

Documentary linguistics and ethical issues

Nick Thieberger and Simon Musgrave

1. Overview

In recent years, there has been an increasing emphasis on documentary linguistics within our discipline. This change of emphasis has been motivated by our concern over the pace of language loss, and has been facilitated by coincidental technological changes. Within this developing field, and especially as a result of the technological resources now available, we suggest that new ethical challenges arise in the professional practice of the linguist. The issues which we wish to raise in this paper stand outside of the area covered by existing institutional ethics procedures.

The practice of documentary linguistics has a greater impact in a community than traditional data collection practice. There are two aspects to this impact. Firstly, a good documentation attempts to record as wide a range of language events as possible, in many genres and in many settings. This implies that the researcher's presence in the community will be more intrusive than if the sole aim is to record sufficient material to prepare a grammatical description. Secondly, the nature of the data captured is also more intrusive, with video recording common and high quality audio recording more or less standard. Language documentation also implies the existence of archival data, that is, high quality data which is intended for persistent storage, which is accompanied by metadata sufficient to allow for the discovery of the resource, and which is under the control of a third party.

Both of these aspects of documentation raise ethical issues. What procedures are appropriate to obtain informed consent to the type of data collection discussed above? What sort of rights and responsibilities does an archive have as another interested party in the negotiation of agreements between researchers and speakers / communities? Given the technological possibilities for dissemination and reproduction, how can ownership rights in recorded material be handled? How far should communities' concepts of ownership be taken into account? How can ownership and access rights be negotiated so that they hold over the time frame which archiving assumes? What may be the consequences for a community when material is returned to them by researchers or archivists, given that the research and archiving

process will inevitably have changed the nature of the material and its status in the community?

We suggest that it is time for linguists to engage with these issues. We will discuss who the interested parties are in these processes, what responsibilities and rights each party may have, and some of the areas of potential conflict between those rights and responsibilities.

2. Introduction

In this paper, we suggest that changes in the field of linguistics over the last twenty years have raised new ethical challenges in our professional practice. We discuss three specific areas in which there has been change: the rise of documentary linguistics, technological developments, and the new importance of archiving data. We suggest that these changes taken together represent a paradigm change in linguistics. Within this new paradigm, we identify some of the new ethical problems which arise, and the new complications which have been added to previously identified ethical issues. Our discussion focuses on the problem of obtaining informed consent to the collection of language data, especially the extent to which archiving of data is legitimated by existing consent agreements; on the complex problems of ownership and rights over language data and the various knowledge products derived from it; and the problems associated with a commitment to making data available to speaker communities and their descendants. Our aim is not to provide answers to these problems, but rather to contribute to the discipline's ongoing engagement with ethical issues.

3. Paradigm Change

3.1 Documentation

Himmelman (1998) sets out desiderata for the documentation of a language. He suggests that this should include, as far as possible, recordings of a wide range of different types of linguistic behaviour, from very formal situations to rather informal situations. Linguists should attempt to document as many different uses of language in the community as they can. Himmelman also advocates the use of video recording as the most complete possible record of language in use. All of these recommendations lead to an increase in the intrusiveness of the linguist's practice, and indeed Himmelman is pessimistic about the likelihood that linguists will regularly obtain permission to make documentation of this sort.

Himmelmann also sets out a strong position with regard to the ethical obligation which should govern the actions of linguistic researchers in their relations with speaker communities: "I presume without further discussion that the interests and rights of contributors and the speech community should take precedence over scientific interests." (1998:172) We take it as given that we have "no special entitlement to study all phenomena; and the advancement of knowledge and the pursuit of information are not in themselves sufficient justifications for overriding the values and ignoring the interests of research participants." (Australian Anthropological Society's ethics guidelines, 3.6.a¹).

It could be considered that documentary linguistics is a new name for what were standard operating practices prior to the dominance of theoretical linguistics in the 1960s. The anthropological linguistic tradition continues to emphasise the importance of recording contextual information and so we can look to ethical guidelines established by anthropologists for guidance. However, while we acknowledge the validity of this position, we also consider that linguistics is undergoing significant changes (a paradigm shift) as a result of changes in technology combined with a reappraisal of the importance of documentation of small languages, especially those with little prospect of being spoken in future. While the practice of recording as much as possible from a variety of speakers and in a variety of contexts may not be new, it is the access to recorded material and its presentation that provides wholly new ways of conducting our analysis. The ability to access arbitrary points within digital recordings allows us, as researchers, to firmly base our analysis in the corpus and to present our results and hypotheses together with the data on which they are based. The data is prepared for reusability by other researchers by being richly annotated, and by having persistent identification and well-formed descriptions that allow it to be located. The need for data prepared in this way is reinforced by exigencies of long-term archiving and accessibility for speakers and their descendants. It is also mandated by large funding agencies (e.g. Hans Rausing Endangered Languages Documentation Programme and the VW Foundation's DoBeS project) who recognise that access to the results of funded research need to be locatable and reusable. All of this is included under the broad rubric of language documentation and illustrates the possibility of a fundamentally revised linguistic practice, which brings with it novel ethical dilemmas.

In this paper we will not address the ethical issues already known from current fieldwork practice, for which see, for example, Rice (2005, to appear), who discusses ethical frameworks for fieldwork and contrasts earlier approaches with advocacy and empowerment frameworks in which the

linguist is actively engaged with the community on their terms (see also Cameron et al 1993 and Craig 1993).

However, we note that for some issues the new paradigm may give a greater intensity to the issues previously faced. For example, fieldwork for language documentation may last longer than that for language description, with a greater possibility of sexual relationships in the field, cf. Newman's (1992:4) observation that, while gender may be discussed in the literature (e.g. women's role as professionals), sex is not. The issue of sexual relationships in the field should be discussed, especially when there is a great power differential in the relationship between the fieldworker and the members of the language community. Newman lists several scenarios involving fieldwork relationships, and fieldworkers (especially students) should be aware of such possibilities in advance of fieldwork. The type of issues raised by Newman include: (1) the question of what kind of sexual behaviour is expected/ agreed upon by a married researcher and their partner during an extended period of fieldwork; (2) the question of the risks faced by a homosexual fieldworker working in a location where homosexuality is not as commonly accepted as it is in the first world; and (3) the question of the contrast between the perceived desirability of a single white person in a poor community, and the way they are seen in their home community. Such issues are also discussed by Kulick and Willson (1995).

3.2 Technology

Technological changes have had an enormous impact in all areas of academic research over the last twenty years, and linguistics has also been affected. The possibilities for data collection have improved significantly in this period. Twenty years ago, audio recording was limited to analogue technology, in the form of either bulky reel-to-reel equipment, or more portable but less accurate cassette recorders. Today, high-quality digital recordings can be made easily with relatively cheap equipment. Video recording is not only desirable, but is now also attainable, again with comparatively cheap equipment.

These digital technologies mean that recorded data is now easily manipulated using the same computers which linguists use for many other tasks. Data can be reproduced with no loss of fidelity (or minimal loss), and the rapid development of the World Wide Web means that it can also be disseminated very widely and very easily. As Anthony Woodbury has said:

With powerful laptops, digital audio and video and the WWW it at least seems that we should be able to capture and store enormous amounts of information.... We should be able to disseminate around the globe the material now collecting dust in attics or

¹ http://www.aas.asn.au/docs/AAS_Code_of_Ethics.rtf (viewed August 28, 2006).

rotting in basements; and we should be able to keep huge amounts of information safe in perpetuity. (Woodbury 2003:36)

These developments must be welcomed, for the new standards of accountability which they are bringing to our discipline, but they also raise problems.

3.3 Archiving

Documentary linguistics implies the existence of linguistic archives. If we have tried to locate information about small languages in archives ourselves then we know how valuable a well-formed set of linguistic data can be for analysis and for language reintroduction. Representation of a language in as much detail as is possible for future generations to access is a goal of language documentation, and, as Johnson has observed:

there is little sense in collecting data on languages that are disappearing if there is no plan for preserving that data ... [and] also [to] support the maintenance and revitalisation of endangered languages by making materials from earlier periods ... available to the speakers and their descendants (Johnson 2004:140)

Even minimal description of data is preferable to none, as "even in imperfect form, ordered, shared data are more useful than no data." (Dwyer 2006:40) Archiving in turn imposes various requirements on the data collector. Data must be available in standard formats which stand some chance of being accessible in the future; they must be accompanied by metadata which is rich enough to make discovering the data possible; and provision must be made, where necessary, to control access to the data over the period of storage.

4. Ethical Challenges

4.1 Informed Consent

Obtaining informed consent from people who are the subjects of research is a central element in the research ethics procedures of institutions. This issue is therefore atypical amongst those which we discuss in this paper, in that institutions such as universities and funding bodies are party to the negotiations. Funding will not be released and research cannot commence before an ethics committee is satisfied that appropriate procedures have been set up to obtain informed consent. We suggest that two aspects of the negotiation are now problematic as a result of the changes discussed above. Firstly, the type of data collection which is now at issue is more intrusive,

both in terms of the range of contexts in which data may be collected and in terms of the technological possibilities for capturing data. Secondly, there are limits to the ability of the data collector to foresee what sort of uses the data may be put to in the future. The increase in the possibilities for exploiting language data which we have witnessed over the last decades cannot be expected to halt now or in the future, and there is little chance of any of us making accurate predictions about the direction of technological development in the future.

The nature of the information which should be available in the negotiation of consent is therefore problematic for both the speakers and for the data collectors. We wonder how the data collector can fully inform the speakers about the nature of the activities to be undertaken, but trust that this is part of the researcher's role, since "mediating each other's cultural imperatives" (Dwyer 2006:32) is integral to the task of language documentation.

Procedures for obtaining consent may satisfy the requirements of the linguist's home institution, but clearly the social relationship which will inevitably exist between the parties impose a further ethical obligation. Speakers may rely on the social relationship, that is, on trust, in their negotiating strategy, in preference to written agreements, which raises the issue of how specific consent forms should be. It may be enough to have a general consent to data collection which the collector can interpret as covering any activity, or there may be activities which require separately negotiated specific permissions, for example recordings of emotional interactions.

What uses of material should consent be taken to cover and when is it necessary to renegotiate? There is a continuum of access and potential use for data, from restricted access archiving, through use in academic analysis, public access (justified perhaps by public funding of the collection and archiving processes), ending with the possible commercial use of data and the knowledge stored in it. What factors influence our judgments about balancing speaker privacy against access to data, and what can initial consent be taken to cover?

There are also difficult questions about the relationship between individual speakers, communities and researchers. Consent can be negotiated with individuals to cover data collected with them, or with representatives of a community (for example a council) to cover all data collection within that community. Himmelmann's precept quoted above in section 1.1 can be taken to imply that speakers should have a right to veto the use of recorded material after the fact. But should a community have the right to veto the use of material recorded with one of its members when that person has given explicit permission for the material to be used?

In the background of all of these questions there is the fact that the data collector cannot know precisely what the data might be used for in the future, despite providing deposit forms for archiving based on consent forms agreed to in the field. To what extent does this consideration make any negotiation of consent illegitimate?

4.2 Archiving and consent

Where data is to be archived, the archive can be considered as a separate party in the negotiation of consents and permissions (see Holton's 2005 discussion of archiving ethics). This raises the question of whether it is sufficient for data collectors to negotiate the consent which they regard as sufficient, or whether specific permissions with regard to archiving also need to be negotiated. Clearly, in the case of heritage data, this issue does not arise, but in the case of data collected now and in the future, the problem must be faced, and must be faced by both data collectors and by archives. The data collector has to inform speakers about the archiving process as fully as possible, but up until the present it has been unusual for the linguist to arrange long-term agreements covering rights and access with the speakers. Presumably it is never going to be feasible for archives to obtain consent on their own behalf. It is up to the archives to make judgments as to the value of consent statements which accompany data lodged with them and how to balance the obvious benefit of securely archiving data, and the less obvious disadvantages of having data whose access status is unclear.

The element of time which is integral to the archiving enterprise brings further problems. It may become necessary or appropriate to review the status of archived material, especially material which has access restrictions imposed on it, but in a situation where the data collector and the speaker are unknown, uncontactable or dead, who can an archive negotiate with? How far do their obligations extend in respect of such material, and should the default assumption be that restrictions continue indefinitely unless their lifting is specified by some party with the right to make that decision (as might be inferred from Himmelmann's position)?

4.3 Ownership and Rights

Many different sorts of material are created in the process of language research. To list a few:

- a recording of the event
- annotation of the recorded event
- non-academic publication (e.g. story books)
- academic publication (e.g. grammar or dictionary)
- archival objects

It seems clear that different sorts of intellectual property are found in these different sorts of material, and it would be difficult to maintain that ownership rights are uniform across these various types of object. Who then can claim rights in each of these various items? The input of the linguist increases as one proceeds down the list, but this intuitive grasp of the situation is not necessarily mirrored in the legal situation. For example, a recording of an event is protected by copyright, and, under Australian copyright law, that legal right is owned by the creator of the object, the person who made the recording, perhaps with some rights assigned to the performer. We need therefore to examine carefully the nature of the intellectual property included in all these objects (and in others, no doubt), and to consider who can lay claim to that property, and what protection they might seek. In some cases, such as copyright, legally enforceable rights exist; in other cases, moral rights exist and are recognised in law in some jurisdictions (e.g. Australia, UK), but not in others.

Should we recognise the existence of ethical obligations protecting intellectual property, beyond these legal protections? How should we respond where institutions wish to claim rights to intellectual property created in their scope, not only for ourselves, but also for those who provide the original data from which new knowledge is derived? Another important question must also be faced: the discussion so far has assumed a Western, legalistic concept of property rights, but this concept is not shared by many other societies. Should we attempt to take indigenous concepts of ownership into account in our thinking on these issues, and if so, how might that be accomplished?

A further complication arises through collaborative research, itself a welcome step and one at the core of the language documentation paradigm. When we work with researchers from other disciplines or even collaboratively with other linguists (perhaps discourse specialists, phoneticians, sociolinguists, typologists or other theoreticians) we will need to pay more careful attention to the rights inherent in the data we have created. There is

some potential in the Creative Commons² approach to declaration of intellectual property rights that allows creators to specify what rights are reserved and what the user can do with the data, as opposed to copyright which claims all rights for the creator with little leeway for the user³.

4.4 Return of Materials

We take it as a given that speakers and their descendants must be able to locate and access language data. As indicated in the quotation given in section 1.3, this possibility is a major motivation for ensuring secure archiving of data. However, this practice also raises problems calling for ethical judgments on the part of linguists and archivists. Once a linguistic event, say a narrative, is recorded, it is fixed in time and this may be contrary to the normal practice of oral cultures. When language fixed in this way is re-introduced to a community, it may have unexpected and unfortunate consequences. For example, what might have seemed an innocuous narrative at the time of recording may be interpreted by descendants of the original speaker as important evidence in a dispute within a community over rights to land or knowledge. It is unclear if data collectors therefore have a responsibility to try to flag the possibility of such interpretation at the time of storage (if they are, in fact, aware of them). Beyond such specific instances, there is the more general concern that time and the use of technology may reify and lend authority to what were originally ephemeral acts. This may have consequences at a purely linguistic level: the choice of what data is collected may over time lend prestige to one dialect or variety in a speech community. Data collectors can not reasonably be expected to be responsible for such effects, but these are issues we will face in repatriation of recordings to home communities. Thus archival preservation, while an important and perhaps crucial aspect of language documentation, is not a politically neutral activity. The researcher will need to mitigate the possible damage caused, but on balance it is better that archival objects exist so that they can eventually be returned to source communities.

² <http://creativecommons.org>

³ As many linguists work in what is termed the 'global South' it is also worth noting the damaging consequences of the first world copyright regime, especially as it is often enforced by so called free trade agreements. Copyright needs to be recognised as a means of control for corporations and is usually not well suited to the intellectual property rights of speakers or linguists (cf. Story, Darch and Halbert 2006).

5. Where to from here?

In the preceding section, we have raised many questions and offered hardly any answers. This should not be surprising at the current stage of investigation, and indeed, we would consider it unhealthy if we (or any other party) were presenting a package of solutions to these very complex problems. We are convinced that in most cases the problems will be and must be resolved by individual decisions guided by a statement of ethics from a professional body, and we would not support a statement or code which went further in attempting to regulate individual practice (cf. Punch 1986:78-84). But we are also convinced that statements of ethics agreed to by a professional association provide a framework within which the association's members can know what is expected by their colleagues – something that is especially important for new researchers entering the field. And finally, we would suggest that the willingness of linguists, as individuals and as a group, to engage with ethical issues in our professional practice is an important part of our response to the technological and methodological changes which are transforming our discipline.

References

- Cameron, Deborah, Elizabeth Frazer, Penelope Harvey, Ben Rampton and Kay Richardson 1993. Ethics, Advocacy and empowerment: issues of method in researching language. *Language and Communication* 13: 81-94.
- Craig, Colette 1993. Fieldwork on endangered languages: a forward look at ethical issues. In André Cochetiere, Jean-Claude Boulanger, and Conrad Ouelton (eds.) *Proceedings of the XVth International Congress of Linguists*, Vol. 1: 33-42. Quebec: Presses de l'Université de Laval.
- Dwyer, Arienne 2006. Ethics and practicalities of cooperative fieldwork and analysis. In Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel (eds.) 2006. *Essentials of Language Documentation (Trends in Linguistics. Studies and Monographs, 178)*, 31-66. Berlin: Mouton De Gruyter.
- Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161-195.
- Holton, Gary. 2005. Ethical practices in language documentation and archiving languages, available at <http://www.language-archives.org/events/olac05/olac-lsa05-holton.pdf>.
- Johnson, Heidi 2004. Language documentation and archiving, or how to build a better corpus. In Peter Austin (ed.) *Language Documentation and Description*, Vol. 2, 140-153. London: School of Oriental and African Studies.
- Kulick, Don and Margaret Willson 1995. *Taboo : sex, identity, and erotic subjectivity in anthropological fieldwork*. London and New York: Routledge.

- Newman, Paul 1992. Fieldwork and field methods in linguistics. *California Linguistic Notes*, 23:3-8.
- Punch, Maurice 1986. *The Politics and Ethics of Fieldwork: Muddy hands and grubby boots*. Sage University Paper series on Qualitative Research Methods, Vol.3. Beverley Hills CA: Sage Publication
- Rice, Keren 2005. The Linguist's Responsibilities to the Community of Speakers. Conference on Language Documentation: Theory, Practice, and Values. LSA Linguistic Institute, 9-10 July 2005.
- Rice, Keren to appear. Ethical issues in linguistic fieldwork: an overview. To appear in W. van den Hoonaard (ed.) *The ethics trapeze*. Vancouver: University of British Columbia Press.
- Story, Alan, Colin Darch and Debora Halbert 2006. *The copy/South dossier*. Canterbury: University of Kent.
- Woodbury, Anthony C. 2003. Defining documentary linguistics. In Peter Austin (ed.) *Language Documentation and Description, Vol 1*, 35-51. London: School of Oriental and African Studies.

Appendix – Some resources on ethics and fieldwork

Ethics statements and codes of practice from professional bodies:

American Anthropological Association
<http://www.aaanet.org/committees/ethics/ethcode.htm>

Australian Anthropological Society
http://www.aas.asn.au/docs/AAS_Code_of_Ethics.pdf

Australian Linguistic Society
<http://www.als.asn.au/> (Link via ALS Policies)

Ethics statements by funding bodies and institutions:

Australian Institute of Aboriginal and Torres Strait Islander Studies –
 Guidelines for Research in Indigenous Studies
http://www.aiatsis.gov.au/__data/assets/pdf_file/2290/ethics_guidelines.pdf

DoBeS Project
http://www.mpi.nl/DOBES/ethical_legal_aspects/

Department of Linguistics, Max Planck Institute for Evolutionary
 Anthropology <http://www.eva.mpg.de/lingua/files/ethics.html>

Other resources:

Ethical issues in linguistic fieldwork, Linguistics Department, University of
 Toronto
<http://www.chass.utoronto.ca/linguistics/LIN458H1F/04.458.ethics.pdf>

American Anthropological Association Ethics Homepage
<http://www.aaanet.org/committees/ethics/ethics.htm>

Interagency Advisory Panel on Research Ethics (Canada)
<http://www.pre.ethics.gc.ca/english/index.cfm>

Linguistic fieldwork at Stanford University – The Ethics of Fieldwork
<http://www.stanford.edu/dept/linguistics/fieldwork/info/ethics.html>