# Ethical challenges in documentary linguistics

Simon Musgrave
Linguistics Program
Monash University
Victoria  3800
Simon.Musgrave@arts.monash.edu.au
http://www.arts.monash.edu.au/ling/staff/musgrave/

Nick Thieberger
Department of Linguistics and Applied Linguistics
The University of Melbourne
Victoria  3010
thien@unimelb.edu.au
http://www.linguistics.unimelb.edu.au/thieberger/

*Abstract*

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.

*Keywords*    Ethics, Language Documentation, Archiving

# Ethical challenges in documentary linguistics

## Introduction

Australian linguistics, with its strength in fieldwork-based research has been notable for discussing and resolving issues that have not been considered elsewhere until more recently. For example, the establishment of regional language centres is a world first and an Australian initiative. AIATSIS provides a long-term secure repository for field recordings and associated material that has no homologue outside of Australia, and its electronic data archive in particular was the first such archive dealing with indigenous languages in the world.

Similarly, the Australian Linguistic Society adopted a statement of ethics (SOE) in 1989. This document is available from the ALS website (and is copied as Appendix A to this paper) and broadly sets out some fairly unexceptionable sentiments, such as the need to behave "in accordance with the highest professional standards" and "in broad conformity with the ethical provisions made by comparable professional bodies". Neither of these aims is specifed in any more detail which presumably allows the linguist to do some 'comparable-professional-body shopping' to determine, if they were so inclined, which suits them best.

The ALS was prompted in part by the distribution of the document *Professional Ethics for Linguists doing Research in Aboriginal communities* prepared by ANU linguistics students around 1984 (copied as Appendix B to this paper). This document was primarily concerned with the responsibilities of researchers working in Australia and was developed in cooperation with the Aboriginal Languages Association. It sets out more detailed suggestions for the responsibilities of fieldworkers in Australia and could be more widely discussed, especially among students of field methods classes or those embarking on fieldwork for the first time.

Unlike the ALS, its British counterpart, the Linguistics Association of Great Britain (LAGB), appears to have adopted no statement of ethics[1], and its US counterpart, the Linguistic Society of America (LSA) has no statement of ethical guidelines. In early 2005 at a meeting organised by an LSA working group on language documentation it was observed that the LSA needed an ethics statement, and the response from several linguists present was that not only was it unnecessary but dangerous to codify ethical

---

[1]  cf. email message to LAGB mailing list from Dan Everett, 4 Sep 2003 (http://www.essex.ac.uk/linguistics/LAGB/mail/2003/031007123217.html),"Does the LAGB have a code of ethics for field research? If there is such a code, does it include anything on ethics of research on endangered languages?" There was no reply on the list, and a search of the LAGB website finds no mention of a code of ethics.

guidelines. It seemed that the danger lay in providing a framework against which a linguist's behaviour could be measured. It is clear that the ALS has little ability to enforce an ethics statement, the norms established by it could only be for guidance of ALS members. This is despite the final sentence of the current statement that "Persons deemed to be conducting research not in accordance with the spirit of this ethical statement may be subject to disciplinary action by the Australian Linguistic Society, according to principles that may from time to time be determined by the Society. "

The Applied Linguistics Association of Australia (ALAA) adopted a 'Statement of Good Practice'[2] in 1997 which outlines a number of responsibilities of practitioners. The Australian Anthropological Society (AAS) has a lengthy Code of Ethics and lays out in the first paragraphs the parameters by which the Code should be understood: "The objectives of the Code are to provide the consensus view of ethical practice among members of the Australian Anthropological Society, to elaborate on the standards we share, to aid members in making informed decisions about their own practices, and to help them communicate their professional positions more clearly to other parties. The Australian Anthropological Society cannot investigate or adjudicate allegations about unethical behaviour."

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 four specific areas in which there has been change: the rise of documentary linguistics, technological developments, the new importance of archiving data, and the increased possibilities for linguists to work outside the academy. 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.

# 1. Paradigm Change

## 1.1 Documentation

Himmelmann (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

---

[2]   http://www.latrobe.edu.au/alaa/goodprac.htm (viewed on 17/1/06)

situations. Linguists should attempt to document as many different uses of language in the community as they can. Himmelmann 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 Himmelmann 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)

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 anthroplogical linguistic tradition continues to emphasise the importance of recording contextual information and so we can look to ethical guidelines established by anthropolgists, as discussed above. However, we do consider that linguistics will undergo 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 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. 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.

## 1.2 Technology

Technological changes have had an enormous impact in all areas of academic research over the last twenty years, and linguistics has been affected also. 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 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.

## 1.3 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)

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.

## 1.4 Beyond academia

It is now not uncommon for linguists in Australia to be employed by bodies other than universities or other educational institutions. For example, a number of linguists are employed by Language Centres under the control of indigenous groups, and others are employed as expert consultants in the preparation of Native Title cases. The professional practice of such linguists is not subject to the check imposed by the ethics processes of academic institutions (to the extent that such processes are relevant to the type of concerns we discuss here).

It is quite common for linguists working in this environment to undertake contractual obligations such as the following (taken from a contract offered by an agency in Central Australia):

3.1    It is the intention of the parties that this contract be consistent with, and not in conflict with, any professional ethics or rules of conduct which may apply to the Consultant.
3.2   In the event of any inconsistency arising between this contract and any professional ethics or rules of conduct applicable to the Consultant, those ethics or rules shall, to the extent of such inconsistency, prevail.

It is a legitimate question to ask whether the current SOE of the ALS is a suitable authority to be appealed to under such provisions.

## 2. Ethical Challenges

### 2.1 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. 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?

## 2.2 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 the data collector to negotiate the consent which she regards 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 Himelmann's position)?

## 2.3 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 Australiaan 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 recognized in law in Australia, but not in other jurisdictions.

Should we recognize 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?

## 2.4 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. 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.

## 3. 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 must be resolved by individual decisions guided by a statement of ethics from a professional body. We would not support a statement or code which went further (cf Punch 1986: 78-84). But we are also convinced that some re-examination of the ALS statement is appropriate at this time, in view of the changes to professional practice which we have discussed, and also changes in community expectations in relation to indigenous communities, an issue not discussed here. We suggest that the willingness of linguists as a group to embark on such a process, and the indication of a willingness to engage with ethical issues in our professional practice, is probably more important than any outcome of the process. That the ALS might be seen as a leader in such a process seems an opportunity to be welcomed.

## Bibliography

Himmelmann, Nikolaus P. 1998. Documentary and descriptive linguistics. *Linguistics* 36: 161-195.

Holton, Gary. 2005. Ethical practices in language documentation and archiving languages. 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.) Volume 2. London: Hans Rausing Endangered Languages Project. 140-153.

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 Publications

Woodbury, Anthony C. 2003. Defining documentary linguistics. In Peter Austin (ed.) Volume 1. London: Hans Rausing Endangered Languages Project. 35-51

## Appendix A – ALS Statement of Ethics

The Australian Linguistic Society, as the principal professional body concerned with the discipline of linguistics in Australia, declares that the following ethical principles apply to its members, and also to persons directly employed by members, in the conduct of linguistic research:

1. Researchers in the field of linguistics have multiple and cross-cutting obligations to their discipline, to their colleagues, to bodies and individuals providing funds and facilities for research, to human communities and individuals who participate in any way in the research, and to themselves. These obligations are to be met by the researcher in accordance with the highest professional standards and consideration of the highest prevailing standards of human ethics, in broad conformity with the ethical provisions made by comparable professional bodies, associations of social sciences and humanities, and committees on medical research and experimentation on human subjects.

2. In particular, the rights of lay persons and communities involved in the research require especial consideration, since such persons and communities may not have the same understandings about the nature and conduct of the research as the researcher. These rights include protection of privacy, the right to withhold cooperation in the research, appropriate remuneration, and access to data collected by the researcher, and to the results of the research. These rights are especially important in small and threatened communities. It is recognised that particular communities may have strong views on the researcher's access to data and dissemination of information; such views should be respected, to the extent that they do not clash with other obligations of the researcher.

3. Nothing in this statement of ethics shall override the conditions for research imposed by particular communities, official bodies, or governments; in agreeing to such conditions before undertaking research, the researcher is deemed to have entered into a contractual obligation.

4. Persons deemed to be conducting research not in accordance with the spirit of this ethical statement may be subject to disciplinary action by the Australian Linguistic Society, according to principles that may from time to time be determined by the Society.

(Adopted at the 1989 Annual General Meeting of the Australian Linguistic Society, published in The Australian Linguistic Society Newsletter no. 90/4, November 1990)

## Appendix B – Draft Statement Of Professional Ethics For Linguists Doing Research In Aboriginal Communities

It is probably impossible to produce a statement of ethics which will apply in a clear and unambiguous way to all possible situations. Applying concepts like 'community'; 'authority structure', and indeed 'se1f-determination', wi11 inevitab1y be problematical, even controversial, in many cases.

Also, there is a wide range of fieldwork methods and styles, from brief fee-for-service interviews to long-term participant-observation, and each will involve its own particular ethical problems. The intention of this statement is to provide the minimal ethical guidelines common to the whole range of research situations.

<u>Relations with communities</u>

(1)    The linguist must always make a point of ensuring that the community understands his or her obligations outside the field situation, the linguist's main motivation for being in the field, the aims of the research and the benefits the linguist stands to gain personally. There is also an obligation to discuss with the community the foreseeable repercussions of the research.

(2)    Where an Aboriginal community organisation has expressed the desire to approve or regulate linguistic research, the linguist has an obligation to work through that body, and in arranging and conducting research, not to act in any way which undermines local Aboriginal authority structures, traditional or otherwise.

(3) Communities have a right to expect that the linguist's work be of practical benefit to the community. In arranging, conducting and writing up research linguists should respond to the expressed needs of Aboriginal people and communities.

(4) Linguists should recognise that their own theoretical and descriptive technical works are in themselves of little value to Aborigina1 communities, and there is therefore an obligation to present results of research in a form as far as possible intelligible to and usable by the people who provided it, and their descendents. This should be recognised, by linguists and their sponsoring bodies, as *a* legitimate part of the research and time and funds should be allocated to it.

(5) Informants, teachers or consultants have a right- to remain anonymous, and the linguist should explicitly ascertain their wishes in this respect. If information is given on a confidential basis it should not be published or otherwise released. Individuals who do not wish to remain anonymous have a right to prominent acknowledgement in proportion to their contribution to the research effort.

(6)    Fair return should be given for all services, recognising that ultimately the judgement of what is 'fair return' rests with individuals and communities. (If work is done on. a strictly fee-for-service basis, ideally this should be at least the minimum wage, and research should be funded with this in mind.)

Control and release of information

(7)  The linguist should not publish or otherwise release material which is likely to have adverse consequences on the community or individuals.   Wherever possible, material should be presented for vetting before it is released.

(8)  Linguists must respect, the right to privacy of individuals referred to in published materials.  Unless there is an explicit understanding to the contrary, material must be presented in such a way that individuals cannot be identified.

(9)  Secret or restricted information should not be released except with the approval of its legitimate custodians, and all copies of such information may be recalled by them if they wish.

(10)  If the linguist published transcriptions or translations of texts; stories, songs, etc, then the source of that material, as determined by the community, should be acknowledged as the primary author and hold copyright.  All royalties from published works of this type should be directed to the appropriate Aboriginal individual or group.

(11)  The linguist must provide the appropriate individuals or groups with an index listing all recorded materials as well as any theses, articles, grammars, dictionaries, maps, etc which are substantially based on research in the community, and be prepared to provide them with copies on request.

(12)  If there are royalties gained through publication of a linguistic work on a particular language, the linguist should ensure a fair return to the appropriate individuals or groups.


(Drafted by Graduate Students in the Linguistics Dept., A.N.U., published in the Aboriginal Languages Association Newsletter, No.5, July 1983.)