Geographical metadata in the California Language Archive

Amy Campbell, Andrew Garrett*, Hannah Haynie, Justin Spence, Ronald Sprouse, and John Sylak (University of California, Berkeley)

General background
The California Language Archive integrates the catalogs and the digital content of two archival repositories at Berkeley:

- Berkeley Language Center (blc.berkeley.edu)
- Survey of California and Other Indian Languages (linguistics.berkeley.edu/~survey)

The CLA is in development and will be online in February 2011. In the future, other language archives may also contribute to the CLA (such as the Bancroft Library and the Hearst Museum at Berkeley).

Berkeley Language Center
An archive of audio recordings deposited since about 1950. From languages throughout the Americas with a special focus on California

- About 1.730 hours of audio recordings (cataloged to date)
- Original media: reel-to-reel tapes, cassette tapes, DATs, CDs, video
- 7,842 audio segments in 221 collections, e.g. 'Catherine A. Callaghan & Mrs. Aza McCauley, 1960-1961'
- Digitization status: all cataloged material is digitized

Survey of California and Other Indian Languages
An archive of paper materials deposited since about 1950. From languages throughout the Americas with a special focus on California

- About 200 linear feet of material
- Contents include field notebooks, unpublished manuscripts, letters and notes, lexical files, etc.
- 2,451 distinct catalog items in 186 collections
- Digitization status: currently >50,000 scanned images from 200 items in 29 collections

Credits
*Corresponding author (garrett@berkeley.edu). BLC and SCOIL metadata collaborators and interlocutors include Lisa Comahan, Marceline Gamett, Susan Stone, and Maziar Toosarvandani. Photos: Halpern MS image — SCOIL, MS Halpern 009.014.0002, Annie Burke photo from Pomo Basketmaking (1972); Sulphur Creek photo from Bill Tuthill, http://www.carpevms.com/sulc-big.htm

Sources of metadata
Contributing repositories primarily hold legacy material, from which we glean geographical information. (New depositories will be asked to supply geographical metadata.)

Berkeley Language Center
Documentation for LA007: Northern Sierra Miwok
Recorded by Catherine A. Callaghan & Mrs. Aza McCauley, 1960-1961

Survey of California and Other Indian Languages
Abraham M. Halpern, field notebook XIV, page 1
Southern Pomo notes with Annie Burke, 7 May 1940

Geographical metadata
Metadata about people
The CLA database includes biographical information about researchers and linguistic consultants:

- Born at [PLACE, DATE]
- Lived at [PLACE, DATE]
- Ethnic affiliation with [PLACE]
- From/raised at [PLACE]
- Father from [PLACE]
- Mother from [PLACE]

One PLACE per individual can be tagged for dialectological priority.

Metadata about items (field notebooks, etc.)
Legacy material often gives no biographical or geographical information about consultants; sometimes they are not even named. We record:

- Fieldwork PLACE

User interface
Map-based searching
Exploring a map interface shows items with geographical metadata in the CLA database:

- The map pins distinguish audio vs. paper material with vs. without digital content.

Maps in search results
A search for Southeastern Pomo material with digital content yields the following partial result:

The map shows places for items with geographical metadata in the CLA database.