Innovating corpus building: Community collaboration, repatriation and restitution

Nicté Fuller Medina
Swarthmore College

Recent scholarship in Caribbean linguistics has called for the creation of online Caribbean corpora (Rickford 2022) and for new approaches to linguistics from the perspective of Southern epistemologies (Pennycook and Makoni 2022). Digital corpora, in particular archival recordings, are an invaluable resource for real-time studies of change over time. Gooden and Drayton (2017), for example, compared contemporary data to archival digitized data from Trinidad and Jamaica and found that the nature of the shifting contact over time between different ethnolinguistic communities was detectable in intricate patterns of linguistic variation suggesting that speakers employ language in creating new sociopolitically driven identities within changing ecologies. The current paper addresses a gap in the availability of Caribbean digital data, in particular from Belize which is often overlooked due to its status as both and, simultaneously, neither Caribbean nor Central American. I report on a collaborative process to digitally preserve legacy recordings of Belize Kriol (English-lexified Creole). These recordings are the second installment of the umbrella project *Language, Culture and History: Belize in a Digital Age,* which aims to digitally preserve and provide broad access to legacy sound recordings within a decolonial framework. This framework draws on principles of post-custodial archival theory (Bastian 2002), the UNESCO principles of the safeguarding of intangible cultural heritage, epistemologies of the South (Pennycook and Makoni 2022) and best practices in sociolinguistic corpus building. From this perspective, as we confront the legal, ethical and technical aspects of corpus creation of a minoritized language varieties, I argue that the goals of linguistics research, in particular, on Caribbean languages, must be balanced with those of repatriation and the creation of a public resource that it is both meaningful and accessible to local communities. I report on the community engaged process which entailed formal collaboration with the Belize Archives and Records Services, as well as engagement with local community organizations involved in language activism, cultural documentation, and digital preservation of sound recordings through a series of workshops and consultations. Additionally, I discuss how I leverage the resources of broader regional organizations such as Digital Libraries of the Caribbean to host digitized recordings online and discuss the value of these legacy recordings for linguistics research. In addition to providing a real-time benchmark for assessing language change, for example, the recordings were created during the pivotal pre-independence period of the 1970's when Kriol was promoted as a unifying tool in the pro-independence movement which potentially impacted the spread of Kriol. Finally, I offer guidance on decolonial approaches for preserving legacy collections based on lessons learned.

Bastian, J. (2002)."Taking Custody, Giving Access: A Postcustodial Role for a New Century," *Archivaria*: *The Journal of the Association of Canadian Archivists* 53 (May), 76-93. https://archivaria.ca/index.php/archivaria/article/view/12838.

Gooden, S., & Drayton, K. A. (2017). The Caribbean. In Raymond Hickey (Ed.) *Listening to the Past: Audio Records of Accents of Englis*h, 414-43. Cambridge University Press.

Pennycook, A., & Makoni, S. (2020). *Innovations and Challenges in Applied Linguistics from the Global South*. Routledge.

Rickford, John R. (2022). 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,* (pp. 213-232). Berlin: Language Science Press.