Bringing archival materials together to expand the Karuk text corpus

Matching up early recordings & transcriptions

Records showed Phoebe Maddux had recorded short texts with J.P. Harrington on a wax cylinder, now broken.

Notes archived in J.P. Harrington Collection at Smithsonian National Museum of Natural History, Washington, DC

Audio archived at California Language Archive, Berkeley, CA

Current speakers help transcribe and translate archival recordings to make texts accessible.

Audio archived at Phoebe Hearst Museum, Berkeley, CA

The recording was also made on disc (not archived at the Smithsonian), and survived!

Translating old recordings

Audio at California Language Archive, Berkeley, CA

Audio archived at American Philosophical Society, Philadelphia, PA

Notes archived at Survey of California and Other Indian Languages, Berkeley, CA

Notes archived in J.P. Harrington Collection at Smithsonian National Museum of Natural History, Washington, DC

Translating old recordings

Current speakers help transcribe and translate archival recordings to make texts accessible.

Translating old recordings

Outcomes

Texts with integrated audio added to parsed corpus, providing valuable material for:
- phonology and morphology
- syntax and narrative structure
- intonation, etc...
- Karuk language learning and teaching

Matching up early recordings & transcriptions

Challenges

- Finding materials
- Identifying materials
- Accessing materials
- Transcription and translation

Solutions

- Reconnecting different media types housed at different archives
- Digitized metadata and content
- Integration with fieldwork
- Lexicon database

Including new recordings & genres

Adding different genres, speakers, and modern recordings broadens the corpus’s potential.

Acknowledgments

I am grateful to all the Karuk elders, teachers, and activists who have worked to preserve that language and who have been willing to share it. Especially thank Vina Smith. Many thanks to the archives at the American Philosophical Society, the Phoebe Hearst Museum, the Smithsonian Institution, and the Survey of California and Other Indian Languages, as well as to Andrew Garrett, Susan Gehr and Ronald Sprouse. And thanks to the Karuk Study Group at UC Berkeley for countless hours of text digitization and audio splitting. This research was supported by a NSF Doctoral Dissertation Research Improvement Grant (BCS-1349075).

Contact:

csandy@berkeley.edu
http://linguistics.berkeley.edu/~csandy/

Dictionary & texts:

http://linguistics.berkeley.edu/~karuk/

References


Contact:

csandy@berkeley.edu
http://linguistics.berkeley.edu/~csandy/

Dictionary & texts:

http://linguistics.berkeley.edu/~karuk/

Outcomes

Texts with integrated audio added to parsed corpus, providing valuable material for:
- phonology and morphology
- syntax and narrative structure
- intonation, etc...
- Karuk language learning and teaching

Matching up early recordings & transcriptions

Challenges

- Finding materials
- Identifying materials
- Accessing materials
- Transcription and translation

Solutions

- Reconnecting different media types housed at different archives
- Digitized metadata and content
- Integration with fieldwork
- Lexicon database

Including new recordings & genres

Adding different genres, speakers, and modern recordings broadens the corpus’s potential.

Acknowledgments

I am grateful to all the Karuk elders, teachers, and activists who have worked to preserve that language and who have been willing to share it. Especially thank Vina Smith. Many thanks to the archives at the American Philosophical Society, the Phoebe Hearst Museum, the Smithsonian Institution, and the Survey of California and Other Indian Languages, as well as to Andrew Garrett, Susan Gehr and Ronald Sprouse. And thanks to the Karuk Study Group at UC Berkeley for countless hours of text digitization and audio splitting. This research was supported by a NSF Doctoral Dissertation Research Improvement Grant (BCS-1349075).

Contact:

csandy@berkeley.edu
http://linguistics.berkeley.edu/~csandy/

Dictionary & texts:

http://linguistics.berkeley.edu/~karuk/

References


