiovisualday/.

often rely on incomplete or nonrepresentative data as historical benchmarks.[23] Crucially, the interviews were collected at a time of ongoing language shift from Maya to Spanish in some Maya-speaking communities in Belize. Some of these communities are now undergoing a second shift, this time to English and Kriol.[24] These interviews will help elucidate the impact of this shift and any needed interventions. Furthermore, Belizean varieties of Spanish are virtually unknown in the field of Hispanic linguistics.[25] Thus, as a representative record of language in Spanish-dominant regions of Belize, these interviews are a key resource for linguistics research.

More than a linguistic benchmark, however, these interviews document a disappearing cultural heritage that includes Indigenous views of animals as shape-shifters and messengers, descriptions of forest spirit beings such as Tata Duende, and stories of the West African trickster spider, Anansi. They archive the relationship between people and the (super)natural world.[26] The narratives also reflect a particular African, Maya, and Spanish syncretism. Anansi traditionally appears in the folklore of African-descendant communities in the region but, in Belize, he surfaces in Mestizo folklore.[27] Likewise, Tata Duende, whose name comes from Maya (*tata*) and Spanish (*duende*), also appears in Creole folklore and that of various Belizean ethnic groups. No publicly available resource currently exists that gives insight into this oral tradition or that captures Spanish as it was spoken in Belize at an earlier time.

Accordingly, the value of the data contained in Older Recordings of Belizean varieties of Spanish (ORBS) underscores the imperative of recovering recordings completed in other communities in Belize. Unfortunately, however, the ORBS is not the only collection to languish in a personal research collection, as several linguists have done fieldwork in Belize.[28] Those recordings are also at risk of being lost, or, in practice, already lost, since these linguists are either retired, deceased, or about to retire. The reality is that even if individual sociolinguists create corpora that are portable and/or interpretable, preservation and long-term access are difficult

---

[23] Juan M. Hernández-Campoy and Natalie Schilling, "The Application of the Quantitative Paradigm to Historical Sociolinguistics: Problems with the Generalizability Principle," in Juan M. Hernández-Campoy and Juan C. Conde-Silvestre, eds., *The Handbook of Historical Sociolinguistics* (West Sussex, UK: John Wiley & Sons, 2012), 63–79.

[24] Angel Cal and Nicté Fuller Medina, "Revitalization of Yucatec Maya in Northern Belize: Training Yucatec Maya Teachers to Teach in Primary Schools of Maya Communities," University of Belize, unpublished manuscript, 2017.

[25] Hagerty, "A Phonological Analysis of the Spanish of Belize"; Fuller Medina, "'Lo Que Hacen Mix es el Kriol y el English.'"

[26] Gerardo Polanco, "Of Gods, Men, and Earth: An Ecocritical Exploration of Maya-Mestizo Alternative Spiritual Traditions in Belize and Their Influence on Belizean Literature," paper presented at Caribbean Studies Forum, University of Belize, 4–7 March 2018; Fuller Medina, *"Cultural Safeguarding through Digital Repatriation and Community Language Mapping."*

[27] Timothy W. Hagerty, "Race and Ethnicity in the Folklore of Belize," *Filología y Lingüística 24, no. 2 (1998)*: 239–43.

[28] These include but are not limited to: Robert B. Le Page and Andrée Tabouret-Keller, *Acts of Identity: Creole-Based Approaches to Language and Ethnicity,* 2nd expanded ed. (1985; repr. Brussels: Modulaires Européennes, 2006); Donald Winford, "Tense and Aspect in Belize Creole," in Hazel Simmons-McDonald and Ian Robertson, eds., *Exploring the Boundaries of Caribbean Creole Languages* (Mona, Jamaica: University of the West Indies Press, 2006), 1–49; Genevieve Escure, "Belizean Creole," in Susanne M. Michaelis, Phillip Maurer, Martin Haspelmath, and Magnus Huber, eds., *The Survey of Pidgin and Creole Languages*, vol. 1 (Oxford: Oxford University Press, 2013), 92–100.

to achieve and, where archives may provide such access, it does not ensure discoverability.[29] This is particularly true where there are no established mechanisms for ingesting research data with an appropriate archival institution. In at least one case, I learned that copies of interviews in the Winford Collection were left in Belize for local use and access.[30] Yet, the Belize copy of this collection was functionally lost as I could not initially locate it. After searching on multiple fieldtrips, through pure serendipity, I located the collection at the National Heritage Library in the captial, Belmopan, cataloged as oral histories in a larger collection of oral histories rather than as sociolinguistic interviews. Thus, the problem is twofold: legacy recordings are functionally lost where they remain hidden in personal research collections or where they are archived but not discoverable, in part, because information on their provenance has been lost over time. Both of these issues can be further tied to default practices concomitant with extractivism since such a model, by definition, does not attend to the concerns of local communities, and local communities themselves often do not have the resources either to enforce countermeasures or to support individual efforts to adopt more equitable approaches, such as ensuring local access. As a consequence, it became evident that multiple collections would benefit from (digital) preservation and strategies to ensure discoverability, access, and longevity.[31] The interest and work of the National Heritage Library in the Winford Collection further confirmed my view that the interviews constitute patrimonial heritage. I subsequently conceptualized the larger project, *Language, Culture, and History: Belize in a Digital Age,* which aims to procure legacy interviews and audiovisual materials currently hidden in linguists' personal research collections and to digitally preserve and repatriate them, while creating broad access for local communities and scholars of language, culture, and history. This reenvisioning of the collection put the need for partnering with community (broadly defined) front and center, in other words, consciously challenging default colonial practices and more specifically those t have shaped traditional sociolinguistic data management practice.

The Older Recordings of Belizean varieties of Spanish are the first installment of this umbrella project. The collection includes forty-two cassette tapes and six open reels. The interviews are mostly in Spanish, though there is at least one narrative in Yucatec Maya and some are in Belize Kriol (an English-lexified Creole). With the exception of four cassette tapes (not numbered), tapes are numbered to 100. Thus, the original collection likely consisted of one hundred tapes, indicating that more than half the cassettes have been lost over the past forty years. The collection also includes two hundred pages of Hagerty's typewritten partial transcripts. From 2016 to 2020, the legacy transcripts, tapes, and open reels were digitized.

---

[29] Gary F. Simons, "The role of metadata in the infrastructure for archival interoperation." Language and Linguistics Compass 8, no. 11 (2014): 486-494. I mean here, interpretable to other sociolinguists, who are generally the intended users in the creation of sociolinguistic corpora. Preservation generally takes places through sociolinguistic labs, which rely on individual research grants. Otherwise materials remain in personal research collections.

[30] Interviews were carried out by Donald Winford (Ohio State University) in the early 1990s. I requested access to the collection for a study or relative clauses in Belize Kriol and learned that analog copies on cassette tapes had been left in Belize.

[31] Other instances I am aware of include Ervin Beck's collection of Creole folk tales and songs archived at Belize Records and Archive Services (Beck, personal communication, 8 October 2018), and more recently researchers have approached the Institute for Social and Cultural Research for assistance in archiving their data in Belize (Rolando Cocom, personal communication, 29 September 2021).

As data was first transferred to me in 2016 while I was a PhD student, I created digital copies of most of the cassette tapes in .wav format using personal equipment. Later that year, I was given access to the open reel player and digitization equipment in the University of Ottawa School of Music and digitized open reels.[32] In 2017, students at the University of Belize in the Service through Research Program (SRP), described further below, created digital copies of the typewritten transcripts. Finally, some cassette tapes needed to be repaired in order to be digitized, digital transfer of open reels had to be redone, and preservation quality copies of tapes were needed. This work was completed in multiple sessions over the course of 2018–20 with assistance from the Library Preservation Studios at the University of California, Los Angeles (UCLA), when I was a CLIR fellow in data curation in Latin American and Caribbean Studies at UCLA Library.

Capacity building

In order to put the corpus online and to repatriate the data, several things needed to happen. This included decisions regarding permissions and/or anonymization of data as is typical in sociolinguistics; extraction of metadata to identify speakers should permissions be sought; linking legacy transcripts[33] to audio, as transcripts often did not include metadata on the corresponding audio file; creating time-aligned transcripts (which would be of most use to linguists); and deciding whether or not, as a legacy dataset, unredacted interviews could be put online. In addition, the work initially had to be completed without a budget and using technology that was locally accessible and available (or available to me personally). Technology was an important consideration since the ORBS is intended to serve as a case study and, in the spirit of centering local communities, local realities needed to be considered. As a consequence, open-source and free versions of applications were used. At the University of Belize, I developed the Service through Research Program (SRP), which allowed students to complete community service hours on the *Language, Culture, and History* project.[34] As a way of attracting students and ensuring that the program could benefit them beyond the obligatory completion of service hours, I modeled the SRP on service-learning principles, which hold that students' academic learning is deepened in structured community service activities that mutually benefit students and communities in meaningful ways.[35] Rather than "structured community service activities," students engaged in research activities that benefit both research projects and students. One key aspect of this program is that students did not need to have any previous research experience or training in sociolinguistics. By virtue of having grown up in Belize and being immersed in multiple aspects of Belizean culture, Belizean students

---

[32] I am grateful to James Law and Maurice Bélanger for their assistance.

[33] These were typewritten partial transcripts of interviews.

[34] Students who are on scholarship must complete a set number of community service hours as a condition of their scholarship. The scholarship program is administered through the office of the Dean of Student Affairs. I owe my thanks to then-dean of student affairs, Jean Perriot, as well to administrators at the time in the Faculty of Education and Arts, Dean Nestor Chan and department chair Ubaldimir Guerra, for supporting the initiative.

[35] Barbara Jacoby, *Service-Learning in Higher Education: Concepts and Practices* (San Francisco: Jossey-Bass, 1996); Janet Eyler and Dwight E. Giles Jr., *Where's the Learning in Service-Learning?* (San Francisco: Jossey-Bass, 1999).

would already be well-positioned to have the cultural and/or language knowledge needed to work with data from Belize, in contrast to technical aspects, which could be taught. Training offered to students included general introduction to research, ethics in minimal risk research, elements of data management, digitizing text documents, and time-aligned transcription. In addition, students were asked to meticulously track time at the start of new tasks in order to give feedback on workflows. In this way, they contributed to refining the workflow and to improving the program in general. Two students and I later summarized this work in a poster presentation discussing the development of the program.[36] Some of the community service hours were also spent attending conference presentations at the Belize National Research Conference and the Caribbean Studies Forum as another way for students to learn about and participate in research. Many of the same formalities of paid student research assistantships were also implemented, as it was always my intention to convert the SRP assistantships into paid positions as soon as funding became available. In 2018, this became possible thanks to a CLIR-Mellon research stipend, and Erson Moreira,[37] who had been working longest on the project, was hired as a research assistant to work remotely from Belize. Although I consider SRP to be a component of capacity building as new or expanded knowledge was shared with students local to Belize, the flow of knowledge was bidirectional rather than from expert to nonexpert. Students' preexisting knowledge of Belizean cultural references and their working knowledge or fluency in Spanish, particularly Belizean varieties of Spanish, were key contributions to the work. I intentionally privileged cultural and language knowledge[38] over linguistics training as a means of countering extractivism and in order to be accountable to the interviews. While enrichment of local communities in this context would generally refer primarily to the interviewee communities, defined more broadly, various sectors of the broader Belizean community (e.g., students described above), the Belizean diaspora, and the Central American diaspora can be thought of as local if seen in contrast to outsider privileged communities. As described further below, I intentionally sought out research assistants from Belize as well as other parts of Central America.

Once at UCLA for a CLIR fellowship, plans were made for a community collaborative process and capacity building in Belize to take place in March 2020, but, due to COVID, this had to be put on hold and the project scaled back.[39] I decided to work on a selection of interviews that corresponded to the narratives published by Timothy Hagerty and Mary Gomez Parham, since these were already in the public domain, albeit in translated, anonymized, and

---

[36] Fuller Medina, Erson Moreira, and Kennion Moreira. 2018. "Building Research Capacity Through Service Learning: Notes from the Language, Culture and History Project." Poster presented at the 1st National Research Conference, University of Belize, Belmopan, Belize, March 21-22.

[37] Erson Moreira is a student athlete at the University of Belize in the Medical Laboratory program. He digitized and organized the bulk of the typewritten transcripts. As a result of having worked on the ORBS since 2017 he has in-depth knowledge of the corpus and is currently a research assistant on the project.

[38] Some interviewees would use Yucatec Maya, English, or Kriol words or phrases in Spanish. Someone who shares the same variety or who also employs similar features is the best candidate to transcribe the interviews.

[39] Nicté Fuller Medina, "Repatriating Sociolinguistic Data to Belize: Can a Decolonial Model Still Work in Times of COVID 19?," CLIR COVID (Re) Collections (blog), 22 June 2020, https://www.clir.org/2020/06/repatriating-sociolinguistic-data-to-belize-can-a-decolonial-model-still-work-in-times-ofcovid-19/.

redacted form.[40] Since there was no metadata linking the published stories to the original audio, sorting this out became the priority in order to identify which audio could be put online. In keeping with capacity-building priorities, UCLA student Karen Contreras later was hired to create time logs and extract metadata.[41] This was followed by the recruitment of additional research assistants supported through UCLA Library in fall 2020. Alexander González Paz[42] and Daniela Hernández Castro[43] were hired to complete time logs, extract metadata, and create time-aligned transcriptions using ELAN.[44] A third research assistant from Belize, Griselly Padron, was hired at the same time.[45] All three are native speakers of Spanish (Guatemalan, Salvadoran, and Northern Belizean, respectively) but had varying skill levels with respect to written Spanish and cultural and language knowledge of Belizean varieties of Spanish. Erson Moreira had the most knowledge about the collection and the project. While the intention was to prioritize language and cultural knowledge in hiring research assistants, in the end all research assistants brought a combination of linguistics training, transcription experience, and other technical knowledge. As a team they complemented each other and we developed workflows that optimized their skills.[46] Our transnational team spanned Los Angeles, northern Belize, western Belize and, in 2021, Toronto. All work was completed virtually, using opensource applications with the exception of Zoom. As in the SRP, the research assistants received training in several areas, including time-aligned transcription, ethics, data management, and the politics of transcription. Three of the students and I synthesized our insights in a copresentation on the political nature of representing minoritized voices through transcription.[47]

------

[40] Timothy W. Hagerty and Mary Gomez Parham, eds., *If di Pin Neva Ben: Folktales and Legends of Belize* (Benque Viejo del Carmen, Belize: Cubola Productions, 2000).

[41] Karen Contreras is a Mestizx Belizean Angeleno whose family hails from western Belize. She is a self-identified heritage Spanish speaker majoring in ethnomusicology at UCLA with previous experience in sound engineering. She worked on the project from August to September 2020. Her research assistantship was supported through the CLIR fellowship research stipend.

[42] Alexander González Paz is a native speaker of Guatemalan Spanish and brought to the table formal training in applied linguistics and Hispanic linguistics as well as some language studies in K'ichee' Maya, which helped us work with some of the Maya lexical items unfamiliar to us.

[43] Daniela Hernández Castro is an Afro-Indigenous Salvadoran and a native speaker of Salvadoran Spanish. She is a premed student with a minor in Spanish at UCLA. She had ample experience in transcription, coursework in Spanish linguistics, and brought a strong language justice perspective to the project.

[44] Han Sloetjes and Peter Wittenburg, "Annotation by Category—ELAN and ISO DCR," in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC), May, 2008* (Marrakech: European Language Resources Association, 2008).

[45] Griselly Padron is a native speaker of Northern Belizean Spanish and holds a BA in anthropology from Galen University, Belize. In addition to her social science background and transcription work, she brings insider knowledge of Belizean varieties of Spanish and local cultural references.

[46] Two other students worked for a very short period on the project: Valeria Sawers (Stanford University) and Julia Tanenbaum (UCLA).

[47] Fuller Medina, Daniela Hernández Castro, Griselly Padron, and Alexander González Paz. 2021. "Towards an Ethics of Care in Representing Minoritized Voices Through Transcription." Presented at the UCLA Library Research Forum, UCLA, Los Angeles, CA, March 23.

In addition to the capacity building described above, another way I aim to disrupt the flow of enrichment is to create teaching guides for use in Belize. As noted above, local varieties of Spanish are stigmatized and not recognized in the education system. Teaching guides will be created to encourage teachers in Belize to use the narratives in classroom instruction and therefore reach younger people. This responds to the lack of teaching materials regarding Belizean Mestizo-Maya culture and the oldest varieties of Spanish spoken in Belize. Teaching guides will list recommended stories appropriate for language and culture classes along with suggested in-class activities, homework assignments, and discussion prompts. Accompanying worksheets for distribution to students will also be included. Since Belize is virtually invisible in Spanish textbooks and Hispanic studies texts, materials will also be more widely shared with language teachers in the United States and Canada.

Ownership and permissions

Ownership is intertwined with permissions since who owns a collection, such as the audio recordings comprising the ORBS, can determine both whose permission is needed and who can grant it. This is further complicated by different types of ownership (institutional, individual, or community), what ownership refers to (physical objects, content of narratives, individual tapes, or the collection of recordings), and legal ownership versus some other form not captured by current law.[48] The nature of the collection is also implicated. Research data is generally governed by stipulations set forth by institutional review boards (IRB) charged with ensuring that research with human participants meets the standard for acceptable ethical practice. But oral histories, for example, even if collected as part of research, are not necessarily categorized as data and have different protocols at the outset regarding permissions, redaction, anonymity, and so forth. Furthermore, funding sources may require that research data be publicly available or even open access. Likewise, soft funding supporting archival collections may have similar stipulations. In addition, copyright law can vary by country. In the case of the ORBS, a truly transnational collection, international, US, and Belizean copyright law are all relevant. According to US copyright, unpublished sound recordings created after 15 February 1972 would remain under copyright in the United States for the period of the life of the author plus seventy years.[49] In Belize copyright for a sound recording expires fifty years from the end of the calendar year in which it was created, with the provision that it was not made available to the public. If it was made publicly available, then the copyright expires fifty years from the end of the calendar year in which it was made available.[50] This means that according to

---

[48] The Western perspective on ownership often comes into conflict with the shared "ownership" of cultural knowledge generally not laid out in copyright or trademarks. As a result, cultural knowledge is often taken without permission, as in the recent attempt to name a rum in the United States "J'ouvert," where the name was to be trademarked based partly on erroneous claims that the word has no meaning in a foreign language. Both the cultural and linguistic patrimony (ownership) were overlooked, no doubt, in part, as a result of Western concepts of both ownership and of language. Jo-Anne Ferreira, "Baré Yo? Baré Yo!," *Language Blag* (blog), 21 June 2021, https://languageblag.wordpress.com/2021/06/21/a-transnational-jouvert-komes-bare-yo/.

[49] "Copyright at Cornell Libraries: Copyright Term and the Public Domain," accessed 1 November 2021, https://guides.library.cornell.edu/copyright/publicdomain.

[50] Government of Belize, "Copyright Act (Cap. 252, Revised Edition, 2000)," accessed 12 November 2021, https://wipolex.wipo.int/en/text/125463.

Belizean copyright law, the recordings would move into the public domain in Belize in less than ten years but remain under US copyright for more than seventy years, since the creator of the recordings is still alive.

Sociolinguistic data, in contrast, is traditionally considered to belong to the researcher who collected it, while library and archival collections belong to the institution where ownership is established in various ways, such as through purchase or donations where legal ownership is transferred to the institution. There are many good reasons for these types of ownership in terms of care, stewardship, access, and, in the case of sociolinguistics, research ethics. The care and management of a collection can be financially burdensome, labor intensive, and require consideration of ethical and legal issues on a case-by-case basis. It then falls on the researcher, research team, or institution to manage access. In some cases, older data is being made publicly available online via Creative Commons licenses. One such example is the Corpus of African American Language (CORAAL), where the creators have made various collections of interviews freely available online via Creative Commons license. This provides an important means of appropriate data citation, which is key to reproducible research. They have also decided to anonymize all interviews; if explicit permission has been granted by an interviewee, credit is attributed by naming them in the acknowledgments section of the user guide but not linked to their interviews.[51] Nonetheless, in linguistics research, there is generally little, if any, mention of returning data to communities. It simply is not a part of standard sociolinguistic practice. Moreover, I am not aware of substantive discussions in sociolinguistics that consider participants to be coauthors of recordings, either from an ethical or legal standpoint. One subfield where there is some departure from this is in language documentation, where it is becoming more and more common to make data available in various ways to communities in the interest of language revitalization and preservation.[52] This access can take the form of pedagogical materials, talking dictionaries, cocuration, and so forth. Libraries tend to do less of this type of community sharing though the practice of community archives, community engagement, and postcustodial collecting, and more open access collections are being created. In tandem with these orientations toward openness and community engagement are calls

---

[51] Tyler Kendall and Charlie Farrington, *CORAAL User Guide*, version 2018.10.06, Corpus of Regional African American Language (Eugene, OR: Online Resources for African American Language Project, 2018); Kendall and Farrington, *The Corpus of Regional African American Language*, version 2020.05 (Eugene, OR: Online Resources for African American Language Project, 2020).

[52] Two examples of large-scale repositories are the Pacific and Regional Archive for Digital Sources in Endangered Cultures and the Endangered Languages Archive. Nick Thieberger, "Curation of Oral Tradition from Legacy Recordings: An Australian Example," *Oral Tradition* 28, no. 2 (May 2013): 253–60; Nathan, "Access and Accessibility at ELAR." Both indicate some motivation for the safeguarding of audiovisual cultural and language data for the benefit of communities; however, in my view, they are essentially scholar- and archive-centered. In addition, the technological capacity required to access an online digital repository is not available in many communities. Florence Piron, Thomas Hervé Mboa Nkoudou, Marie Sophie Dibounje Madiba, Judicaël Alladatin, Hamissou Rhissa Achaffert, and Anderson Pierre, "Toward African and Haitian Universities in Service to Sustainable Local Development: The Contribution of Fair Open Science," in Leslie Chan, Angela Okune, Rebecca Hillyer, Denisse Albornoz, and Alejandro Posada, eds., *Contextualizing Openness: Situating Open Science* (Ottawa: University of Ottawa Press, 2019), 311–31.

for institutions to employ contributive justice-based frameworks, particularly with respect to transnational postcustodialism.[53]

For researcher-scholars like myself, these decisions are largely individual ones. For me as an insider-outsider researcher,[54] the extractivism inherent in data collection from minoritized communities was always problematic, as were the standard sociolinguistic practices previously described. Consequently, the moment I conceptualized the ORBS interviews as patrimony, I viewed ownership as shared, as dictated by the very definition of patrimony. I began to think of ways the creation of a sociolinguistic corpus could give equal or primary consideration to communities. While Hagerty had passed the physical collection on to me and told me it now belonged to me, based on my experience in libraries and memory institutions, I chose to appeal to copyright law, and in 2019 I requested that Hagerty transfer copyright to me. This would give me some legal standing to work with the data in the ways I wanted as well as potentially facilitate collaboration with memory institutions in Belize, as I would have adequate documentation for accessioning the collection. While I viewed this as offering some protection both for me and the data in a system that almost exclusively recognizes only Western forms of individual ownership, copyright transfer is not without its problems. First, the recordings were created in Belize but had been located in the United States for forty years. They will be accessed online through a US-based institution and potentially on site in Belize while I, the current copyright holder, am located in the United States. As noted above, any differences in copyright law between the two contexts need to be considered. Additionally, a number of ethical issues remain to be considered. Presuming legal copyright permits all decisions to be made by the copyright holder, such a stance would align with colonial practices rather than any of the models referenced earlier, such as the UNESCO guidelines on intangible cultural heritage.

In a community-centered model, obtaining permission from interviewees is the best-case scenario but is not necessarily easy to do or well supported. Sociolinguistic data is generally collected under ethical guidelines assuring interviewees of confidentiality and anonymity. Consequently, the practice has generally been to restrict access to sociolinguistic data. If the data is being made public, identifying information is redacted, pseudonyms may be used for speakers and/or the community where data was collected, or, more recently, researchers obtain explicit consent to put that data online. Thus, solid ethical precedent supports making redacted interviews available without explicit consent, provided that recordings are anonymized. This is bolstered by Western ownership, which gives me, as copyright holder, standing to make various decisions regarding the interviews. In addition, given that the project has been partially funded through UCLA Library, adherence to and not just alignment with its policy

---

[53] T.-Kay Sangwand, "Preservation Is Political: Enacting Contributive Justice and Decolonizing Transnational Archival Collaborations," *KULA: Knowledge Creation, Dissemination, and Preservation Studies* 2, no. 1 (2018): 1–14.

[54] A researcher's positioning is constantly being negotiated, relativized, and coconstructed in interaction with research participants and collaborators. As a Belizean, I am an insider in a pan-Belizean sense who does share language and culture with Belizean communities. Yet I cannot be fully an insider in all communities within Belize. This positions me at different times as an outsider, as does the researcher role. For a discussion of insider/outsider positioning in a creolophone context, see Shondel Nero, "Language, Identity, and Insider/Outsider Positionality in Caribbean Creole English Research," *Applied Linguistics Review* 6, no. 3 (2015): 341–68.

on open access has also been an important consideration.[55] Finally, a crucial aspect that militates against community permission-seeking and collaborative decision-making is the time and financial investment required. Tracking down interviewees is best done on the ground in Belize, which costs both time and money.[56] Together, these issues illustrate how default practices, even those that are most progressive and consider the protection of interviewees, may ultimately fall short, particularly where minoritized communities are involved and where there is a transnational dimension. This creates immense pressure to conform to prevailing practices, but the scholarship on community archives, postcustodialism, critical archival studies, and the protocols and guidelines noted above make explicit the need to hold the line. The process for achieving representation is as important as representation itself. I am periodically reminded, for example, of the kind of impact failure to obtain permissions could have. Amanda Delgado recounts her surprise at coming across a recording of a family member on social media, likely shared with the best of intentions but lacking consideration of the ethical dilemmas: "Cuando queramos compartir material a una audiencia más amplia no olvidemos realizar el consentimiento. Si la persona ya falleció se debe pedir a las familias involucradas. Esto les tomará mucho tiempo pero pues ni modo, y como recompensa las familias aprecian el gesto de respeto."[57] Moreover, family members of interviewees or interviewees themselves may wish to be credited for their narratives or memorialized in an online corpus. To be included in an archive could be a source of pride while simultaneously offering representational belonging.[58] Anonymization of interviews, generally a safeguarding measure to *protect* interviewees, could well compound existing erasures or present barriers to interviewees or their descendants who may wish to locate their interviews.[59] In sum, neither the ethical nor the legal issues related to ownership of recordings are clear-cut, highlighting the importance of navigating this work intentionally and in conversation with interviewee communities, local memory institutions, and cultural workers. Accordingly, we are making every effort to locate interviewees through my own social and professional network in order to obtain permissions, determine whether

---

[55] The UCLA Library Open Collections and Scholarship Policy states, in part, that their strategy "Creates and cultivates distinctive collections of unique materials, both those that are international in scope and those that document the histories and cultures of Southern California's diverse communities, and makes them discoverable and deliverable to scholars worldwide" https://www.library.ucla.edu/about/about-collections/open-scholarship-collections-policy

[56] Ultimately the pandemic made this portion impossible to carry out in any substantive way.

[57] "When we want to share with a wider audience let's not forget to get consent. If the person is deceased then the families involved should be asked for consent. This may take you a lot of time but that's just the way it is, and in return, the families appreciate the show of respect"; Amanda Delgado, "Ixamanda Twitter post," 27 October 2020, https://twitter.com/Amanda_DelgadoG/status/1321300865320439809; my translation.

[58] Michelle Caswell, Marika Cifor, and Mario H. Ramírez, "'To Suddenly Discover Yourself Existing': Uncovering the Impact of Community Archives," *American Archivist* 79, no. 1 (Summer 2016): 56–81.

[59] Thieberger argues that researchers have an obligation to ensure that speakers can locate their records or those created with their ancestors, underscoring the importance of a well-constructed corpus; Thieberger, "Curation of Oral Tradition from Legacy Recordings." I would add that easily searching for a name is also key. There is certainly precedent for crediting narrators for their stories. In Newfoundland, Canada, for example, creators of vernacular culture are traditionally credited for their contributions; Becky Childs, Gerard Van Herk, and Jennifer Thorburn, "Safe Harbour: Ethics and Accessibility in Sociolinguistic Corpus Building," *Corpus Linguistics and Linguistic Theory* 7–1 (2011): 163–80.

interviews should be anonymized, and work with local communities and institutions. Once it is safe to travel to and within Belize, more of this work will be done on the ground.

Voice

In addition to countering symbolic annihilation,[60] this work aims to make it possible for these underrepresented voices to participate and intervene in, as well as contribute to, the production of knowledge. In the first instance this is meant literally, by allowing the interviewees' voices themselves to speak as these recordings "provide an empirical basis of evidence on which to assert communities' historical presence; . . . to prove the facts of their presence in the face of silencing, marginalization, and misrepresentation."[61] To some extent Hagerty and Parham have facilitated this already through the publication of stories in *If di Pin Neva Bend,* where they were careful to stay faithful to the narratives as they were originally given.[62] Accordingly, the ORBS will form part of the nascent Belizean literature and serve as a testament to orality in literature. In addition, the very writing of this essay, outlining the choices made and the work itself, has given me voice as a Belizeanist insider/outsider researcher, potentially giving voice as well to other insider/outsider researchers by offering a different path to working with data and legacy collections.

A second way that the narratives give voice is in providing the basis for potential empowerment of local scholars to interpret local epistemologies and build locally relevant and useful knowledge.[63] Hagerty and Parham's choice to publish the narratives in Belize is once again key, as this made the work available locally and contributes to enriching the local literary tradition.[64] Belizean scholars and writers have had the opportunity to provide Belizeanist perspectives rather than outsider interpretations on the local folklore, spiritual beliefs, and spirit beings

---

60 Caswell, Cifor, and Ramirez, 'To Suddenly Discover Yourself Existing,'" reinterprets this term, originally used by feminist media scholars to describe the ways mainstream media render invisible, malign, or otherwise marginalize minoritized groups, to refer to a parallel process in memory institutions.

61 Caswell, Cifor, and Ramirez, 74.

62 Timothy W. Hagerty, personal communication, 11 December 2021. Mary Gomez Parham, coeditor of *If di Pin Neva Bend*, is a Belizean scholar.

63 Thomas Hervé Mboa Nkoudou, "Libraries in the Age of Technocoloniality: Epistemic Alienation in African Scholarly Communications," paper presented at the Critical Approaches to Libraries Conference*, online, 5–6 May 2021*, https://doi.org/10.5281/ZENODO.4739658.

64 I am unaware of the factors which led to the choice to publish the narratives in English translation, but while a Spanish-language publication may have been more closely aligned with the original narratives, an English-language publication is arguably more widely accessible in Belize since English is the official language. This conundrum deeply reflects the particularities of the Belizean linguistic ecology, a dilemma that, in the specific case of this book, was not salient to me as someone embedded in and shaped by that ecology despite my training in sociolinguistics and my language-justice commitments. I am grateful to research assistants for raising this issue of the publication language choice. The choice to publish in Belize, however, was clearly intentional despite Hagerty and Parham's being academics based in the United States. Hagerty also published articles in Belizean journals; e.g., Hagerty, "Historical Perspective on the Spanish Language of Belize," *Belizean Studies* 20 (1992): 16–21. When asked about this choice, he indicated that in his view, if the work was about Belize, then it should be accessible to Belize (personal communication, 11 December 2020).

described in the collection of stories.[65] Likewise, these narratives will provide the basis for creating teaching materials to support recognition of Belizean varieties of Spanish and to fill gaps in various subfields of linguistics, where contributions from Belizean varieties of Spanish have been largely absent.

Finally, this project gives voice by returning this language data to the public domain in digital form as Caribbean language data tends to be less widely available.[66] This presents a challenge to computational methods for linguistic analysis and applied contexts such as automated speech recognition systems and speech pathology. In addition, quantitative sociolinguistics has largely developed on a relatively narrow and nonreproducible base undergirded by a monolingual bias, yet knowledge produced is assumed to be universal, particularly where quantitative and/or computational methods are used. One way to address this is to make more data available to be tested by existing models.[67] I see this corpus as supporting open science and reproducibility since competing theories and or methods can be tested on the same dataset and replication is more feasible.[^69]

In order for this work to give voice, and keeping in mind the priority to center communities while being guided by an ethics of care, it was important to consider where the narratives would "live." The Digital Library of the Caribbean (dLOC) was the obvious choice for an online access point. dLOC is described on its homepage as "a cooperative digital library for resources from and about the Caribbean and circum-Caribbean."[68] Content is contributed by over sixty partners in the United States, the Caribbean, Canada, Central and South America, and Europe. Several elements make dLOC ideal: partners and contributors retain all rights to materials and dLOC supports collaborative projects and funding initiatives as well as capacity building. Consequently, dLOC is a known entity in Belize. The Belize National Library Service and Information System has been a partner since 2008 and Belizean cultural heritage workers have benefited from support from dLOC in various ways.[69] That dLOC is known in the region and in Belize is an important consideration in terms of discoverability by those users for whom the collection of interviews may be most relevant. Furthermore, the dLOC model, which centers the Caribbean, the partners, and the collections themselves, aligns with the vision of the *Language,*

---

[65] Polanco, "Of Gods, Men, and Earth"; Christopher Lloyd De Shield and Gerardo Polanco, "Succouring an Ixtabai: Zee Edgell's Deployment of Belizean Folklore in *The Festival of San Joaquin* (1997)," *Revista Canadiense de Estudios Hispánicos (2019)*: 23–42.

[66] John Russell Rickford, "The Value of Online Corpora for the Analysis of Variation and Change in the Caribbean," in Shelome Gooden and Bettina Migge, eds., *Social and Structural Aspects of Language Contact and Change* (Berlin: Language Science, 2021).

[67] Aria Adli and Gregory Guy, "Widening Horizons: Cross-cultural Approaches to Linguistic Variation," workshop presented at the *conference "New Ways of Analyzing Variation 45,"* Vancouver, November 2016; Andrea L. Berez-Kroeker, Lauren Gawne, Susan Smythe Kung, Barbara F. Kelly, Tyler Heston, Gary Holton, Peter Pulsifer, et al., "Reproducible Research in Linguistics: A Position Statement on Data Citation and Attribution in Our Field," *Linguistics* 56, no. 1 (2018): 1–18; Miriam Meyerhoff and Naomi Nagy, eds., *Social Lives in Language Sociolinguistics and Multilingual Speech Communities: Celebrating the Work of Gillian Sankoff*, vol. 24 (Philadelphia: John Benjamins, 2008).

[68] Digital Library of the Caribbean (dLOC), home page, accessed 25 May 2021, https://www.dloc.com.

[69] For further information on dLOC's work in outreach, training, exhibits, and projects, see the dLOC Factsheet, https://dloc.com/AA00001499/00004.

Culture, and History project. It permits a virtual and symbolic repatriation to the Caribbean, keeping in mind of course Belize's peculiar status between the Caribbean and Central America. Through this online platform, interviews can be safely archived while accessible to diaspora Belizeans (or anyone) with the technological capacity. Since this technological capacity is not guaranteed in Belize, on-site local access will have to be determined through a community collaborative process.

The (Uncertain) Way Forward

Several issues remain to be disentangled, many of which must wait for when it is safe to travel to and within Belize for in-person community and individual meetings. These relate to metadata, on-site access, repatriation of physical objects, levels of access, and a community consultation model. Metadata is currently in English as is the description of transcription decisions and markup. These will have to be provided in Spanish in keeping with the goal of ensuring access and making visible Belizean varieties of Spanish. Furthermore, metadata may be expanded and edited (e.g., redacting names) depending on community input. Although the process for obtaining community contributions on description of the collection, anonymization, and on-site access is on hold due to COVID, in the interim, the interviews can be archived with dLOC without being made public. With respect to the physical objects, the intention is to return them to Belize to be archived at an appropriate memory institution as determined in collaboration with interviewee communities and stakeholders from local institutions. However, this is not without its challenges. Some of the tapes are already deteriorating and legacy equipment such as a reel-to-reel player may not be available to listen to original recordings. In addition, current institutional capacity to accession the collection may have been impacted by economic fallout from COVID.[70] This also relates to on-site access. Community members may wish to have copies of the interviews located in their community, but there may not be a community organization with the infrastructure to steward a collection. This could present a challenge, since locating the recordings as close to their record creators as possible is not simply a matter of justice but also an important aspect of longevity, discoverability, and use. If users have to travel to a central location to access the collection, for example, then it is unlikely that it will be regularly accessed. The potential for the collection to give voice will be impeded if it is functionally lost to communities and scholars. One way to increase visibility and use, regardless of access points, is through the creation of resources for teachers in formats that are accessible and useable in local contexts. This will be a priority in the next phase of the project.

With respect to digital access, I have envisioned an online open access repository with the caveat that not all interviews should be available or available in their entirety and that interviewees may not want their interviews (redacted or unredacted) to be part of an online open access resource. Community members may also wish to provide instruction on how they wish the narratives to be used if they will be in the public domain, whether that be online or local. At the same time, the copyright and ownership issues discussed above will have to be satisfactorily resolved. While existing publicly available sociolinguistic corpora

---

[70] Fuller Medina, "Repatriating Sociolinguistic Data to Belize."

such as the CORAAL already depart from traditional sociolinguistic data practices by making data open access via a Creative Commons license, this achieves the goals of a more open science but not necessarily those of communities and may not be relevant for a transnational collection. In fact, Creative Commons did attempt licenses for the Global South but the endeavor was not successful.[71] The work of Local Contexts may provide a key starting point together with current practices in Belize in the safeguarding of intangible cultural heritage.[72] Local Contexts reenvisions intellectual property frameworks for Indigenous materials to provide both legal licenses and nonlegal labels, Traditional Knowledge Licenses and Labels, which could help facilitate management of cultural knowledge in the sense of extending community best practices to those outside the community. Such a model both provides an alternative to Western copyright while still providing a license as well as pathways for agency that other communities could follow.

Discussion

This project challenges the traditional approaches of memory institutions and the field of (socio)linguistics by centering communities. Archives are generally institutional repositories, which are often inaccessible to the very communities they document. Likewise, sociolinguistic interviews are rarely returned or accessible to communities. The decolonial approach of this project insists on the recognition that a "collection or documentation of knowledge is made possible by a range of acts of generosity and sharing that should not culminate with the singular authorship/ownership of that material."[73] Likewise, I insist on alternative paths to working with legacy sociolinguistic collections by recognizing language data as cultural patrimony. This is accomplished through repatriation, envisioning a shared decision-making process, cocreating space for community voice and agency in the preservation of its stories, and confronting many of the legal, technical, and ethical issues in the creation of an open repository of narratives from the Global South.

Yet, in the words of Audre Lorde: the master's tools will not dismantle the master's house. This collection sits, as do I, in the matrix of coloniality and all that it implies. Despite my stance, for example, that narratives, language, or stories cannot be owned by a researcher, copyright law indicates that this ownership is possible. I myself appealed to this Western ownership by requesting transfer of copyright in my name as an attempt to function within the Western context and hold the recordings safe. In addition, as Hannah Alpert-Abrams and colleagues point out, neoliberalism is implicated even in practices, such as postcustodialism, deemed to be anticolonial.[74] I must question if enough steps have been taken in this project to guard against the commodification of the ORBS and any collections to follow. And what of neoliberal labor conditions that may share similarities with those described by Alpert-Abrams

---

[71] Jane Anderson and Kimberly Christen, "'Chuck a Copyright on It': Dilemmas of Digital Return and the Possibilities for Traditional Knowledge Licenses and Labels," *Museum Anthropology Review* 7, no. 1–2 (2013): 105–26, esp. 110.

[72] Anderson and Christen.

[73] Anderson and Christen, 120.

[74] Alpert-Abrams et al., "Post-custodialism for the Collective Good."

and colleagues?[75] Much work was unpaid and completed by students in Belize even if within a formalized training and mentorship program that intentionally aimed to be equitable. I interpret this circumstance to be an outcome of the historical impoverishment that leaves the national university with a nascent and unstable research infrastructure with little to no internal funding. This is also reflected at the national level, where no research funding bodies exist. The work later continued in a well-resourced context but in circumstances of precarious temporary employment, along with multiple barriers to transnational work. The work itself suffered several delays as a result of my relocations to new institutions: first to the University of Belize then to UCLA and as of 1 August 2021 to a third institution.[76] While the slowing down of time is in itself an anticolonial stance, delays due to external circumstances are not a choice.[77] Nonetheless, there is some choice in the freedom afforded by the margins. The University of Belize is primarily a teaching institution, and while there is an expectation of research, this level of expectation had historically varied, so I was not subject to the publish-or-perish pressures so common in tenure-track positions at research-intensive institutions. Similarly, at UCLA, my work in the postdoctoral fellowship was centered largely on this project and was, in fact, one of the reasons I chose the fellowship in data curation. The fellowship allowed me to deprioritize the linguistic research goals and center the community collaborative aspect, even when this got side-tracked due to COVID. The academic clock is different from the community clock, so there was still risk. The former exerts pressure to prioritize analysis and publication (of both scientific publication and the corpus itself) over community collaboration. But in the end, centering communities, capacity building, and a decolonial praxis "disrupts Eurocentric notions of time that colonize our academic lives."[78]

Returning to the concept of agency, the curator or researcher and communities are not the only agents in bringing forth this collection. Jessica Tai and colleagues note a somewhat "radical and contested concept in the humanities": the "notion that material objects hold agency, . . . the capacity to exert power independently of their human creators or users."[79] This would perhaps be even more radical and highly contested in linguistics, a science or social science that privileges the scientist, itself a Eurocentric view. However, the nature of this collection, consisting of actual voices, some of which are ancestor voices, makes this notion appear less radical. There are several ways I view the ORBS as sentient and dynamic. First, it exists in and interacts with a multidimensional space connecting the past to the present and the diaspora to Belize. It is both constrained by and, in opposition to coloniality, hidden for forty years, speaking only in translation in English or through linguistic analysis soon after interviews were completed. Second, the collection has taken on a life of its own, guiding its

---

[75] Alpert-Abrams et al., "Post-custodialism for the Collective Good."

[76] Having learned from previous moves, I have put several things in place to transition with minimal delays. This includes using available research funds to support current research assistants whose experience on the corpus now makes them particularly invaluable.

[77] Riyad A. Shahjahan, "Being 'Lazy' and Slowing Down: Toward Decolonizing Time, Our Body, and Pedagogy," *Educational Philosophy and Theory* 47, no. 5 (2015): 488–501.

[78] Shahjahan, 2.

[79] Jessica Tai, Jimmy Zavala, Joyce Gabiola, Gracen Brilmyer, and Michelle Caswell, "Summoning the Ghosts: Records as Agents in Community Archives," *Journal of Contemporary Archival Studies* 6, no. 1 (2019): 1–20, quote on 5.

own trajectory and social life. I can only take partial credit for how this trajectory unfolded in the time since the interviews were first transferred to me in 2016. Here I depart from traditional research perspectives, aligning more with what Glenda A. Leung speaks of as a possible fifth research paradigm, one that welcomes the "interplay between the seen and unseen, the material and nonmaterial, the rational and emotive, the sensing mind and the intuitive mind." Pushing further, Leung asks, "What if the status quo shifted and it were acceptable for intuition and synchronicity to be part of research design? What if the unseen informed the unfolding of one's research?"[80] My work with the ORBS unfolded in a way that I can only begin to appreciate and embrace in hindsight. I am left to wonder, for example, if my career trajectory had been different, would the interviews have entered the public domain too soon? Each career decision I made was with the ORBS in mind. One reason I accepted a position at Belize's national university was because I felt this would facilitate the repatriation and capacity-building efforts. When I found that I needed more access to technology and archival expertise, the fellowship in data curation presented itself and I moved to UCLA. Much of the work described here, and the writing of this essay, takes place at UCLA, the very university where Hagerty was a PhD student in the 1970s when he collected the data in Belize in order to complete his doctoral dissertation. When I accepted the fellowship at UCLA, I had not made this connection but perhaps the unseen, the ghosts in the archive, the ancestors had. Finally, throughout the fellowship, I gained much more insight into the value of the Digital Library of the Caribbean and how it operates. Might the synchronicities and nudges reflect the collection exerting agency and sentience?

Adding to the questions Leung poses, I would also ask, what if this fifth research paradigm included the ancestors' voices, their interventions, and it became acceptable to acknowledge them? Tai and colleagues posit that the archivist is accountable to ghosts, those who are missing and who by their absence are ghosts in the archive, and, further, that we are obligated to portray them in an ethical way, in a way that honors and acknowledges them.[81] What if in abiding by these ethics we are honoring ancestors and not (only) the ghosts? I am not the first to call on, name, or acknowledge ancestors in academic work, but Western epistemologies fail us in allowing for the adequate mechanisms to cite them.[82]

While much work remains before the Older Recordings of Belizean varieties of Spanish enter the public domain via the Digital Library of the Caribbean, several imperatives have emerged. First, the priorities of sociolinguists to understand linguistic variability in its social contexts can no longer be the central concern when it comes to data from the Global South and other marginalized communities; rather, this must be balanced with community expectations regarding the curation of their own stories and oral histories. It may also be time to reflect on ancestors as part of contemporary communities and as part of archives. Second, "Data is Patrimony." I initially titled this essay "Data *as* Patrimony," but upon reading the current title,

---

[80] Glenda A. Leung, epilogue to Leung and Loschky, *When Creole and Spanish Collide*, 322. Leung lists quantitative, qualitative, mixed methods, and big data as the current four research paradigms.

[81] Tai et al., "Summoning the Ghosts," 16.

[82] At the 2018 Caribbean Studies Forum hosted by the University of Belize, a deep discussion emerged in the final question period regarding interventions from ancestors and the challenges of citing those interventions in an academic essay. We found no solutions.

I realized that this stronger assertion is more accurate in this case. Third, it matters whose hands work on the collection; again, the right people seemed to come together in alignment in a Belizean, diasporic Central American transnational team. Finally, the voices in the ORBS are ready to speak, the data should be released from the confines of sociolinguistic research to their descendants, to speak to multiple disciplines, to generate local knowledge, to expand our understanding of language and culture, and to be reused but always in ways dictated by an ethics of care.

Acknowledgements

Nicté Fuller Medina

Nicté Fuller Medina (PhD, University of Ottawa) is a quantitative linguist. She studies the mechanisms underlying multilingual speech to both inform linguistic science and to challenge narratives of deficiency with respect to racialized language varieties. She leads two on-going projects: *The Language Contact in Belize* project which examines Spanish-Kriol-English contact and *Language, Culture and History: Belize in a Digital Age* which focuses on the creation of digital corpora and repatriation of legacy sociolinguistic data within a decolonial framework. Her most recent publications appear in the volume *When Creole and Spanish Collide: Language and Cultural Contact in the Caribbean* and in the *Canadian Journal of Linguistics*. She is an insider/outsider researcher, a Belizeanist, Central Americanist and Caribbeanist.