Ethical Concerns in Documentary Linguistics

With Special Attention to Language Endangerment

K. David Harrison
Swarthmore College
Living Tongues Institute for Endangered Languages

Sitting in a remote Siberian village listening to Marta Kangarayeva, one of the last fluent speakers of the Tofa language, I wondered where her story would go from here. A lifelong hunter, reindeer herder, and enthusiastic linguistic consultant, Marta devoted hours to sharing her knowledge with me. As she regaled me with stories of reindeer herding, I recorded her speech in digital video and audio.

I later played Marta’s audio tracks to other Tofa speakers to solicit their commentary and help with transcribing the story. Marta does not read but owns a cassette player, so I sent her audio copies of her recordings. I knew I would later be showing her videos to students in my linguistic anthropology seminar. Sentences from Marta’s story, with digital soundfiles, would be included in a published grammar of Tofa. Her story, with my corpus of digital Tofa materials, is now archived in the Netherlands, for future researchers to access, long after Marta and I are gone, and Tofa ceases to be spoken.

But what ethical and legal rules constrain this complex web, extended in space and time, surrounding Marta’s story? As languages vanish and linguists hasten to document them, new paradigms and standards emerge for collaborative research, archiving, data sharing and safeguarding of intellectual property. These may supersede earlier models both in their technical dimensions and in their human and ethical ones.

See Ethical Concerns on page 32
Providing Identifiers

Language documentation requires the recording of people's voices and images in digital media. Anonymity becomes moot, as it would in fact be unethical (unless the speaker requested to anonymize data). In writing grammars of endangered languages, emerging best practice requires that every single datum be identified with the initials (if not the name) of the speaker, often accompanied by a CD audio recording. A recent example is Nick Thieberger's A Grammar of South Etape: An Oceanic Language of Vanuatu (2006). So-called "thick" metadata—maximally identifying when, where and who produced the data—also allows linguists to observe the kinds of change and hyper-variation found in small or demographically contracting speech communities, which pose theoretical challenges to current models of language change and obsolescence.

Informed Consent

Informed consent can be a delicate negotiation. Asking speakers to sign consent forms is often unreasonable—especially if they cannot read—or culturally inappropriate. Linguists and anthropologists typically employ recorded, on-camera discussions in which the purpose and use of the data is discussed (this discussion itself may take place in an endangered language) and the participant agrees to be recorded. Explaining possible ends-uses is a tricky business. I recall a discussion with nomadic Monchak consultants in Western Mongolia in which I attempted to explain how their data would be archived and accessed remotely over the "Internet." To describe the web, I referenced technologies they already knew about (library, telegraph, radio), and I trusted that their consent based on my explanation was sufficient.

Data Repatriation

Even more important to the Monchak community than data dissemination, I learned, was data repatriation, an emerging ethical standard in documentary linguistics. In 2004 I visited the Monchak for a third time, carrying on my laptop digital recordings from previous visits. One community elder had died, I learned, and I hesitated to show his recordings, for fear of upsetting the family or transgressing some cultural norm. But the family themselves approached me saying, "Don't you have pictures of our father from last time?" I can only imagine my chagrin if I had not brought the recordings along. The family watched, transfixed and in tears, as I played the video. As soon as it was finished they summoned neighbors and requested repeated showings.

Though I had no way to leave copies of the data with the Monchak, I felt that they deserved to freely view the videos during my visit, and that in what I was prepared to provide, even if it meant running down all my batteries and foregoing more recordings. Language data must be repatriated whenever possible and, in whatever format the community is able to receive it, whether a paper transcript, a public screening, a videotape or a CD. This is yet another area of ethical practice in which standards and expectations seem amicably to be higher for endangered language data than for other data.

Questions of Preservation and Ownership

The liberal role as curator of primary ethnographic data—the intellectual property of others—is a precarious one. This has intensified as linguistics moves from an "anomalous" model in which informants were asked to produce long lists of predetermined words or sentences (lacking any cultural impotence and cultural traits), to a more organic, holistic model in which consultants are invited to tell their stories. The resulting narratives are not merely material to be mined for linguistic data, but are cultural productions, wholly owned by their originators (both as individuals, and in some cases communities) and entrusted to linguists for analysis and preservation.

Working on languages with minimal and shrinking speaker bases, we find that attitudes towards language ownership can differ from those of larger speech communities. A community like the Tofa (fewer than 30 speakers in 2002) can feel ownership over their language—one in which every single speaker personally knows every other speaker—that is not at all comparable to how France or the Francophone world "owns" French. Heightened feelings of ownership by speakers problematize the issues noted above, and have even led to lawsuits. (In 2006 the Mapuche of Chile sued Microsoft, alleging it had translated Windows software into Mapudungun without tribal leaders' permission.) Communal ownership of intellectual property by non-incorporated groups is not a formal category under international or national copyright regimes. Recent policy initiatives such as UNESCO's recognition of intangible cultural heritage provide only discretionary or advisory protection, not equivalent to formal legal regimes of patent and copyright. Language data viewed by a community as its intangible property requires extra care, for example, archiving with graded access for different users, so that culturally or personally sensitive materials are protected. This places a greater burden on linguists to collect metadata noting speakers' wishes for granting or limiting access in space or time.

There is a consensus among documentary linguists that we are still far from addressing the ethical dilemmas posed by digital media, data longevity and collaborative models. We confront the globalization of our data, for example, as narratives told by last speakers of Tofa in Siberia are recorded by me, digitized and archived in the Netherlands, annotated in the US, funded by a foundation in Germany, potentially accessible worldwide on the Internet (but not in the originating community, who lacks such technology), and eventually published in print/audio format in the US or Europe. The meaning of what the documentary linguist or ethnographer collects in a particular speech community and cultural context may shift when it is transferred to and through these other communities and media. The question of ownership over endangered languages becomes highly fraught, since copyright as legally conceptualized has to do with the expression of ideas and knowledge, not necessarily the ownership of the idea and knowledge itself, no matter how related. Without the legal means to control the fate and future of their language—often one of the last vestiges of their ancestral culture left to them—small communities risk losing access to their own culture and heritage.

The question of who owns endangered language data, and how we are to ethically and responsibly curate it, remains very pressing indeed. Linguists will continue to look to anthropology for guidance. Linguists may also have something to share with anthropologists from our own experience, and from guidance we may get from last speakers of languages, who are, after all, uniquely invested in the fate of their own utterances.

David Harrison studies endangered languages in Siberia, Mongolia and India. He is the author of When Languages Die: The Extinction of the World's Languages and the Erosion of Human Knowledge (2007).