Ethics issues in the archiving of sign language acquisition data

IPROSLA

Integrating and Publishing Resources of Sign Language Acquisition

Anne Baker

ACLC, University of Amsterdam

28 February 2012

1. Introduction

The discussion on ethical issues in the archiving of data has been a hot topic for several decades, and increasingly so with the considerable changes that are taking place in data access through the internet e.g., Mauthner & Parry (2010) in a paper on a joint UK and Dutch eResearch Ethics website. In the area of language documentation the ethical debate has a precise focus since informants are the direct source of all documented data. Language documentation in fact only makes sense when there is good archiving as Horton (2005) points out. The data are relevant for future research but also for the community of language users from whom they have come. For many endangered languages such documentation and archiving is crucial. The Assembly of Alaska Native Educators stated "Linguists can help strengthen heritage languages through … assist(ing) in conservation and preservation of heritage language materials including appropriate media and storage facilities." (quoted in Horton 2005).

The purpose of this brief paper is to indicate the ethical issues that are particularly relevant to the project IPROSLA in which an archive of sign language acquisition data will be created.

Williams et al (2009) identify three main areas for which there should be a clear ethical code in any archiving enterprise:

1. Continuation of participant protection beyond original purpose
2. Multi-media data integrity
3. Restriction of access to archived data

These will be dealt with in turn with respect to this project.

2. Participant protection

2.1 Legal situation

It has to be clear in any national context who has the copyright for the material. In some European countries, e.g. in Great Britain, there is a split between the copyright of the informant or interviewee and those of the researcher. There the informant holds the intellectual rights to the material and to any transcript; the institution that at the time of recording employed the interviewer or researcher holds the copyright to the actual recording. In Dutch law the rights to the recording and transcripts lie with the institution that employed the researcher at the time of the recording, but for audio- and video-recordings the portrait right is applicable (c.f. DANS web information). This right requires that explicit permission must be obtained from the interviewee in order to make a recording publicly accessible (see 2.2 and 4). The precise legal situation needs to be established in any specific country since this can vary and permission forms signed by the informant need to explicitly waive copyright. The institution also must explicitly hand over copyright to the physical recording when it is archived by an institution other than the one that made the recording. In the case of the IPROSLA project this agreement will be signed by both universities and the Language Archive of the Max Planck Institute for Psycholinguistics, Nijmegen (see Appendix 1). Dutch law is applicable here since the data were collected and archived in The Netherlands.

2.2 Informed consent

Before beginning recordings consent for the recording has to be obtained. This must be given in writing and is usually done for the series of recordings involved. There are different levels of so-called informed consent' that need to be explicit, such as for example whether the recordings may be viewed by anyone or only researchers, whether the recordings may be publicly used for scientific presentations, whether the real name of the informant may be given. It is crucial that the person providing the consent is aware of the nature of the permission given. Crasborn (2010) discusses this issue with regard to sign language data where it is of particular relevance that the communication with the deaf people involved is optimal. Signing a consent form means reading text that is formulated in the spoken language, here Dutch, and this is usually not the first language of the deaf adult. The permission therefore needs to be interpreted into the sign language of the adult, e.g., Sign Language of the Netherlands (NGT).

In the situation of language acquisition data, whether spoken or signed, a carer, usually the parent, gives consent for the recording to be made where he/she is involved but also his/her child. The child is a minor and so consent is given on his/her behalf. It could be the case that when the child reaches the age of 18, that he or she may wish to withdraw this consent. This possibility has to be available (see 2.1). The VSNU (2005) has set out guidelines for dealing with personal information in scientific research including all the definitions in Dutch law pertinent to this topic.

3 Multi-media data integrity

 3.1 Anonymization

In spoken language acquisition data, and the transcripts thereof, it is common for a pseudonym to be used. This is usual for example in the CHILDES database (MacWhinney 2000). However it often occurs in the recordings that the original name of the child and other family members are used. Such instances are rarely acoustically camouflaged in the CHILDES database and other such databases, although the guidelines of DANS suggest that this should be done. The use of pseudonyms is however to be recommended in all transcripts and meta-information for the recordings, and is standard practice.

In video recordings it is far more difficult to preserve anonymity. Techniques can be used such as blurring the face, blacking out the eyes as is illustrated in Fig. 1. In sign language data the use of such techniques would mean loss of essential information for the researcher since a considerable amount of linguistic information is provided by face movements such as eyebrow raise etc. it can also be important to analyze eye gaze. Such masking techniques are therefore not an option for sign language research. In the IPROSLA project pseudonyms are used in all transcripts related to the vide-recordings but the recordings themselves do not use any masking techniques.

It must also be pointed out that, although every care should be taken to maintain anonymity as far as possible, technology for face recognition is developing so quickly that an unethical user will soon be able to search on internet for similar faces and discover the identity of the informants.

Fig. 1.   Some options for masking identity in video recording: top: original image; middle: blurring of face; bottom : eyes masked. (Courtesy of Onno Crasborn, Radboud University Nijmegen)

3.2 Tweaking the data

In filming young children episodes can occur that are not suitable for public exposure, such as bathing etc. Unusable data also occur for example when the child goes out of camera. There may also be sections where other children can be seen for whom no permission has been obtained. For the archiving process such episodes should be removed. However for the integrity of research it should be recorded why such episodes have been removed so that no doubt arises about the scientific integrity of the researcher. In the IPROSLA project it has been noted in the meta-data which fragments of the original film have been excluded and why.

4. Restriction of access to archived data

Although it is desirable that archived data be made widely available, it is important also to try to prevent any abuse of the material. Most archives have a restricted access such that potential users must request access and when so doing, indicate their reason and qualifications for gaining access. The granting of access is usually in the hands of the archiving institution, since it is not feasible in the long term that individual researchers be consulted in this way. Researchers leave the field, retire and die and it is essential that there be continuity in this process. It might be desirable to allow the researcher to be consulted when access it requested for a period after the initial archiving. The DANS guidelines for the Netherlands recommend that the copyright of the data be handed over to the archiving institution in a written agreement. Clearly then it is the task of the archiving institution to organize and administrate access. The archiving institution needs to have the possibility of keeping such necessary records with respect to the permissions granted. It also needs to be alert for any instances of non-compliance to the conditions of use.

5. Summary

A number of recommendations have been made in this paper:

1. The legal situation with respect to copyright needs to be determined for the country involved and in the case of international projects for all countries involved.
2. Informed consent needs to be understood by the adult informants with all its implications in the internet age.
3. Recordings should be made anonymous by using pseudonyms in the transcripts and meta-data. Further measures such as the blurring of faces etc. is not possible for sign language research.
4. Sensitive or unusable episodes should be removed but indicated in the met-data with the relevant reason.
5. Restricted access should be standard procedure to protect the use of the data. Tnis should be essentially in the hands of the archiving institution.

References

Crasborn, O. (2010)   What does "informed consent" mean in an internet age? Publishing sign language corpora as open content. *Sign Language Studies, 10/2:* 276-290.

Crasborn, O., Zwitserlood, I., & Ros, J.. 2008. *Het Corpus NGT. A digital* open access *corpus of movies and annotations of Sign Language of the Netherlands*. Centre for Language Studies, Radboud Universiteit Nijmegen. URL: http://www.ru.nl/corpusngt/.

Data Archiving and Networking Services (DANS): www.dans.knaw.nl/data-archief/juridische-informatie. (Consulted February 2012.

Crasborn, O. & Zwitserlood,I. (2008) The Corpus NGT: an online corpus for professionals and laymen, In:*Construction and Exploitation of Sign Language Corpora. 3rd Workshop on the Representation and Processing of Sign Languages*,Crasborn, Hanke, Efthimiou, Zwitserlood & Thoutenhoofd, eds. ELDA, Parijs. pp 44-49.

Holton, G. (2005)   Ethical practices in language documentation and archiving. Paper to the Linguistic Society of America. Jan. 2005.

MacWhinney, B. (2000)   The CHILDES project. Tools for analyzing talk. Hillsdale, NJ: Lawrence Erlbaum,

Mauthner, M. & Parry, N. (2010)   Ethical issues in digital data archiving and sharing. On eResearch Ethics , http://eresearch-ethics.org/position/ethical-issues-in-digital-data-archiving-and-sharing/ October 2010. Consulted February 2012.

Vereniging van Universiteiten (VSNU) (2005) Gedragscode voor gebruik van persoonsgegevens in wetenschappelijk onderzoek. (zie ook bijlage). Web publication: http://www.vsnu.nl/Media-item/Gedragscode-voor-gebruik-van-persoonsgegevens-in-wetenschappelijk-onderzoek.htm. Consulted February 2012.

Williams, M., Dicks, B., Coffey, A., & Mason, B. (2009) Methodological issues in qualitative data sharing and archiving. Briefing Paper. University of Cardiff, UK. http://www.cardiff.ac.uk/socsi/hyper/QUADS/Briefing%20paper%20ethics.pdf.