

Announcing the curated collections of Terrence Kaufman at AILLA

Susan Smythe Kung and J. Ryan Sullivant
Archive of the Indigenous Languages of Latin America,
University of Texas at Austin

January 3, 2019

SSILA Annual Meeting, New York City, Jan. 3-6, 2019

Introduction

AILLA is a dynamic digital repository of multimedia resources in and about the indigenous languages of Latin America. It was founded in 2001 by Joel Sherzer and Anthony Woodbury at the University of Texas at Austin. It has interfaces in both English and Spanish, making it accessible to a large part of Latin America. AILLA's holdings include audio and video recordings; transcriptions, translations, and other annotated documents; photos, maps, and sketches; field notes, grammars, dictionaries, ethnographies, manuscripts and educational materials, all in digital formats.

AILLA is completely free to use. Metadata is open access, meaning that anyone connected to the internet can read the catalog entries. However, to view or download files, a user must create a free account. AILLA has both restricted and unrestricted access levels, and the level of restrictions is listed in the filelist. Level 1 is Public Access, meaning that there are no restrictions on viewing or downloading these files after the user logs in. Restricted access levels are 2, 3 and 4. You can find more information about these access levels in the Frequently Asked Questions section, which is where you'll also find information about creating an account, changing your password or navigating the site.

Archiving the Terrence Kaufman Collection

In 2012, AILLA was awarded an NSF-DEL grant entitled "Archiving the Terrence Kaufman Collection".¹ This was a 3-year grant and we had 3 no-cost extensions, so the project ran from 2012 to 2018. It kicked off with an acquisition trip to Pittsburgh in Fall 2012 to get the first batch of materials; a second trip followed in Spring 2013 to get the rest of the materials. Once the collection was at AILLA, Kung did all of the initial processing and organization, and then most of the digitization, metadata collection, and data ingestion was done by eight graduate research assistants and seven undergraduate interns. The final push to finish included working with

¹ NSF Acknowledgement: This material is based upon work supported by the National Science Foundation under Grant Number BCS-1157867. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Kaufman to determine access, contacting all linguists who had created some of the materials, writing collection descriptions, and setting access levels.

The eight graduate research assistants in chronological order were Stephanie Villard, Ryan Sullivan, Justin McIntosh, Telma Can, Natalia Bermúdez, Wikaliler Smith, Jaime Pérez González, and Ambrocio Gutiérrez Lorenzo. The seven undergraduate interns in chronological order were William Vásquez, Paul Aguilar, Natali Arellano, Nishmet Montelongo, Judith Lara, Rebeca Martínez, and Flor Rodríguez

The physical assemblage of materials included 138 boxes of fileslips, 1067 CDs, 501 cassette tapes, 94 mini disks, 66 reel-to-reel tapes, 18 Hi-8 tapes, 7 mini digital video tapes, 393 floppy disks, 18 linear feet of paper (approx. 36,000 sheets), 320 digital files, and 3 oversized maps.

As we worked to digitize and organize the physical assemblage, we realized that the materials, which Kaufman had collected over the course of 50 years, were too diverse to put into just one collection in AILLA. Thus, over the course of the next 5 years, we built 13 different collections, which I will summarize.

The 13 collections include data on 119 distinct languages, as well as approximately 75 dialectal varieties. These languages and varieties represent 26 languages families, 4 isolates, and 1 pidgin, and they come from 22 countries extending from Canada to Argentina. Language families with the highest representation are Mayan (Mexico and Guatemala), Mixe-Zoquean (Mexico), Otomanguean (Mexico), and Uto-Aztecan (Mexico and USA).

1. Mayan Languages Collection of Terrence Kaufman

AILLA PID: [243664](#)

This collection consists of materials collected or compiled by Kaufman as he conducted research on Mayan languages in Mexico and Guatemala. These materials were gathered between 1959 and 2004, with most materials dating from 1960 to 1993. The holdings include audio recordings, transcriptions, fileslips, survey questionnaires, articles and manuscripts. Some of the primary data were gathered by Kaufman himself (especially for Huasteco, Mochó, Tzeltal, and Tzotzil), while other primary sources were collected by other people and given to Kaufman (e.g. the Mayan Vocabulary Survey that he wrote in 1962 and revised in 1964, then sent out to everyone he knew of that was working on a Mayan language). Kaufman also compiled fileslips and data sets from published and unpublished sources.

Most Mayan languages are represented in this collection, however, many are only represented by a relatively small number of forms in comparative works. The collection contains 593 folders of 410 audio recordings and 950 documents. For a more detailed breakdown of the format types for each language, as well as a chronology of Kaufman's Mayan research, you can consult the collection description page.

This collection does not include any Mayan materials that resulted from either of the large-scale organized projects that Kaufman directed, specifically the Francisco Marroquín Linguistic Project or the Project for the Documentation of the Languages of MesoAmerica Collection.

Most of the items in this collection are Public Access, but a few are restricted.

2. Francisco Marroquín Linguistic Project

AILLA PID: [124507](#)

Next is the Francisco Marroquín Linguistic Project collection. This project ran from 1970 to 1978 in Guatemala and Kaufman was the director. It is often referred to in English by its Spanish acronym PLFM (Proyecto Lingüístico Francisco Marroquín). The PLFM was a private, autonomous, Guatemalan non-profit association; and the US linguists who worked on the project were Peace Corps volunteers. This collection contains 102 folders of 420 digital documents. The data in this collection generally fall into one of three categories: (1) Materials related to the Guatemalan Mayan Dialect Survey, (2) materials about the PLFM itself, such as memos and organizational documents, and (3) notes and maps about particular Mayan languages or subsets of languages. All materials in this collection were digitized from paper documents only. No audio recordings from the PLFM are known to have survived the civil war in Guatemala.

All items in this collection are Public Access.

3. Huasteca Nawa Collection of Terrence Kaufman

AILLA PID: [247464](#)

This collection contains materials related to Kaufman's self-funded research on the Nawa language varieties spoken in the Huasteca, a region of Mexico encompassing parts of San Luis Potosí, Veracruz and Tamaulipas states. Kaufman collected these materials between 1983 and 1993, with the majority of them dating from 1987 and 1990. This collection includes 104 folders with 240 audio recordings and 292 documents, including hand-written transcriptions, file slips and manuscripts.

All audio recordings in this collection are Public Access while all other files (including transcriptions, file slips and manuscripts) are currently Restricted, as Kaufman is still actively working with these data. He will consider granting access on a case-by-case basis.

4. Project for the Documentation of the Languages of MesoAmerica

AILLA PID: [119927](#)

Next is the collection from the Project for the Documentation of the Languages of MesoAmerica (also known as the PDLMA), which was a large-scale effort to document many of the languages of southern Mexico. From 1993 until 2010, the PDLMA brought together speakers of indigenous Mesoamerican languages with academic linguists for summer field seasons in various locations in Mexico, including Fortín de las Flores and Catemaco, Veracruz; Oaxaca City, Oaxaca; and San Cristóbal de las Casas, Chiapas.

The PDLMA documented 30 different language varieties from the Mije-Sokean, Nawa, Mayan, Totonac-Tepehua, and Otomanguan (including Sapoteko, Chatino, Tlawika, and Matlatzinka) language families. Only a subset of the total languages were worked on each summer (anywhere from 3 to 11). The PDLMA employed approximately 45 different linguists all told and at least as many native speaker consultants. For a detailed list of the PDLMA languages that includes the names the linguists who worked on them and the years they were worked on, please see the collection description page.

This is by-far the largest single collection in AILLA with 1,596 different folders and 11,126 individual files that include audio and video recordings, scans of file slips, elicitation notebooks, project memos, and analyses in the form of handouts from presentations that were given at the weekly language group meetings. There are also recordings and handouts from various workshops that Kaufman conducted each summer, on topics such as sound symbolism, Spanish loanwords, ethnobiology, and ethnomedicine.

Many items in this collection are Public Access, while others are Restricted. Public access materials include all project memos, handouts from individual presentations at Group Meetings, and all workshop recordings and handouts. All audio and video that Kaufman was responsible for recording are public access. Kaufman left the decision to lift restrictions on audio, video and fieldnotes up to the individual linguists who made these data; of the linguists that we were able to find, most of these people have given us permission to lift restrictions; others are reviewing the materials before making a decision (either alone or in collaboration with members of the speech community); still others have never responded to our emails. Finally, all file slips are restricted as Kaufman is negotiating the publication of associated dictionaries.

PDLMA-ASSOCIATED LANGUAGE SURVEYS

Starting in 2003, Kaufman devoted some of the PDLMA resources each summer to conducting dialect surveys for 5 different language groups. Each of these dialect surveys is contained in its own collection. Chronologically, the first of these was the ...

5. Totonac-Tepehua Dialect Survey

AILLA PID: [124508](#)

This collection was planned and written by Carolyn MacKay, Frank Trechsel, and Kaufman and conducted from 2003 to 2005. The questionnaire was administered and recorded by Benigno Robles Reyes, who is a Huasteco Maya speaker, in a total of 29 Totonac-speaking towns and 1 Tepehua-speaking town in the states of Veracruz, Puebla and Hidalgo. During the summers of 2004 and 2005, the recorded surveys were transcribed by José Santiago and Miguel Gerónimo, both of whom are Totonac speakers from Filomeno Mata, Veracruz. The majority of the Totonac varieties found in this collection do not have ISO 639-3 codes, so many of the materials are cataloged generically under Totonac.

This collection features 62 folders that contain either the audio recordings and transcriptions of the survey questionnaire or audio recordings of narratives, conversations, and songs collected in the communities where the surveys were done. There are also photos of the Totonac-Tepehua region, a PDF master list of the communities, and a PDF of a blank questionnaire.

All files are Public Access except for the transcriptions of the questionnaires; the questionnaire transcriptions will remain restricted while MacKay and Trechsel check the accuracy of the transcriptions and data. Though the transcribers, Santiago and Gerónimo, are both native speakers of Filomeno Mata Totonac, they did not mark vowel length or laryngealization. Furthermore, since the interviewer, Robles Reyes, does not speak either Totonac or Tepehua, there are instances of misunderstandings between interviewer and interviewee.

6. The Survey of Zapotec and Chatino Languages Collection

AILLA PID: [243980](#)

The Survey of Zapotec and Chatino languages was planned and written by Kaufman and Mark Sicoli from 2005 to 2007. In 2007 and 2008, they trained approximately 35 Zapotec speakers to administer and transcribe the survey, which was then administered in 104 Zapotec or Chatino-speaking towns of Oaxaca in 2008 and 2009. The transcriptions were completed in 2010. The data collection for this dialect survey was done under the auspices of the PDLMA, which was funded by the National Institute of Indigenous Languages of Mexico (INALI) from 2007 to 2010. Additionally, Mark Sicoli had an NSF DEL grant (#BCS-1263671) to finalize the transcriptions and to organize these materials for deposit into AILLA. Of the 148 folders in this collection, 122 folders contain audio-recorded and transcribed questionnaires for Zapotec languages and 23 of them are questionnaires for Chatino. Additionally, there is a PDF of a blank questionnaire, and a PDF of the grant proposal to INALI.

All materials are Public access.

Three dialect surveys were carried out at the PDLMA in 2010.

7. Gulf Nawa DS

AILLA PID: [124449](#)

The first of these three collections is the Gulf Nawa Dialect Survey, which was planned by Una Canger, Roberto Zavala, Valentín Peralta, and Terrence Kaufman to study the lexical and grammatical variation among the varieties of Nawat spoken on Mexico's gulf coast in Tabasco and Veracruz states. Surveys were conducted in 3 municipalities in Tabasco and in 8 municipalities in Veracruz. This collection features 26 folders containing mostly audio recordings. There are a few video recordings and photographs of the towns and people represented in the surveys. Also included in this collection is dialect variation research done by Canger in 2004 prior to the creation of the Gulf Nawa Dialect Survey project itself (notes on Paso de Cupilco Nawat phonology).

All items in this collection are Public Access.

8. Yokot'an (Tabasco Chontal) Dialect Survey

AILLA PID: [124448](#)

The second of the three dialect surveys that were done in 2010 is the Yokot'an Dialect Survey, which was planned by Kaufman and Brad Montgomery-Anderson (who was the only linguist to work on this language for the PDLMA). Yokot'an is a Mayan language also known as Tabasco Chontal, and it is frequently confused with the unrelated Chontal languages of Oaxaca. Surveys were conducted in 10 communities in the Mexican state of Tabasco. The collection consists of 14 folders containing audio recordings, document files, and 26 digital images of the towns and people represented in the survey.

All items in this collection are restricted at this time as we are waiting to hear back from Montgomery-Anderson.

9. Soke Dialect Survey

AILLA PID: [124516](#)

The third and final dialect survey done at the PDLMA in 2010 was the Soke Dialect Survey. It was planned by Roberto Zavala and Kaufman. The Soke Dialect Survey concerns itself only with the Oaxaca Zoque, Chiapas Zoque, and Jitotoltec and does not contain material related to the Gulf Zoquean languages of Sotepan, Ayapan, or Texistepec. Surveys were conducted in 3 communities in Tabasco, 3 in Oaxaca, and 18 in Chiapas. The collection consists of 30 folders containing audio recordings, photos of the towns and people represented in the survey, and a blank questionnaire.

All items in this collection are Public Access.

Now we are leaving behind the PDLMA and moving on to the other collections, starting with ...

10. Educational Materials Collection of Terrence Kaufman

AILLA PID: [247465](#)

This is a collection of syllabi, course descriptions, and class assignments and activities about indigenous languages that were created and used by Kaufman during the course of his teaching career. While the list of languages mentioned in these materials is quite extensive and ranges from Canada to Paraguay, most of them are represented by reproductions of small data sets. The materials are from 1964 to 2006, with most of the materials dating from the 1990s and later. The collection has 16 folders containing audio recordings of lectures and digital documents of course syllabi, course proposals, lecture handouts, and class exercises. Each folder contains materials related to a single course offered by Kaufman. When courses were offered multiple years, the materials for each instance have been placed in the same folder in this collection.

All items in this collection are Public Access.

11. Latin American Languages Collection of Terrence Kaufman

AILLA PID: [124495](#)

This collection contains materials that do not fit well into one of the other collections of language data. These materials were collected, compiled, or created by Kaufman over the course of his career, from 1963 through 2013. These materials were not collected under the auspices of an organized project nor do they include languages that were part of Kaufman's sustained independent research. In total, there are 43 folders containing 80 text documents, 202 files containing scanned images of file slips, one photo and one video.

The 43 folders in this collection contain either analytical documents written by Kaufman about one or more Latin American languages or scanned images of file slips that were collected or compiled by Kaufman. The data on these file slips derive from 2 sources: First, there is data that Kaufman collected during isolated elicitation work (as in the Suruí and Dení file slips). Second, there is data that Kaufman compiled from other sources; in these cases Kaufman grouped together perceived cognates and often made annotation on the file slips themselves (as with the Hokan file slips).

Many of the languages represented in this collection are found in only one resource. For a detailed listing of the languages and language groups that appear multiple times, please see the collection page.

Many items in this collection are Public Access, whereas others are Restricted. (Kaufman will consider granting access on a case-by-case basis.)

12. Epigraphy Collection of Terrence Kaufman and John Justeson

AILLA PID: [248167](#)

This collection contains materials related to Terrence Kaufman and John Justeson's work on the writing systems of ancient MesoAmerica, and in particular the Epi-Olmec script. These materials were created between 1993 and 2004. This collection contains 9 folders of 15 files containing documents about Epi-Olmec decipherment, Zapotec hieroglyphic writing, and the texts at Xochicalco.

Some items in this collection are Public Access, and others are Restricted while Kaufman and Justeson continue their analysis.

13. Archiving the Terrence Kaufman Collection

AILLA PID: [245023](#)

This collection includes all materials that were the direct result of this NSF grant project. It includes the grant proposal, requests for supplemental funds and no-cost extensions, conference papers and posters about the collection and its history, and photos of some of the acquisition and curation processes.

All files in this collection are public access.

AILLA Collections Related to the PDLMA

Many people who worked on the PDLMA had already archived the data that they had collected during the PDLMA before AILLA staff started curating Kaufman's vast assemblage. These collections are listed in the order that they were deposited into AILLA:

- San Miguel Chimalapa Zoque Collection of Heidi Johnson (ailla:124411)
- Huehuetla Tepehua Collection of Susan Smythe Kung (ailla:124437)
- Chiapas Zoque Collection of Daniel Suslak (ailla:124438)
- Sierra Popoluca Collection of Lynda Boudreault (ailla:124386)
- Southern Zapotec Languages Collection of Rosemary Beam de Azcona (ailla:124376)
- MesoAmerican Languages Collection of Roberto Zavala Maldonado (ailla:124461)
- Isthmus Zapotec Collection of Gabriela Pérez Báez (ailla:124428)
- Northern Zapotec Collection of Erin Donnelly (ailla:124501)

AILLA Collections of (some of) Kaufman's collaborators

AILLA also holds the collections of many of Kaufman's collaborators; again these collections are listed in the order that they were deposited into AILLA:

- Mayan Languages Collection of Nicholas Hopkins (ailla:124472)
- Mayan Languages Collection of Laura Martin (ailla:124421)
- MesoAmerican Languages Collection of Lyle Campbell (ailla:124470)
- Mayan Languages Collection of Nora England (ailla:124483)
- MesoAmerican Languages Collection of Kathryn Josserand (ailla:124466)
- Acquisition of Mayan Languages Collection of Clifton Pye (ailla:124375)
- The Works of Thomas Cedric Smith Stark (ailla:124510)
- Sochiápam Chinantec Whistled Speech Collection of Mark Sicoli (ailla:242622)

Please note that the data that Kaufman collected on Xinca (a non-Mayan language of Guatemala) were previously deposited by Lyle Campbell in the MesoAmerican Languages Collection (ailla:124470).

La'ay la'tziwch
Thank you
Gracias