Language archives: An introduction

Andrew Garrett
University of California, Berkeley

Breath of Life Workshop
4 June 2012

Outline

What's an archive?
> Two misconceptions about archives
> Actual archives and archivists

Major language archives for California languages
> Archives at Berkeley
> Archives around the US
> Tribal archives: The future of language archiving

Access, ethics, and cultural property

What's an archive?

- ▶ Two misconceptions about archives
    - ▶ Sometimes these misconceptions are actually true ...
    - ▶ ... but mostly they aren't!
- ▶ Actual archives and archivists

Misconception #1: You can't come in!

Misconception #2: We couldn't find it even if we let you in

Actual archives and archivists

What archivists do

- ▶ Access
    - ▶ Make materials available to interested people

- ▶ Description
    - ▶ Figure out contents: language, subject matter, time & place

- ▶ Preservation
    - ▶ Keep materials safe for the next generation
    - ▶ . . . for the next 100 years
    - ▶ . . . for the next 500 years . . .

Archivists are devoted to all three — including **broad access**!

Actual archives and archivists

What archives are like

- ▶ Access

    - ▶ Complex registration process, because **anyone** can use an archive: no library card required

- ▶ Description

    - ▶ Usually a good catalog description of materials

- ▶ Preservation

    - ▶ What you seek is actually there!
      . . . and in good condition!
      . . . even if it's 200 or 500 or 1000 years old . . .

Major language archives for California languages

Documentation about California (and related) languages:

- ▶ Four major archives at UC Berkeley

    - ▶ Bancroft Library
    - ▶ Phoebe A. Hearst Museum of Anthropology
    - ▶ Berkeley Language Center
    - ▶ Survey of California and Other Indian Languages

- ▶ Three major archives elsewhere in the US

    - ▶ National Anthropological Archives, Smithsonian Institution
    - ▶ American Philosophical Society (Philadelphia)
    - ▶ Santa Barbara Museum of Natural History

Archives at Berkeley

- ▶ Bancroft Library
  - ▶ paper records from about 1800 to the mid-1950s (and later)

- ▶ Phoebe A. Hearst Museum of Anthropology
  - ▶ sound recordings from 1901 to about 1940 (and later)

- ▶ Berkeley Language Center
  - ▶ sound recordings from about 1950 to today

- ▶ Survey of California and Other Indian Languages
  - ▶ paper records from about 1950 to today

Bancroft Library

Two major collections

- ▶ Papers of A. L. Kroeber
  - ▶ Online: 'Guide to the A. L. Kroeber Papers'

- ▶ Papers of the Department of Anthropology
  - ▶ Online: 'Finding Aid to the Ethnological Documents of the Department and Museum of Anthropology'
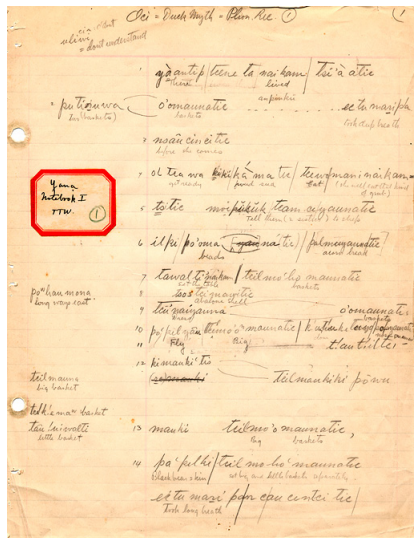
- ▶ C. Hart Merriam Collection

Many other Bancroft collections are relevant for specific languages and people, but the above collections contain language material for almost all California languages.

Bancroft Library

Representative collection:
T. T. Waterman's field notebooks

- ▶ TTW (Ph.D. Columbia 1913)

  - ▶ at Berkeley 1907–1918,
    1920–1921
  - ▶ d. 1936 while Territorial
    Archivist of Hawai'i

- ▶ Linguistic fieldwork

  - ▶ 6 Yahi notebooks
  - ▶ 16 Yurok notebooks
  - ▶ also Kumeyaay, Luiseño,
    Makah, Tubatulabal

- ▶ Related material: wax cylinder
  recordings in the Hearst Museum

Waterman's 1914 transcript of Ishi's wood duck story,
p. 1 (Yahi language; Bancroft Library CU-23.1 #35.1)

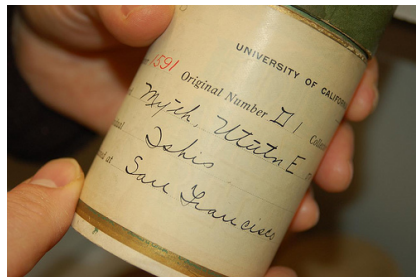# Phoebe A. Hearst Museum of Anthropology

Collection includes sound recordings created beginning in 1902, mainly before 1940:

- ▶ c. 400 hours of recordings from c. 65 languages
- ▶ Organization:
    - ▶ Wax cylinders: 2713 cylinders = c. 2200 texts in 113 collections
    - ▶ Wire recordings and tapes: c. 160 recordings
- ▶ Status: most content is digitized (no public access)
- ▶ Catalog: California Language Archive (**cla.berkeley.edu**)

Phoebe A. Hearst Museum of Anthropology

Representative collection: E. W. Gifford's 1926 Mono recordings

- ▶ Gifford (1887-1959) recorded three Mono speakers in August 1926
- ▶ Contents (on 59 wax cylinders)
  - ▶ 7 songs
  - ▶ 5 narratives
  - ▶ 1 'explanation' (of a dance)
- ▶ Duration: 1 hour, 22 minutes

A Hearst Museum wax cylinder in its container
(but Yahi language, not Mono)

Berkeley Language Center &
Survey of California and Other Indian Languages

Contents

- ▶ Paper records and sound recordings
    - ▶ Languages too numerous to specify ...
    - ▶ But mainly focusing on central and northern California
- ▶ Mostly created since about 1950
- ▶ Mostly associated with the Linguistics Department (student and faculty documentation) at UC Berkeley
- ▶ Original documentation content mostly digitized & online
- ▶ Catalog: California Language Archive (**cla.berkeley.edu**)

The most exciting development: collaborations on archiving with Native people donating their own recordings — linguist-free!

Survey of California and Other Indian Languages

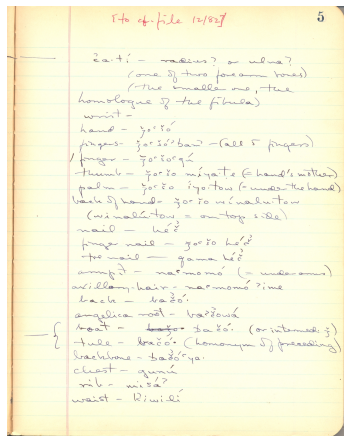An archive of paper documents deposited since about 1950:

- ▶ c. 270 linear feet of material from 296 languages
- ▶ Contents: field notebooks, unpublished manuscripts, lexical fileslips, letters and notes, microfilms, etc.
- ▶ Organization: 186 collections, $> 2{,}450$ distinct catalog items
- ▶ Digitization status: currently c. 58,000 scanned images from 303 items in 48 collections

Survey of California and Other Indian Languages

Representative collection:
Abraham M. Halpern (1914–1985)

- ▶ Halpern (Ph.D. Chicago 1947) studied with A. L. Kroeber and Harry Hoijer, then had a career in international relations
- ▶ collection has 80 items, mainly concerning Pomoan languages
- ▶ includes 23 field notebooks dated 1939–1940 and 1983
- ▶ also lexical file slips, linguistic notes, unpublished manuscripts, correspondence, etc.

Pomo field notebook 18, p. 5: Kashaya Pomo

Berkeley Language Center

An archive of sound recordings deposited since about 1950:

▶ 1,730 hours of recordings from 136 languages (cataloged to date)

▶ Media: wire recordings, tape reels, cassette tapes, DATs, CDs, videotapes, DVDs, digital files

▶ Organization: 221 collections, 7842 audio segments

▶ Digitization status: all cataloged material is digitized (but there is a large backlog)

Berkeley Language Center

Representative collection: Sylvia M. Broadbent, Southern Sierra Miwok

- ▶ Broadbent (Ph.D. Berkeley 1960) recorded three languages between 1955 and 1957, including SSM

- ▶ Contents
  - ▶ 20 songs
  - ▶ 18 elicitation sessions
  - ▶ 14 narratives
  - ▶ 3 conversations

- ▶ Duration: 6 hours, 22 minutes

- ▶ Related material: field notebooks in the Survey of California and Other Indian Languages

Castro Johnson and Charlie Rohan recording Southern Sierra Miwok with Sylvia Broadbent (source: Broadbent, *The Southern Sierra Miwok language*, 1964, p. 214)

Archives around the US

- ▶ National Anthropological Archives, Smithsonian Institution
- ▶ American Philosophical Society (Philadelphia)
- ▶ Santa Barbara Museum of Natural History

National Anthropological Archives, Smithsonian Institution

National Anthropological Archives, Smithsonian Institution

Some major collections (for California)

- ▶ Papers of J. P. Harrington (Chumash, Karuk, many others)
- ▶ Sound recordings made by J. P. Harrington
- ▶ Bureau of American Ethnology: 19th century documentation
- ▶ Selected papers of C. Hart Merriam

Access:

- ▶ Website: **nmnh.si.edu/naa**
- ▶ Catalog: **siris-archives.si.edu**

American Philosophical Society (Philadelphia)

American Philosophical Society (Philadelphia)

Some major collections (for California)

- ▶ Papers of Harvey Pitkin (Patwin, Wintu, Yuki)
- ▶ Papers of Jaime de Angulo and L. S. Freeland (Achumawi, Atsugewi, Pomoan, Shasta)
- ▶ Papers of P. E. Goddard (Hupa and California Athabaskan)
- ▶ Materials from Edward Sapir, Mary Haas, and many others

Access:

- ▶ Website: **amphilsoc.org**
- ▶ Catalog: **amphilsoc.org/mole/search**

Santa Barbara Museum of Natural History

- ▶ Chumash linguistic materials
  - ▶ J. P. Harrington
  - ▶ Madison Beeler
  - ▶ Richard Applegate

- ▶ Website: **sbnature.org**

Tribal archives: The future of language archiving

Many community groups and tribes possess extensive collections:

- ▶ Recordings from organized classes
- ▶ Community and tribal documentation programs
- ▶ Donations from individuals and families

For some languages, these collections are much larger than any university-based collections.

Tribal archives: Examples and resources

Tribal archives

- ▶ Alutiiq Museum, Kodiak, Alaska (alutiiqmuseum.org)
- ▶ Cherokee Heritage Center (cherokeeheritage.org)
- ▶ Colorado River Indian Tribes Library/Archives (critlibrary.com)

National associations and museums

- ▶ National Museum of the American Indian (nmai.si.edu)
- ▶ Association of Tribal Archives, Libraries, and Museums (atalm.org)

But mostly these entities do **not** focus on language materials.

Tribal archives: Advantages and challenges

Advantages

- ▶ Community relations and trust
- ▶ Controls over access
- ▶ Locally defined culturally-sensitive protocols

Challenges

- ▶ Permanence: Will the physical and technical infrastructure still be present in 100–200 years?
- ▶ Technical requirements: Is there good physical preservation and curation of materials?

Communities will make different choices based on the options.

Native archival principles

- ▶ Protocols for Native American Archival Materials

  - ▶ 'Some materials may have been collected or later restricted by a donor in contravention of community rights and laws . . . In all of these cases the rights of a Native American community must take precedence.'
  - ▶ 'For Native American communities the public release of or access to specialized information or knowledge . . . can cause irreparable harm.'

- ▶ Galarrwuy Yunupingu, 1986 (via Aboriginal and Torres Strait Islander Library and Information Resources Network)

  - ▶ 'There is information that is restricted, that our children cannot learn about, there is information that is restricted even to adults, there is information that is of a secret or sacred nature, that many people have no knowledge of or access to.'

Access, ethics, and cultural property

A constant tension:

- ▶ Archivists like material to be **open** and **accessible** to anybody who is interested in using it.
- ▶ University archives are usually interested in **broad** access, for example reaching users outside the university.
- ▶ Depositors sometimes want material to be **restricted** in some way.

Specific problems are generally of two types . . .

Overly restricted materials

Disputes

- ► . . . between linguists and community members
- ► . . . among community or tribal members

Fears

- ► Linguists who are slowly working on dictionaries & grammars may fear that other linguists will make use of their work.
- ► Community members may fear that accessible materials about their culture or language will be used for profit or in culturally inappropriate ways.

Personal sensitivity concerns

- ► In one case (not in CA) a whole collection is embargoed until 2035 because parts of the collection include gossip and private personal information. The collection includes valuable cultural and language materials.

Overly accessible materials

Ideally, before any material is deposited in an archive, the depositor and other creators would study it to decide on any restrictions:

▶ Culturally inappropriate or sensitive material

  ▶ Owned content (private songs, etc.)
  ▶ Recordings that should not be heard by unauthorized people
  ▶ Sacred content

  Sometimes stakeholders will differ in their judgments.

▶ Personal material

  ▶ Gossip or slander
  ▶ Personal details (address, birthdate, SSN)
  ▶ Private aspects of life or family history

Careful preparation is time-consuming and was rarely done in the past, so it is possible to find material that should **not** be accessible.