Building a Database while Considering *Research Ethics in Sign Language Communities*

Julie A. Hochgesang[1], Pedro Pascual Villanueva[2], Gaurav Mathur[1], Diane Lillo-Martin[2]
[1]Gallaudet University, 800 Florida Avenue NE, Washington, DC 20002
[2]University of Connecticut, Storrs, CT 06269
E-mail: julie.hochgesang@gallaudet.edu, pedrodpedro@gmail.com,
gaurav.mathur@gallaudet.edu, lillo.martin@uconn.edu

Abstract

We are constructing an American Sign Language ID-gloss Database, which will enable sign language researchers and Deaf community members to access standard glosses for common signs. Since we are working with a language used by a community that has historically been marginalized during the research process, we feel the need to include an ethical framework for working with the Sign Language community as we consider best practices for developing sign language corpora. We will refer to the guidelines, Sign Language Communities' Terms of Reference (SLCTR), outlined in Harris, Holmes & Mertens (2009). Before making the database available to the ASL community, we plan to evaluate how members will use it and what they need from the research team to facilitate such use. This evaluation will go a long way towards ensuring that ownership of the research data lies with the ASL community. Such a reflexive evaluation of ethical practices is crucial from the beginning stages and throughout the research process. This means the ASL community is directly involved in the research process, is able to access aspects of the entire process, and can have a hand in the construction of knowledge about their own language, community and culture.

1. Introduction

We are constructing an American Sign Language ID-gloss Database, which will enable sign language researchers and Deaf community members to access standard glosses for common signs, as found in corpora such as those we are currently building. Our aim is to create a database which is flexible and powerful enough to be used by people in varying fields (e.g., linguistics, language teaching, interpreter training, preservation of Deaf heritage, etc.). As we start our work, we wish to consider not only the technical aspects of the endeavor (e.g., database design, transcription decisions, representative issues) but the ethical ones as well. We are working with a language that is used by a community that has historically been marginalized during the research process (Harris, Holmes and Mertens, 2009). It is established in spoken language corpora work that researchers need to be reflexive of ethical issues from the planning stage to publication and to be explicit about this process (Dwyer, 2006). As we consider best practices for developing sign language corpora, we feel it is necessary to also consider ethical frameworks for working with the Sign Language community. With this in mind, we are using the guidelines, Sign Language Communities Terms' of Reference (SLCTR), outlined in Harris, Holmes & Mertens (2009). This framework emphasizes "the need for the researchers to establish trust with the participants in the community and to ensure that the participants view the research as collaborative and culturally valued" (pp. 107).

2. Background – ID-gloss Database

For optimal usability, the corpora of sign languages should make data more accessible and useful; provide comprehensive and robust features for querying data; and be in a format that is automatically searchable and retrievable. Different uses require different levels of detail in transcription, but all require consistency in notation. For this reason, we have chosen to represent signs in our corpora using ID glosses, written English words which stand for sign lemmata (Johnston, 2008; see also section 4.1 below). In order to achieve the goal of transcription using consistent ID glosses, we need a common set of sign-gloss correspondences, easily searchable, accessible, and understandable. For this reason, we are constructing an ID-gloss Database (Alkoby et al. to appear).

The ASL ID-gloss Database will consist of two main components. The first is the 'global site', which contains a pool of video files and database field templates (such as those used to describe the sign's gloss, alternative uses, morpho-syntactic category, phonological descriptions, etc.). The second component consists of multiple 'local sites', in which user groups store their own group's information about each video file, organized according to the templates chosen by that group. Due to the structure of the database, each user group has the independent ability to determine how best to structure the glosses used by that group, and which information to include in addition to the gloss itself. Furthermore, the program will allow users to see (but not modify) the glosses used by other user groups. In this way, users may choose to adopt conventions followed by other groups, possibly leading eventually to a greater degree of consistency across research groups within the United States.

The first local site will contain the glosses and additional information used by the group of Deaf and signing hearing researchers developing this project, including (in alphabetical order) Karen Alkoby, Jeffrey Bernath, Paul Dudis, Julie Hochgesang, Diane Lillo-Martin, Gaurav Mathur, Gene Mirus, and Pedro Pascual Villanueva.

3. Sign Language Communities' Terms of Reference (SLCTR)

The set of SLCTR principles is unique in that it is among the first attempts to formally draft principles towards ethical conduct for research regarding the Deaf community. While most researchers working with the Sign Language community in the past may have been mindful of how they worked with the research subjects, there has been no consistent set of principles specific to the Deaf Community that could be used by the researchers. In other words, general research ethics tend not to take into consideration specific research ethics for certain communities, including the Sign Language community. Such a lack, Harris, Holmes & Mertens (2009) claim, has led to a lack of awareness of the particular cultural issues of the Sign Language community which sometimes subsequently results in harm to the Deaf community and therefore a reluctance in the Deaf community to further collaborate with researchers. In response to this, Harris, Holmes & Mertens drafted guidelines, adapted from the Indigenous Terms of Reference (Osborne and McPhee, 2000), in order to indicate respect for, show sensitivity to, address the importance of culturally appropriate research guidelines for, and acknowledge the culturally complexity of the Sign Language community. The guidelines are reproduced in Table 1 below.

1. The authority for the construction of meanings and knowledge within the Sign Language community rests with the community's members.
2. Investigators should acknowledge that Sign Language community members have the right to have those things that they value to be fully considered in all interactions.
3. Investigators should take into account the worldviews of the Sign Language community in all negotiations or dealings that impact on the community's members.
4. In the application of Sign Language communities' terms of reference, investigators should recognize the diverse experiences, understandings, and way of life (in sign language societies) that reflect their contemporary cultures.
5. Investigators should ensure that the views and perceptions of the critical reference group (the sign language group) is reflected in any process of validating and evaluating the extent to which Sign Language communities' terms of reference have been taken into account.
6. Investigators should negotiate within and among sign language groups to establish appropriate processes to consider and determine the criteria for deciding how to meet cultural imperatives, social needs, and priorities.

Table 1: Sign Language Communities Terms of Reference Principles (Harris, Holmes, & Mertens 2009)

4. Issues Related to our Project

As we begin work on the ASL ID-gloss Database Project, we have started to consider the project-specific issues that may arise throughout the course of our work. The three that we identify in this short paper are decisions related to gloss standardization, uses of the ASL ID-gloss Database, and transparency. We discuss each in turn in the following subsections. In general, we share the opinion that … "the formation of partnerships with researchers and the Sign Language communities is an important step in addressing methodological questions in research" (Harris, Holmes & Mertens, 2009, pp. 111). This guides our proposed solutions, aided by the SLCTR principles, to the issues discussed in the following subsections.

4.1 Glosses – Who Decides?

Glosses are the written representations of signs using the dominant spoken language of the Sign Language community. For example, in the United States, English is used in glossing ASL. There are problems related to glossing of Sign Language data, including inconsistency and incompleteness of representations (e.g., Johnston, 2008; 1991; Slobin, 2008; Mulrooney, 2006; Pizzuto and Pietandrea, 2001), yet the practice persists. Some linguists (e.g., Johnston, 2008; 2001; 1991) propose the use of ID glosses, consistent and unique labels for signs, to take some steps toward alleviating the well-documented problems associated with traditional glosses. We agree with this proposal and have begun to establish a database in which we will maintain a catalog of ASL glosses for the research community. As we undertake this project, we are fully aware that the data we work with comes from the ASL community. We feel we have a responsibility to consult the community while constructing written representations for signs from their own language.

Principle one of the SLCTR holds that "the authority for the construction of meanings and knowledge" rests with the Sign Language community. In that vein, we plan to survey community members in determining the ID glosses included in the database. Input from the community members will help to establish the optimal gloss we will use for each sign. We will target members of different sub-communities, including those with different backgrounds and those with different possible uses of the database (cf. section 4.2) in order to get a representative response.

Principle 5, in which the complexity of the cultural make-up of the Sign Language community is considered during the research process, is inherent in our treatment of the glosses as equal representations of as many ASL signs as we can feasibly include. Variation based on region, age, gender, education and other social factors will not be used to include or exclude any certain ASL sign. If the signs are linguistically different (based on our ultimate set of criteria), they will receive different

ID glosses. We will not intentionally exclude signs that may be considered by some to be used by a minority of the Sign Language community. In this treatment of the data, we avoid highlighting certain ASL signs as representative of the entire Sign Language community. We will stress in the literature regarding our database that any unintentional exclusion is due to our being unaware of such signs, as well as our limitations by time and funding to including only a subset of all signs.

We are also mindful of the fact that glosses are not cultural artifacts (as pointed out by our collaborator Paul Dudis) but tools of the scientific realm. This means that ultimately factors including the goals of the research project, the issues well discussed in the field regarding glosses and representation of data, and the input from the Sign Language community will all be considered as we make our final decisions in selecting the ID glosses to be used in our component of the database. All of the factors discussed here have also entered into our decisions regarding the design of the database, and in particular our implementation of a system which will allow different user groups to construct their own catalog of ID glosses which are best suited for their own purposes.

4.2 How the Database Will Be Used

The Amsterdam Manifesto, prepared by a group of sign linguists following the meeting of the conference on Theoretical Issues in Sign Language Research in Amsterdam in 2000, raises the point that much of sign language research is dependent on Deaf research assistants as well as data from Deaf native signers. The manifesto suggests that one way to acknowledge the contributions from these sign language communities is to give something back to them.

The ID-gloss database as described above clearly draws on and describes data from Sign Language community members. The question raised by the Amsterdam Manifesto and SLCTR regarding the database is, then, what can the investigators give back to the Sign Language communities in exchange for establishing this database? Is it sufficient to allow access to the database by the Sign Language community members? These questions ultimately depend on the issue of how the database is to be used.

The second and sixth principles of the SLCTR provide guidance in addressing these concerns. In their discussion of the second principle, Harris, Holmes & Mertens (2009) talk about how important it is to publish some of the research in sign language, rather than publishing in written language all the time. The underlying premise of this principle is that Sign Language community members should have access to the research, and publishing some of the work in sign language is one way to provide that access. The sixth principle says, in essence, that investigators should work with Sign Language groups to establish processes so that the research would meet the Sign Language communities' priorities. These principles can be applied in the context of the ID-gloss database. Here, we outline two ways that we do this.

First, we make the database as accessible as possible to the Sign Language community members. It is important to bear in mind that the database is intended to be a research tool that enables easier and more consistent transcription. It is not intended to be a dictionary, even though it shares some elements in common with one (e.g., an entry will include an image of the sign, a corresponding gloss, its meaning and its phonological description, among others). However, this intended use does not mean that we cannot share the database with Sign Language community members, and that they would not find appropriate uses for it. We could, for example, design a user interface specifically for Sign Language community members that would permit them to understand clearly the purpose of the ID-gloss database. This would address the second principle, in which we acknowledge their right to ensure that what they culturally value as a Sign Language community is included.

Another way to address the sixth principle is to set up guiding principles, in close consultation with Sign Language community members, on how to use the database. The guiding principles should clarify, for example, whether users are allowed to download and/or disseminate the information from the database. The guidelines should also specify who can add and modify entries in the database, and for what purposes the database can be used, e.g., for a conference presentation, for classroom instruction, and/or for purely research purposes.

By opening up the ID-gloss database to Sign Language community members, issues of ownership and researchers giving something back to the community are at least partially addressed.

4.3 Transparency

Transparency requires that researchers are open and reflexive about their information regarding the community being studied. In terms of the Sign Language community, researchers must adhere to transparency in a way that is accessible, i.e., in the community members' own sign language. Being transparent is a factor in meeting most of the SLCTR principles.

On the website where our ASL ID-gloss Database is hosted, we will provide signed ASL text wherever there is written English text. This practice of providing Sign Language text has been established by some other signed language corpora (e.g., the BSL corpus which can be found at: www.bslcorpusproject.org, last accessed March 20, 2010). We intend to adopt this

practice. In addition, the specialized user interface, as introduced in section 4.2, should allow Sign Language community members opportunities to provide feedback on aspects of the database, e.g., through comment boxes that accept video media (therefore signed input) and through polling. This is directly concerned with the sixth principle of the SLCTR, in which the Sign Language community helps establish research procedures. We will provide community members with the accessible opportunity to give input on the design and content of the database in a way that reflects their priorities.

By being transparent, we indicate our respect and understanding for practices culturally appropriate to the ASL community.

5. Discussion

We would like to emphasize that while we deem it extremely important that the Sign Language community be involved in the research process, we are aware that they do not possess the same scientific training or knowledge as sign language linguists do. We plan to honor the SLCTR, Amsterdam Manifesto, and the Sign Language community by being reflexive of and transparent about our practices and collaborating with the Sign Language Community, while simultaneously meeting the requirements of the research community. In fact, the membership of the Sign Language community and the research community overlaps, as there are some sign language linguists who are Deaf or otherwise members of the Sign Language community; there are of course also some sign language linguists who are not members of the Sign Language community. The SLCTR principles apply equally to all sign language researchers.

In this paper, we have discussed a few particular strategies regarding how we are implementing the SLCTR principles, including our actions and which SLCTR principles they reflected. We plan to continue consulting the SLCTR, including the principles we did not address in this paper, throughout the process of our research project.

6. Conclusion

As researchers, our focus is usually on theoretical, experimental, and/or technical aspects of our projects. However, it is important for us to bear in mind that the language we are so deeply involved in studying has a rich and important cultural value to the members of the Sign Language community. To appropriately follow relevant ethical considerations as we conduct our research, we must consciously consider and implement principles which have been determined to be suitable and applicable for studies in this area. Such a reflexive evaluation of ethical practices is crucial from the beginning stages and throughout the research process. This means the Sign Language community, in our case the ASL community, is directly involved in the research process, is able to access aspects of the entire process, and can have a hand in the construction of knowledge about their own language, community and culture.

7. Acknowledgements

The project described in this article was supported by Award Number R01DC009263 from the National Institute on Deafness and Other Communication Disorders. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders or the National Institutes of Health.

Thanks to Paul Dudis and Raychelle Harris for input on this article.

8. References

Alkoby, K., Bernath, J., Hochgesang, J.A., Mirus, G., & Pascual Villaneuva, P. (to appear). Construction of an ID Gloss Database for American Sign Language. To be presented at the conference on *Theoretical Issues in Sign Language Research*, Purdue University, October 2010.

Dwyer, A.M. (2006). Ethics and Practicalities of Cooperative Fieldwork and Analysis. In J. Gippert, N.P. Himmelmann, & U. Mosel (Eds.), *Essentials of Language Documentation*. Berlin: Mouton de Gruyter, pp. 31--66.

Harris, R., Holmes, H.M., & Mertens, D.M. (2009). Research Ethics in Sign Language Communities. *Sign Language Studies,* 9(2), pp. 104--131.

Johnston, T. (2008). Corpus Linguistics and Signed Languages: No Lemmata, No Corpus. In O. Crasborn, E. Efthimiou T. Hanke, E.D. Thoutenhoofd, I. Zwitserlood (Eds.), *Proceedings of the Third Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora*, pp. 82--87.

Johnston, T. (2001). The Lexical Database of Auslan (Australian Sign Language). *Sign Language and Linguistics*, 4(1/2), pp. 145--169.

Johnston, T. (1991). Transcription and Glossing of Sign Language Text: Examples from Auslan (Australian Sign Language). *International Journal of Sign Linguistics*, 2(1), pp. 3--28.

Mulrooney, K. (2006). *The Structure of Personal Narratives in American Sign Language*. Doctoral dissertation. Gallaudet University: Washington, DC.

Osborn, R. & McPhee, R. (2000). Indigenous Terms of Reference (ITR). Presented at the 6th UNESCO-ACEID International Conference on Education, Bangkok, December 12-15.

Pizzuto, E & Pietandrea, P. (2001). The Notation of Signed Texts: Open Questions and Indications for Further Research. *Sign Language and Linguistics*, 4(1/2), pp. 29--45.

Rathmann, C., Mathur, G. & Boudreault, P. (2000). Amsterdam Manifesto. *Das Zeichen* 14, pp. 654--655.

Slobin, D. (2008). Breaking the Molds: Signed Languages and the Nature of Human Language. *Sign Language Studies*, 8(2), pp. 114--130.