Oral Histories: Linguistic Documentation as Social Media

Mark Liberman

Linguistic Data Consortium. University of Pennsylvania

3600 Market Street, Suite 810, Philadelphia, PA. 19104 USA

myl AT ldc.upenn.edu

Abstract

Oral history recordings were pioneered by anthropologists in the early 20th century, collected by Alan Lomax and by the Federal Writers' Project during the 1930s and 1940s, and popularized by authors like Oscar Lewis and Studs Terkel in the 1950s and 1960s. Inexpensive tape recorders allowed the form to spread in the 1960s and 1970s. Now a new opportunity is provided by the combination of ubiquitous multimedia-capable digital devices, inexpensive mass storage, and universally accessible networking. The potential popularity of oral history-like recordings is demonstrated by the tens of thousands of people who have made recordings for StoryCorps. However, there is still no easy way for a motivated group – a family, an athletic group, a school class, a business, a scholarly discipline a club, a church -- to create and publish a collection of oral histories or similar forms of cultural documentation. But the software required to make and edit such recordings, and to transcribe, index, document, publish, and comment on them, is relatively simple and easy to create. And with the right infrastructure, millions of people around the world would participate, creating linguistic and cultural documentation on an unprecedented scale.

Keywords: language resources, collection, annotation, transcription, distribution.

1. Introduction

Oral history recordings were pioneered by anthropologists not long after the first primitive recording devices became available. John Lomax taught the Federal Writers' Project to use oral-history interview techniques in collecting thousands of first-person life stories of ordinary people during the Great Depression. His son Alan Lomax collected oral history interviews along with his folk-music recordings in the 1930s and 1940s. Books and radio programs by authors like Oscar Lewis and Studs Terkel popularized the form in the 1950s and 1960s. Inexpensive tape recorders allowed the form to spread in the 1960s and 1970s, and enabled sociolinguists and dialectologists to record tens of thousands of hours of interviews, though there was still no easy way to reproduce and share the results.

Now a new opportunity is provided by the combination of ubiquitous multimedia-capable digital devices, inexpensive mass storage, and universally accessible networking. Anyone with a smartphone can make good-quality recordings and upload them to the cloud. And the potential popularity of such recordings is demonstrated by the fact that tens of thousands of people have made (alas mostly inaccessible) recordings for the StoryCorps project, and millions of people have listened to selected and edited samples on NPR.

However, there is still no easy way for a motivated group – a family, an athletic group, a school class, a business, a union, a political campaign, a scholarly discipline, a club, a church -- to create and publish a collection of oral histories or similar forms of cultural documentation. But the software required to make and edit such recordings, and to transcribe, index, document, publish, and comment on them, is relatively simple and (where not already existing) will be easy to create. And with the right infrastructure, millions of people around the world would participate, creating linguistic and cultural documentation on an unprecedented scale.

1.1. Existing archives

There are hundreds of thousands of hours of existing untranscribed recordings. Many of these have been preserved in digital form, and others are being digitized – the question of how to salvage and preserve this material is an issue for another time. Tens of thousands of hours of sociolinguistic and dialect-survey recordings are now available in digital form, or soon will be.

For nearly all of these recordings, there's a natural constituency of interested people – ordinary citizens as well as linguists, historians or sociologists – who could be enlisted to help with transcription and annotation, if there were an efficient way for them to participate in distributed web-based projects.

In addition, there are many formal oral history projects that have been completed over the years, but are in many cases not easily available due to lack of any easy way to publish them, or because of outdated and inappropriate attitudes towards intellectual property or "human subjects" restrictions.

1.2. New collections

Recently, a colleague was initially interested in gathering oral texts for a lexicographic project in a widely-spoken but under-documented language. I showed her how to use a cell phone or tablet to make recordings. The first person that she interviewed was someone who had participated in an important series of historical events half a century ago.

At the end of an enthusiastic hour-long interview, the subject said "But I have so many more stories to tell, and my friends have so many more!" The interviewer found that first hour of stories so interesting that she decided to spend the summer making similar interviews in

her home country, and to look for ways to publish the results.

This is a common reaction. Most people have interesting stories to tell, if you give them a chance; and given the opportunity, most people like listening to other people's stories. And there are many social groups and events that motivate thematic collections of stories: intellectual or political movements, athletic teams, neighborhoods, school or church groups, professional gatherings, family reunions, and on and on. Whatever the topic or the occasion, there are millions of individuals and thousands of groups who would create accessible archives of oral texts if it were easier for them to do it.

A lot of this is already happening on tumblr, YouTube, Facebook and so on. But that material is generally shorter, less permanent, and/or less organized than the sort of thing envisioned here.

2. What we need

Some of the needed components already exist. Would-be collectors can easily find

- ways to record and edit audio and video
- ways to share audio, video and text online
- ways to organize online discussions

Some things that are not as easily available are

- good tools for transcription and alignment
- good ways to control access
- models for informed consent and authorship attribution
- methods for anonymization where needed
- inspirational model projects
- a flexible environment for group projects

Transcriptions are important partly because some people prefer to read, but mostly because they make searching and skimming possible. In particular, we badly need

- a well-designed transcription tool that works in web browsers and can read and write distributed audio, video and text
- easy-to-use systems for creating multi-media albums with flexible control over reading, writing, and commenting
- programs for aligning text and audio, and for presenting the results in a way that facilitates multimedia searching, browsing, and skimming.

2.1. Shared goals

There are a large number of other enterprises that share needs with the activities under discussion here. For example, an efficient, flexible, and easy-to-learn web-based transcription system is badly needed for many purposes. And an easy-to-configure system for managing distributed transcription and annotation – assigning tasks,

keeping track of progress, etc. – would be widely used if it were available.

3. How to get there

The Oral History Association has a web site[1] on "Oral History in the Digital Age", which promises "the latest information on best practices in collecting, curating, and disseminating oral histories." Content of this type is important as background to the discussion. But the OHA site, with its nearly 100 subsidiary pages on topics like "Metadata: Best Practices for Oral History Access and Preservation," "Transcripts, Time Coding, and You," and "File Naming In the Digital Age," is clearly aimed at a professional audience. The point of this presentation is that with the right environments and the right models, the "collection, curation and dissemination" of such recordings can become a mass-market enterprise.

Many groups and individuals around the world are working on various aspects of this problem. But along with some duplication of effort, there are pieces of the puzzle that no one seems to be working on.

I hope that this workshop will begin a discussion of ways and means, which can continue as an online discussion that informs participants about ideas, techniques and tools, and helps to enlist others in the process.

1 http://www.oralhistory.org/ohda-essays/