FAMSI © 2002: Jonathan D. Amith

Cultural and Pedagogical Lexicography of Modern Náhuatl

Research Year: 2001
Culture: Guerrero
Chronology: Contemporary
Location: Central Guerrero, México
Site: Balsas River Valley

Table of Contents

Abstract

The grant award was for the digital recording of headword entries for a dictionary of several dialects of Náhuatl spoken in the Balsas River Valley of central Guerrero, México. The first stage of recording had been completed during summer 2001 in New Haven with the participation of two speakers from San Agustín Oapan. The second stage of recording was completed during summer 2002 in New Haven with Cristino Flores, a Náhuatl consultant from the neighboring village of Ameyaltepec.

Resumen

La asignación de la beca fue para la grabación digital de entradas de palabras principales para un diccionario de varios dialectos del Náhuatl hablado en el Valle del río Balsas de Guerrero Central, México. La primera faceta de la grabación había sido completada durante el verano del 2001 en New Haven con la participación de dos portavoces de San Agustín Oapan. La segunda faceta de la grabación se completó durante el verano del 2002 en New Haven con Cristino Flores, un consultor del náhuatl del vecino pueblo de Ameyaltepec.

Submitted 09/16/2002 by:
Jonathan D. Amith
jonathan.amith@yale.edu

Introduction

As with the 2001 recordings, given the high cost of renting professional sound recording studios in México (approximately $20–40 US/hour) and the necessity of close contact with the computational linguists at the Linguistic Data Consortium, University of Pennsylvania, who designed the program for time-coding the transcript of each recording session and for subsequent segmenation, tagging, and linking, it was determined that the most cost-efficient and secure way to ensure high quality taping would be to take advantage of the professional recording facilities and technical assistance offered pro bono by Haskins Laboratory in New Haven. This year (2002) the recording was done directly to hard disk and the results burned on CD. This facilitated the transfer to the LDC server and subsequent segmentation and tagging of the sound files. Given the high cost of the project (approximately $10,000, not including any reimbursement to Amith), to complete the Ameyaltepec recordings it was necessary to supplement the remaining money of the FAMSI grant ($1,500) with additional funds from the following institutions and grants:

- Center for Latin American Studies, University of Chicago, which earmarked part of its Title VI funding for Náhuatl recordings (this covered transportation, housing and per diem expenses of Cristino Flores);

- Jonathan Amith, who applied part of his funds from an International Research and Fellowship Program, U.S. Dept. of Education, to the recording project (this covered additional honorarium for Cristino Flores as well as some travel and per diem);

- Steven Bird, Linguistic Data Consortium, University of Pennsylvia, who applied part of his National Science Foundation grant (this covered the entire cost of uploading, segmentation, tagging, linking, CD burning, etc.).

Since the time the Oapan Náhuatl recordings of the summer 2001 were carried out (the status of research when the interim report was submitted), the following advancements have been made, as outlined in the next two sections of this final status report.

Oapan Sound Recordings (post-recording processing)

- The Oapan audio recording material (on digital audio tape) was uploaded to a central computer at the LDC;

- Filters were applied to insure the best possible sound for internet distribution while the original files were maintained for phonetic analysis;

- Original and filtered sound files were segmented and the segmented files (approximately 25,000 for the original and filtered versions) were tagged with unique references to each dictionary headword;

- Segmented recordings were normalized and downsampled (from 44,100 to 22,050); this set (approximately 25,000 sound files) was then converted to .mp3, yielding an additional 25,000 recording files;

- From approximately 6 tokens of each Oapan word (3 by a female consultant and 3 by a male), two were chosen (one female and one male) for linking to the online dictionary;

- The female and male forms were concatenated into a single file and versions of this file were placed on a server in both .mp3 and downsampled .wav formats.

Ameyaltepec Sound Recordings (recording and post-recording processing)

- The Ameyaltepec material was recorded directly onto a hard disk and burned on CDs (approximately 60 hours of recording); this material was uploaded to a central computer at LDC;

- Sound files were segmented and the segmented files (approximately 22,000) were tagged with unique references to each dictionary headword;

- Segmented recordings were normalized and downsampled (from 44,100 to 22,050); this set was then converted to .mp3, yielding an additional 22,000 recording files;

- From approximately 3 tokens of each Ameyaltepec word one was chosen for linking to the online dictionary;

- This linked set (in downsampled .wav format) was then converted to .mp3 and both formats were placed on a web-accessible server at the University of Pennsylvania.

The Linguistic Data Consortium paid for a computer programmer who wrote a PHP script to present the Náhuatl dictionary online and automatically create links to the Oapan and Ameyaltepec sound files. This dictionary was placed online and will soon be made operational at http://www.ldc.upenn.edu/nahuatl by clicking on the link to Hyperlex2.

Conclusion

Thus, as proposed last year, the final recordings were completed in the summer of 2002 and the original terms of the FAMSI grant were achieved. There were also additional developments not foreseen in the original proposal. Susan Guion, a linguistics professor at the University of Oregon, has been given a complete copy of the segmented sound files (approximately 47,000 files each with a unique name that reflects the reference number of the headword that they exemplify). She will use this material in her phonetics class and begin work on a phonetic description of Oapan and Ameyaltepec Náhuatl. The same sound files were also used by Chris Manning, a linguist and computer scientist at Stanford, in developing a prototype model of Kitlkitl, an adaptation for Náhuatl of Kirrkirr, which has been successfully used for indigenous language instruction in Australia. Finally, the sound files will be integrated into a language documentation and literacy program coordinated by Amith that the Ford Foundation is sponsoring in the Náhuatl-speaking villages of the Balsas River Valley.

Again, thanks very much to FAMSI for the kind support for this project.