
Plan Overview

A Data Management Plan created using DMPTool

Title: The Jakalteko language of Guadalupe Victoria: documenting a highly endangered dialect of the Mayan language Jakalteko-Popti' [jac] in Chiapas, Mexico

Creator: GRANT ARMSTRONG

Affiliation: University of Wisconsin–Madison (wisc.edu)

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: 44070

Template: NSF-SBE: Social, Behavioral, Economic Sciences

Project abstract:

The Mayan language Jakalteko-Popti' is spoken by approximately 34,000 people in Guatemala and around 400 in Mexico. While ample documentation and analysis of the language has been done by researchers and national linguistic organizations in Guatemala, the dialect spoken in Mexico has not been recorded and analyzed. Since nearly all of the roughly 400 speakers in Mexico, who reside primarily in the town of Guadalupe Victoria in Chiapas, are elderly, the language is classified as nearly extinct or highly endangered according to institutions such as Ethnologue and the National Institute of Indigenous Languages in Mexico (INALI). For this reason, producing quality recordings of this dialect and performing basic linguistic analyses to pinpoint any differences that might have evolved between Guatemalan and Mexican dialects is of paramount importance. The objective of this project is to record video and audio of the Jakalteko language of Guadalupe Victoria. These recordings will have two important uses. First, they will be archived for public use and will serve as a chronicle of the linguistic and cultural history of the town. Second, they will be used to identify any linguistic differences between this speech community and those in Guatemala from which the Jakalteko speakers of Guadalupe Victoria have been separated for approximately 130 years, thus providing a unique data set for linguists to study linguistic change and dialect divergence.

Last modified: 08-26-2019

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

The Jakalteko language of Guadalupe Victoria: documenting a highly endangered dialect of the Mayan language Jakalteko-Popti' [jac] in Chiapas, Mexico

Roles and responsibilities

The DMP should outline the rights and obligations of all parties as to their roles and responsibilities in the management and retention of research data. It should also consider changes to roles and responsibilities that will occur should a principal investigator or co-PI leave the institution or project. Any costs should be explained in the Budget Justification pages.

I, Grant Armstrong, am the sole PI of this project. I will be responsible for recording video and audio, preparing elicitation questionnaires, writing up field notes and producing metadata sheets. All video and audio will be transcribed and translated using the ELAN (EUDICO Linguistic Annotator), which I will do with the help of a native speaker assistant in the field. I will be responsible for saving all media and text files in the field on SD cards and external hard drives, distributing these in the community in a convenient medium for local use, and then depositing all files in an archive once all accompanying text files for media have been completed. The fellowship stipend will be used to purchase video and audio recording equipment, a durable laptop computer for use in the field, noise-cancelling headphones and splitter for transcription work with native speaker assistants and SD cards and external hard drives with excess of 200 GB of storage space.

Expected data

The DMP should describe the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project. It should then describe the expected types of data to be retained.

Three types of data will be produced as a result of this project: (i) approximately 30 hours of video, (ii) approximately 150 hours of audio and (iii) accompanying text files that will contain transcriptions, translations and any other relevant annotations related to the media files. It is anticipated that all of this data will be retained, archived and made publicly accessible with no restrictions. Video and audio recordings and accompanying transcriptions will serve as the first documentation of this kind from Guadalupe Victoria. They will be archived in the community library and with each participant who produced the recording in more accessible formats such as CD/MP3 for audio, DVD/MP4 for video and physical print outs/txt files for transcriptions and translations. The main reason for using these formats for local use is because internet access is not reliable in the community due to its geographic location. The files will be deposited as uncompressed media file types (MPEG video and WAV for audio) and eaf-xml for accompanying text in an archive. The responses to the linguistic questionnaires complement existing work on dialectal variation done on Jakalteko-Popti' within Guatemala. They will also be stored in the archive and made publicly available so that cross-dialectal comparison can be easily done by speakers and researchers in both countries.

Period of data retention

SBE is committed to timely and rapid data distribution. However, it recognizes that types of data can vary widely and that acceptable norms also vary by scientific discipline. It is strongly committed, however, to the underlying principle of timely access, and applicants should address how this will be met in their DMP statement.

This project will have two phases. The first phase involves recording narratives, demonstrations, and personal histories and other speech genres. I estimate that recording, transcribing and translating these texts should take three months. All recordings will be timely deposited in an archive for public use. The second phase involves asking speakers to complete a set of linguistic questionnaires designed to identify any potential differences between the dialect under study in Mexico and other dialects of Jakalteko-Popti' spoken in Guatemala. This phase will last approximately 5 months. I plan to deposit results of these questionnaires in 4 blocks that correspond to different topics of interest: phonetics-phonology, morphology, syntax and lexicon. In sum, I anticipate that the project will involve 8 months of work and that all data will be accessible through an archive after 12 months. The data will be made available incrementally as the project progresses.

Data format and dissemination

The DMP should describe data formats, media, and dissemination approaches that will be used to make data and metadata available to others. Policies for public access and sharing should be described, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. Research centers and major partnerships with industry or other user communities must also address how data are to be shared and managed with partners, center members, and other major stakeholders.

All video files will be saved in MPEG-2 format, audio in WAV format. Text files that accompany the media files will be saved in eaf-xml format or as txt files. These files will be uniquely labeled using metadata sheets that contain information regarding the identity of all parties involved in each recording, place and time, and any other information pertinent to specific recording. All recordings, accompanying text files and associated metadata will be archived in publicly available folders. All media files and accompanying text files will be produced with informed consent of all participants. Part of obtaining informed consent will involve asking each participant if he or she wishes to remain anonymous or be identified by name in the recordings. For those wishing to remain anonymous, names will be removed from all publicly available data linked to their recordings, while those who wish to be identified will be publicly acknowledged in the archive or any publications for their contributions. It is anticipated that all data produced as a result of the project will have no access restrictions.

Data storage and preservation of access

The DMP should describe physical and cyber resources and facilities that will be used for the effective preservation and storage of research data. These can include third party facilities and repositories.

While collecting data in the field, media files will be stored on SD cards and backed up on 3 external hard drives. Accompanying text files will be created using a laptop computer and saved to 3 external hard drives. All external hard drives will be password protected. Once accompanying text files have been created and appropriately linked to media files, these will be deposited to a linguistic archive that will preserve and manage access to the files. An additional backup will be stored on the hard drive of a desktop computer owned by the University of Wisconsin-Madison, which is password protected. All other copies will be subsequently deleted.

Additional possible data management requirements

More stringent data management requirements may be specified in particular NSF solicitations or result from local policies and best practices at the PI's home institution. Additional requirements will be specified in the program solicitation and award conditions. Principal Investigators to be supported by such programs must discuss how they will meet these additional requirements in their Data Management Plans.