The Archive of the Indigenous Languages of Latin America: An Overview

Susan Smythe Kung and Joel Sherzer

The Archive of the Indigenous Languages of Latin America (AILLA) is a repository of primarily linguistic and anthropological data about the indigenous languages of Latin America and the Caribbean. In this article we give a brief description of the archive and its mission in Section 1, and we discuss the predecessors and precursors to AILLA in Section 2, and the importance of AILLA in Section 3. In Section 4 we highlight a few of the large and publicly accessible collections, and in Section 5 we illustrate some of the ways in which teachers, professors, researchers, and indigenous community members have used data archived at AILLA.

1. Description and Mission of the Archive

AILLA was founded at the University of Texas at Austin (UT-Austin) in 2001 with funding from the National Endowment for the Humanities and the National Science Foundation.¹ Today AILLA is directed by Joel Sherzer, Emeritus Professor of Anthropology, along with codirectors Anthony C. Woodbury and Patience Epps, both Professors of Linguistics; and it is managed by Susan Smythe Kung, who holds a Ph.D. in Linguistics.

AILLA has no physical presentation space because it is a completely digital repository whose collections are accessible only through its website at http://www.ailla.utexas.org. As of this writing, the collection includes samples of 282 languages from 22 Latin American and Caribbean Countries. There are 16,370 audio recordings, 2,155 video recordings, 4,604 digital texts, and 4,289 images. 126 depositors from North, Central, and South America, as well as Europe have collaborated by archiving their data, the majority of which are raw, unanalyzed audio and/or video recordings. Often the raw recordings are accompanied by images, transcriptions, translations, and interlinearized morphological analyses.

¹ AILLA was launched as a pilot project with a seed grant from the College of Liberal Arts at UT-Austin in 1999. It is now chiefly supported by the LLILAS Benson Latin American Institute and Collections, a collaboration between UT-Austin's College of Liberal Arts and the University of Texas Libraries (UTL), with crucial additional support from the National Science Foundation. We are grateful to all these institutions for financial, moral, and technical support over the years. AILLA's founding directors are Joel Sherzer and Anthony C. Woodbury, mentioned below, and Mark McFarland, the former Director of the Digital Library Services Division of UTL. The founding manager of the archive is Heidi Johnson (Ph.D. in Linguistics), who retired in 2012.

The collection consists of multimedia files preserved on servers managed and backed up by the University of Texas Libraries (UTL) Digital Services. The AILLA website has parallel interfaces in English and Spanish. Visitors are free to browse the catalog and read the general information pages without registering with the archive. However, if they wish to access any of the files, they must register and create a user account. This process is free, but it requires that users agree to the Conditions for Use of Archived Resources (http://www.ailla.utexas.org/site/ use_conditions.html), which include, among other things, a prohibition against commercial use of the files, and an expectation that the visitor will demonstrate respect for the cultures and peoples whose languages, cultures, and work are represented in the archive. Once users have agreed to these terms, they are able to access any media file that has been archived at the public access level (level 1). Three other levels (levels 2-4) involve restricted access; these restrictedaccess levels allow creators or depositors of archived materials, or the communities in which these materials were collected, to control access to these data by means of passwords and time limits (that is, dates when the restricted materials will convert to public access).²

AILLA's primary mission is the preservation of irreplaceable linguistic and cultural resources in and about the indigenous languages of Latin America, most of which are endangered. Most archived resources are deposited by linguists and anthropologists for whom audio and video recordings are a central part of their research methodology. Many indigenous groups, such as the Maya linguistic research organization Oxlaiuui Keej Maya' Ajtz'iib' (http://www.ailla.utexas.org/search/collection.html?c_id=67), have also archived the results of their investigations with AILLA.

The majority of the materials in the repository are audio recordings that were originally created on media ranging from open-reel tapes to digital recorders. Analog materials are digitized either in AILLA's lab or in the UTL Digitization Services labs. The audio and video recordings consist of a wide range of discourse genres, including conversations, many types of narratives, songs, political oratory, traditional myths, curing ceremonies, and so on. Some recordings are accompanied by transcriptions and/or translations in media ranging from scans of handwritten notebooks to time-aligned XML files. Other textual resources include dictionaries, grammars, ethnographic sketches, field notes, articles, handouts, and presentations. The collection also contains many hundreds of photographs. If it can be digitized and is specifically relevant to an indigenous language and culture of Latin America, it is fully acceptable for AILLA.

AILLA's secondary mission is to make these valuable and useful resources maximally accessible via the Internet. However, we simultaneously endeavor to protect from inappropriate use any materials that are personally, culturally, or politically sensitive and to support the intellectual property rights of the creators. The system of access levels discussed above lets creators and depositors have fine-grained control over their materials, allowing them to specify different levels of access for each file in their collections or a single level of access for an entire collection. For example, audio recordings might be public, while video recordings might be restricted. Names of speakers or research participants can be kept anonymous or made public.

² A full description of the system of access levels is outside the scope of this article, but more information about the different levels of access is available on the Access Restrictions page of our website (<u>http://www.ailla.utexas.org/site/access_restrict.html</u>).

Historically, very little of the fruit of linguistic and anthropological research has been genuinely available to the public, to other researchers, or to the indigenous communities in which the research was done. The reasons for this lack of accessibility are threefold. First, very little of the linguistic and anthropological research on indigenous languages that was conducted during the twentieth century was archived. A linguist would go to the field for several months or years, record various narrative genres, take handwritten notes during elicitation sessions and interviews, and take photographs or make sketches. Upon returning home from the field, one would analyze as much of the data and publish as many analyses as possible, but there was not enough time to be able to transcribe every recording or analyze every speech event, and there was no precedent for sharing the raw data that had been collected. In very rare cases, for example when the linguist or ethnographer was extremely prolific or famous or affiliated with a particular institution, the collection might end up in a brick and mortar archive such as the American Philosophical Society or the Indiana University Archive of Traditional Music, among others, where the data would be archived and preserved for future researchers, students, community members, or interested persons. However, the large majority of the raw data on indigenous languages that was collected in the twentieth century has simply disappeared; it has deteriorated to the point of uselessness, or it has been thrown out by surviving family members. AILLA tries to rectify this problem by providing free digitization services for analog data, uploading both digitized analog data and born-digital data to secure servers, providing a user interface (website) to allow the public access to these data via the Internet, and returning the original analog data to the depositors.

The second reason that very little research data on indigenous languages has been available to the public has to do with the accessibility of data that are stored in brick and mortar archives. Even when raw, primary data are physically stored at a traditional archive, these data are still not easily accessed by either researchers or members of the indigenous communities in which the data were collected. In order to have access to these materials, community members or researchers must travel to the location of the archive, where they will have to pay for food and lodging for several days or even weeks in order to browse and/or study the materials. This physical barrier is non-existent for AILLA because all of the data are available online to anyone with Internet access. Thus, interested persons can browse the archive and download data from anywhere in the world.

Finally, even when interested persons travel to a traditional brick and mortar archive, they still might not be able to view or handle the data because the materials are restricted. They find themselves in a frustrating situation in which they are denied access to data even after they have traveled a long distance and spent a lot of money to study it. Though sensitive materials are protected, AILLA's directors, manager, and advisors strongly believe that accessibility is extremely important. Restrictions tend to keep the language community members out, while researchers might be able to gain access to archival materials through academic networks. Resources that are publicly accessible can be heard, seen, and read by all speakers, even those living outside of their native communities. AILLA's policy is that if a resource can be made public, it should be made public; if it is sensitive, it should be protected. Our goal is to ensure that the unique and wonderful resources preserved at AILLA can be used to maintain, revitalize, and enrich the communities from which they came.

AILLA was intended from the outset to function as a partner with its depositors, providing them with a means of both preserving and sharing, under appropriate terms, the data they collected during their fieldwork and research with the indigenous peoples of Latin America. The archive accepts any legitimate resources that can be housed in a digital format.

2. Predecessors and Precursors of AILLA

AILLA has roots in the Americanist tradition of the documentation of indigenous languages of the Americas, beginning with Franz Boas and Edward Sapir and continuing with their students. According to this tradition, grammatical features of a language and the culture of its speakers are observed and documented by means of the collection of texts. Of course, Boas and Sapir did not have tape recorders, so they collected texts through elicitation. The invention of the portable tape recorder revolutionized the field of linguistic anthropology. Tape recording—first on reel-to-reel tapes and later on cassette tapes—made it possible to collect, as well as accurately transcribe and translate, actual performances of verbal discourse. These texts, both the manuscripts and/or the audio recordings, were sometimes archived in such places as the National Anthropological Archives of the Smithsonian Institution, the American Philosophical Society, the Library of Congress, or the Indiana University Archive of Traditional Music, among others. However, most researchers never archived their text collections.

During the 1970s and 1980s, the University of Texas at Austin became the center of what has come to be called "the discourse centered approach" to language and culture. Two important conferences at UT-Austin developed this approach and led to the publication of two books, *Native South American Discourse* (1986), edited by Joel Sherzer and Greg Urban, and *Native American Discourse: Poetics and Rhetoric* (1987), edited by Joel Sherzer and Anthony Woodbury. During these conferences, tape recordings were played and analyzed, and these audio recordings were included in the resulting books. In the years that followed, recordings became essential components of both field research and the publication of its results.

Though there was already a tradition of archiving recordings along with field notes, especially in the field of ethnomusicology, the emergence of the Internet revolutionized the possibilities of archiving verbal discourse. Recordings could be digitized and archived on servers that were available to the public at large. Access, especially open access, became possible. All of this led to the founding of AILLA.

3. Importance of AILLA

Hundreds of native languages are still spoken in Latin America, but they are all endangered because of massive migration of the speakers away from their original homelands and drastic changes in sociocultural and economic conditions. Even in communities where the languages are still vibrant, culturally important ways of speaking—such as ceremonial dialogues, traditional narratives and songs, and curing practices—are being lost. Saving recordings of these speech genres enables future generations to remember and perhaps relearn their traditions. Linguists and anthropologists have been making recordings of indigenous languages for decades. Collections of magnetic tapes recorded in the mid- to late-twentieth century still survive, but even the original researchers find these materials difficult to access today. One of AILLA's most valuable contributions has been to digitize these fragile collections and to make them available to the public by storing them on a secure and stable server. Many well-known senior researchers have deposited recordings with AILLA, and new generations of researchers do so as well because they recognize that this is the best way to preserve the materials that they have so carefully collected and protected. Because the current generation of researchers is collecting digital audio and video, the collections are growing exponentially and will continue to do so.

It is especially urgent to support the survival of these endangered, indigenous languages and to help their speakers maintain them. AILLA enables the traditions of the past to be preserved for the present and the future. Current and future generations will have access to the cultural, moral, and aesthetic components that are part of their indigenous heritage. AILLA contributes to this goal in a major way, and is recognized for this contribution by the speakers and communities involved. In several cases, treasured curing chants and other valuable verbal performances thought to be extinct have been recovered, thanks to the recordings provided by the depositors.

Once recordings are safely housed in standard digital formats, they are available to anyone who has access to the Internet.³ AILLA is especially dedicated to making the collection available to members of indigenous communities in Latin America. It strives to keep the website sleek and swift so it will work properly in small-town internet cafés, as well as in big-city universities, and it uses only formats that can be listened to or viewed using common software programs that are easily downloaded free of charge. Many recordings are accompanied by transcriptions and/or translations in English, Spanish, or Portuguese.

AILLA plays a crucial role for indigenous communities. The recordings give credibility to their languages and cultures and confer prestige to them. Rather than being considered inferior dialects, these languages stand on a par with the major languages of the world. In addition, each indigenous community becomes aware of being part of a larger set of communities that has its own unique and special language. Each indigenous group is proud to have its voices represented along with other indigenous groups. For these reasons community leaders and organizations support AILLA.

4. Featured AILLA Collections

AILLA currently consists of 10,736 resources spread across 140 collections, some of which are quite large.⁴ Obviously it is impossible to highlight all 140 collections here, though they all are worthy of special attention. As our goal for this article is to encourage the reader to

³ Excluded from this availability are files that have been restricted from public access.

⁴ In order to access any of the files contained in the collections mentioned in this section (including files for which links have been provided), the reader must be a registered AILLA user and must agree to AILLA's Terms and Conditions of Use (<u>http://www.ailla.utexas.org/site/use_conditions.html</u>).

visit and explore the archive, here we mention some collections that are completely or mostly unrestricted. Nevertheless, even when individual files are restricted, the AILLA visitor can still read the descriptions of these files, as well as descriptions of the collections in which they are found, in the metadata notes. Some AILLA collections that are completely unrestricted include Jonathan Amith's Nahuatl Collection (http://www.ailla.utexas.org/search/collection.html? c id=1), Amith and Rey Castillo García's Mixteco Language Documentation Project (http:// www.ailla.utexas.org/search/collection.html?c_id=101), Alan and Pamela Sandstrom's Nahuatl Collection (http://www.ailla.utexas.org/search/collection.html?c_id=117), Yolanda Lastra's Mexican Languages Collection (http://www.ailla.utexas.org/search/collection.html?c_id=39), Kathryn Josserand's MesoAmerican Languages Collection (http://www.ailla.utexas.org/search/ collection.html?c id=79), Nicholas Hopkins' Maya Languages Collection (http:// www.ailla.utexas.org/search/collection.html?c id=86), Maurizio Gnerre's Jivaron Languages Collection (http://www.ailla.utexas.org/search/collection.html?c_id=18), H. Dieter Heinen and Werner Wilbert's Warao Language and Culture Collection (http://www.ailla.utexas.org/search/ collection.html?c_id=29), and Marie Claude Mattei Müller's Venezuelan Languages Collection (http://www.ailla.utexas.org/search/collection.html?c_id=45).

Many of the collections that are archived with AILLA and are thus available on the website contain the raw, primary data that are the foundations upon which many journal articles, dissertations, and books have been based. Here we highlight two such collections that are relevant to specific books.

Jonathan Hill's Curripaco⁵ Collection (http://www.ailla.utexas.org/search/ collection.html?c_id=68) includes his recordings of all of the stories and music, as well as the photographs and some of the transcriptions of stories, that he included in his 2009 book Made from Bone: Trickster Myths, Music, and History from the Amazon. Appendix B in the book provides a chapter-by-chapter list of all of the archived resources and their corresponding AILLA resource numbers. Though the book itself is not archived at AILLA, much of Hill's raw data on which the book is based is. Figure 1 shows a screenshot of the AILLA resource KPC003R000. At the bottom of the resource information is a list of filenames. Each filename contains the resource number plus an item number followed by the file format (for example, KPC003R000I001.pdf). When AILLA users click on and open the very first file, KPC003R000I001.pdf, they find a draft version of the table of contents from the book, reproduced with a list of corresponding AILLA resource numbers. Items 2 and 3 (I002 and I003) in this resource are translations of the table of contents into Spanish and Curripaco, respectively. The reader can consult these lists in any of these three languages in order to quickly and easily find the corresponding recording for a particular narrative. Also included in this same resource is an introduction to the collection in Spanish (I004), as well as two versions of the introduction in Curripaco (I005 and I006).

⁵ This language is also called Wakuénai.

	ation				
Resource ID	KPC003R000				
Collection	Curripaco Collection				
anguage(s)	Сиггірасо				
anguage Community					
itle					
nglish Title	Made from Bone				
panish Title	Hecho de Hueso				
ountry	United States				
lace	Illinois				
ate Created	2008				
escription	KPC003R000I01: Table of contents from book, English titles KPC003R000I002: Spanish titles KPC003R000I003: Curripaco titles KPC003R000I004: Introduction in Spanish KPC003R000I005: Introduction in Curripaco KPC003R000I006: Another version of the introduction in Curripaco				
ienres	Book				
o cite this esource	Hill, Jonathan (Author), University of Illinois Press (Publisher). (2008). "Made from Bone". Curripaco Collection. The Archive of the Indigenous Languages of Latin Am www.ailla.utexas.org. Media. text, image. Access: public. Resource: KPC003R000.				
ontributors mathan Hill mathan Hill		Colle	ositor ector		
onathan Hill niversity of Illinoi	s Press	Auth Publ	lisher		

Fig. 1. AILLA Resource Information for KPC003R000 (<u>http://www.ailla.utexas.org/search/resource.html?r_id=1644</u>).

Victoria Bricker's Mayan Language Collection (<u>http://www.ailla.utexas.org/search/</u> <u>collection.html?c_id=81</u>) includes the primary data that she used to write her 1973 book *Ritual Humor in Highland Chiapas*. Two sets of Tzotzil resources will be of particular interest to readers of the book. The first set, which includes the resources TZO004R003 (shown in Figure 2) and TZO004R007, encompasses a ritual performance on New Year's Day that is described in great detail in Chapter 2. Both of these resources contain recordings of the ritual, as well as

Resource Infor	mation				
Resource ID	TZO004R003				
Collection	Mayan Languages Collection				
Language(s)	Tzotzil				
Language Community	San Lorenzo Zinacantán, Chiapas				
Title					
English Title	Joking interactions during visit of Grandfathers and Grandmothers at house of Second Prefect				
Spanish Title	Interacciones en broma durante visita de Abuelos y Abuelas en casa de Prefecto Segundo				
Country	Mexico				
Place	San Lorenzo Zinacantán, Chiapas				
Date Created	1966-01-01				
Description	Beginning on Christmas Eve (24 December) and lasting through Epiphany (6 January), men costumed as Grandfathers, Grandmothers, Angels, and a Bull perform humorous rituals in several places in Zinacantan Center, including the Church of Saint Lawrence and the homes of cargoholders. The recording documents their performance in the home of the Second Prefect on New Year's Day in 1966. See Bricker (1973: Chapter 2) for a description of these rituals				
Genres	Speech play				
Original ID	TAPE 3A, 3B				
To cite this resource	Bricker, Victoria (Researcher), Anonymous (Transcriber). (1966). "Joking interactions during visit of Grandfathers and Grandmothers at house of Second Prefect". Mayan Languages Collection. The Archive of the Indigenous Languages of Latin America: www.ailla.utexas.org. Media: audio, text, image. Access: public, restricted. Resource: TZC004R003.				
Contributors Victoria Bricker Victoria Bricker Victoria Bricker LR		Depositor Collector Researcher Transcriber			
TZO004R00310	Filename	Type sample	Language	Access	Details
TZO004R00310		primary text	tzo	Level 1	Details
TZO004R00310		guide	eng	Level 2	Details
TZO004R00310		primary text	tzo	Level 1	Details
TZO004R00310		sample	tzo	Level 1	Details
TZO004R00310 TZO004R00310		primary text	tzo	Level 1 Level 1	Details Details
TZO004R00310		primary text transcription & translation	tzo tzo ena	Level 1	Details
	o tipat	u anacipuon o u analauon	120 ong	LOVOIT	Dorand

Fig. 2. AILLA Resource Information for TZO004R003 (<u>http://www.ailla.utexas.org/search/resource.html?r_id=2209</u>).

Tzotzil transcriptions accompanied by English translations that are organized like scripts for a play. The recordings found in the first resource (TZO004R003) were made in 1966 while the recordings found in the second resource (TZO004R007) were made in 1969. One can read the description of the ritual in the book (which is archived as resource TZO005R001 [http://www.ailla.utexas.org/search/resource.html?r_id=3062]), listen to the recordings from the two years, and compare the transcriptions to each other and to the corresponding passages (pp. 20-29) in the book.

Parts of Chapter 3, "The Festival of Saint Sebastian," in this same book are based on Bricker's raw data that are archived in four different AILLA resources. Two resources were recorded in 1966, TZO004R004 (http://www.ailla.utexas.org/search/resource.html?r_id=2210) and TZO004R005 (http://www.ailla.utexas.org/search/resource.html?r_id=2211), and two were recorded in 1969, TZO004R009 (http://www.ailla.utexas.org/search/resource.html?r_id=2215) and TZO004R010 (http://www.ailla.utexas.org/search/resource.html?r_id=2216). These resources also contain the recordings, along with Tzotzil transcriptions and English translations, and they correspond to the ritual speech described on pages 57-65 and 50-53 of the book, respectively.

When depositing her collection with AILLA, Bricker chose to keep anonymous the identities of the performers in this particular ritual. Thus, the only restricted files in these six resources are the ones that are listed as "guide" under the "Type" column (see Figure 2). The rest of the files in these resources can be viewed by anyone who is registered with AILLA and logged onto the website.

The collections described here are just the tip of the iceberg. AILLA contains a wide variety of materials on many different Latin American indigenous languages. We hope that the reader will take the time to explore the virtual archive and contact us with any questions about accessing or depositing files.

5. How AILLA Visitors Use the Data

The AILLA website does not track site users' activities or viewing histories, so we have no way of knowing who has viewed what. Instead we rely on the visitors themselves to tell us how and for what purposes they use the archive. In this section, we will describe some anecdotes that have been shared with us about ways in which AILLA users have used the archived data.⁶

Early in the archive's history, the first manager, Heidi Johnson, received a call from a local Austin kindergarten teacher. Her class was making paper versions of Kuna *molas*⁷ as a cultural enrichment project, and she wanted to find some Kuna music for the children to listen to during this activity. Following the recommendation of Dr. Johnson, the teacher downloaded

⁶ Detailed instructions for AILLA visitors that explain how to use the AILLA website can be found on the AILLA "How to Use the Archive" page (<u>http://www.ailla.utexas.org/site/howto_use.html</u>), so we will not go into those details here.

⁷ A *mola* is a geometric design that Kuna women embroider onto fabric panels. The term refers to both the design and the embroidered panels.

several songs and chants from Joel Sherzer's Kuna Collection (<u>http://www.ailla.utexas.org/</u> search/collection.html?c_id=54).

Many researchers use data archived at AILLA either to supplement their own data corpus or to find data on languages related to their own research languages. One linguistic researcher listened to recordings of survey interviews, searching for all occurrences of a particular morpheme, which he then compared to the cognate morpheme in his own research language. The same researcher has combed through transcribed narratives in related languages for occurrences of particular sequences of consonants. Other linguistic researchers have used transcriptions of archived recordings to search for particular grammatical constructions, while others have used some of the higher-quality recordings to do acoustic analyses.

Professors of both linguistics and anthropology have used archived recordings as examples of various types of speech events, such as whistled speech, ceremonial or ritual speech, storytelling, humorous speech, and so on. Similarly, professors who teach linguistic or anthropological field methods courses have used archived data for examples of both best and worst practices for such field necessities as metadata collection and audio and video recording techniques. Some professors who teach courses on endangered and/or indigenous languages require their students to utilize materials that are archived at AILLA when conducting research for their term papers.

A few linguists who work on language revitalization projects in Central and South America have told us that members of some indigenous communities use the data archived at AILLA to create word lists, dictionaries, story books, and other teaching materials. However, the most touching anecdotes come from the archive users who are also members of indigenous communities. They browse the archive to hear the voices and words of their ancestors, family, friends, and neighbors. Several times a year, we receive emails from community members who have just listened to the voices of their loved ones who have passed on; they write to tell us how thankful they are that AILLA has preserved these voices for the present and the future.

> Archive of the Indigenous Languages of Latin America LLILAS Benson Latin American Studies and Collections The University of Texas at Austin

References

Bricker 1973	Victoria Bricker. <i>Ritual Humor in Highland Chiapas</i> . Austin: University of Texas Press.
Hill 2009	Jonathan Hill. Made from Bone: Trickster Myths, Music, and History from the Amazon. Urbana: University of Illinois Press.
Sherzer and Urban 1986	Joel Sherzer and Greg Urban, eds. <i>Native South American Discourse</i> . Berlin: Mouton de Gruyter.

Sherzer and Woodbury 1987

Joel Sherzer and Anthony C. Woodbury, eds. *Native American Discourse: Poetics and Rhetoric*. Cambridge: Cambridge University Press.