Linguistic Fieldwork and IRB Human Subjects Protocols

Denise DiPersio, Linguistic Data Consortium

Most linguists interested in planning or conducting fieldwork (particularly if they are in a University setting) are familiar with the need to develop a study protocol that must be approved by their institutions' Institutional Review Board (IRB). Preparing that protocol can be daunting: understanding the process, completing forms, answering questions from the IRB. Most universities offer some level of support to faculty and students for IRB reviews, and the web is a good (though sometimes overwhelming) resource for information.

Below is a brief review of the the current IRB system presented for background purposes only. This is followed by a discussion of the shortcomings of the IRB process with respect to social science research generally, the Linguistic Data Consortium's (LDC) IRB experience as an example of how linguists can approach the IRB process for a protocol that meets federal requirements and serves research goals and a list of IRB resources for linguists available online.

Background -- IRB History and Practice The IRB system as it stands today grew out of the 1979 Belmont Report, a study performed by a US government-appointed blue-ribbon commission that articulated ethical principles and guidelines for protecting human subjects participating in research. Those principles and guidelines were designed to combat past heinous abuses in scientific research experiments using humans, including work conducted by the German Nazi regime in the 1930s and 1940s and the US Public Health Service Tuskegee syphilis study conducted between 1932 and 1972.

The Belmont Report identified three fundamental principles for all human subjects research: respect for persons, beneficence and justice. Respect for persons means preserving individual autonomy, obtaining participants' informed consent and being truthful with participants about the study. The idea of beneficence is to do nothing to harm subjects; the goal should be to maximize research benefits and to minimize harm to individuals. Finally, justice refers to procedures that are administered fairly, are reasonable and are not exploitative; if subjects are compensated for participation, that compensation should be fair.

The current federal regulations governing human subjects research, sometimes referred to as the "Common Rule", represent the implementation of those ideas. In practice, IRB protocols are generally required to address the following questions:

1. Will the study require the participation of vulnerable populations? Those are defined in the regulations as children, pregnant women, fetuses, neonates and prisoners and as populations vulnerable to undue influence or coercion. In the case of the latter group, which may include fieldwork subjects, IRBs will want to be assured that subject participation is voluntary and that there are no circumstances that could be considered coercive. Supporting examples on the latter point include providing subjects with the right to withdraw from the study at any time and/or to designate certain information provided as not usable.

2. How will informed consent be obtained? Written consent is preferred, but some study conditions are not conducive for obtaining written consent. The regulations allow for the waiver of written consent under appropriate circumstances.

3. How will confidentiality be maintained? This refers to both any personal identifying information (PII) collected and to study data. With respect to PII, for what purpose is it being collected and how will it be segregated from the study data? For instance, PII may be collected for compensation purposes only; IRBs will want to know that is the case, that PII is not stored with the study data and is not otherwise linked to it, or if it is, how it is linked. With respect to fieldwork, IRBs

may want to know how the custody of such information will be maintained, e.g., in a log book in the researcher's possession while in the field, uploaded daily to a laptop and/or mass storage device, transferred to a secure network upon return from the field.

IRBs and Social Science Research

Those who have submitted protocols to their IRBs for linguistic fieldwork know that the process is geared to the consideration and approval of clinical (medical) research studies. The information requested in the submission form is not always relevant or written with social science applications in mind. Linguists must be prepared to be flexible and sometimes creative in preparing an application.

Moreover, despite the presence of the federal regulations, IRBs are not uniform across institutions in their approach to human subjects research issues in the social sciences. An example is the use of crowd-sourcing options as a survey method or as a means for collecting linguistic data (e.g., transcriptions, translations). Some IRBs have declared that crowd-sourcing does not constitute research with human subjects; others have said that it does, but that such studies are exempt from review; yet others require review and analyze the proposed use on a case-by-case basis. Practices like these inhibit collaboration across institutions as well as the ability of researchers to develop uniform standards for a particular kind of study.

Finally, there are researchers (linguists among them) who perceive IRBs as obstacles to research. The validity of that claim is open to some doubt. In a recent anonymous survey undertaken to examine that perception, linguists reported that their protocols were generally approved with little problem, inconvenience or adverse effects on the study (Bowern, 2010). It may be that the bigger problem is marshalling resources to prepare and submit the protocol in a timely manner. In a 1992 survey of IRB members probing issues in social science and behavioral research, respondents reported the following as among the "major" problems with protocols: "[p]rotocol is treated as a bureaucratic evil, not as a planning tool"; and "[p]rotocol is unclear to the non-specialist" (Sieber and Baluyot, 1992: 10). Linguists should bear comments like these in mind as we prepare IRB submissions.

LDC's Human Subjects Protocol

LDC has maintained a human subjects collection protocol with its host institution the University of Pennsylvania's IRB for almost twenty years. Issued initially for multilingual conversational telephone studies, the protocol has expanded to include various kinds of speech, text, handwriting and language-related judgments through on-site studies at LDC, fieldwork and crowd-sourcing. LDC seeks to design its protocol for flexibility so that individual modification requests to the IRB -- to add new studies, to modify existing studies, to approve new or revised consent forms -- can be handled under expedited review. That strategy has been largely successful but is challenged by fast-changing modes of collection and increased IRB attention on social science research.

LDC's protocol contains a description of study design that covers its general procedures and describes the current universe of, and judgments about, linguistic behavior that it collects. Some of the features of this protocol are listed here:

- Record linguistic performance – speech, writing, typing, dictation – by subjects in person, over the phone, using a computer mediated device, on a writing surface (paper, chalkboard) with no interlocutor or a human or machine interlocuter; subjects may optionally wear a headset transmitting silence or noise.

• Collect judgments about linguistic behavior and decisions involving linguistic data including but not limited to: auditing speech recordings, judging handwriting legibility, summarizing written text and reading comprehension.

• Collect linguistic performance information on human subjects including but not limited to: gaze tracking and keystrokes per minute.

Participants provide demographic information possibly including date of birth, sex, height, weight, native language, foreign language competency, literacy, handedness, education, occupation and regional –residency background. Where specific categories of demographic information are not strictly required for research purposes, questions involving those categories are optional or excluded. Collected linguistic data and demographic information which provides /permits speaker- anonymity will be retained for use in developing corpora which are used for language-related research, education and technology development.

Personal identifying information (PII) (name, address, phone number, social security number, email address) is collected for compensation purposes only, is stored separately, and is not linked or associated with collected linguistic data. Each participant is assigned a unique Personal Identification Number (PIN) to track his or her participation in the study and ensure appropriate compensation. Personal identifying information is not retained for corpus development and is not included in research corpora. Personal identifying information may be retained as required to comply with federal regulations and to provide contact information for potential participants in future studies.

The data collected is retained at LDC and used to develop linguistic corpora which LDC makes available to the larger research community through its public catalog.

Study methods, data management plans and consent procedures are revised as needed for particular work. For instance, with respect to fieldwork, LDC's protocol contains descriptions for the following:

(1) Non-remote field locations (*e.g.,* Philadelphia, PA; Seoul, Korea): speech data collected from participants is recorded directly to digital recorders and/or computers used by field workers and copied to a database at LDC as soon as practicable.

(2) Remote field locations (*e.g.,* Papua New Guinea): Speech data collected from participants is recorded directly to digital recorders used by bilingual native speakers, uploaded into a database on a laptop computer and then backed up on mass storage devices in the field location; that information will be uploaded to a database at LDC immediately following each field trip.

(3) Collecting personal identifying information (name, address, phone number, social security number, email address) in the field (*e.g.*, Philadelphia, PA; Seoul, Korea; Papua New Guinea): personal identifying information is collected in field locations using logbooks that are captured into a spreadsheet on a laptop computer, backed up on a mass storage device and then uploaded to secure storage at LDC following each field trip.

(4) Protecting data collected in the field: Mass storage devices and logbooks are stored securely, and the PII spreadsheet is encrypted; they are under the field workers' exclusive control until they and their contents are transferred to LDC where the data is stored on a secure network and the logbooks are kept in a locked file cabinet.

(5) Consent (Papua New Guinea): In field locations involving unwritten languages and/or speakers not literate in their native languages, participants provide verbal consent that will be recorded.

These provisions are examples of how fieldwork can be described to address the issues about which most IRBs will be principally concerned. During the modification process, LDC has been requested by its IRB to provide specific examples of questions that participants will be asked (e.g., for linguistic judgments). In the case of the consent waiver for fieldwork in Papua New Guinea, LDC submitted the English "script" of the consent that would be read to participants by bilingual native speakers.

In LDC's experience, preparation is the key to successful IRB review. Linguists will need at some point to prepare questions (or question types), consider alternatives to consent and plan data collection and management in the field. Indeed some or all of that information was likely necessary for the initial funding proposal. Having that material already in hand and incorporated into the proposal for the IRB will increase the likelihood of a relatively prompt and smooth review process.

IRB Resources for Linguistic Fieldwork

Template protocols for linguistic fieldwork approved by IRBs at various institutions are available online. These templates can facilitate the approval procedure to the extent they already address the significant issues and include formulae to explain the importance of sharing among researchers. Local boards may be persuaded to favorably consider a proposal based on similar studies approved by other boards.

Sites of note include:

TalkBank, http://talkbank.org/. The goal of this project, coordinated by Carnegie Mellon University (CMU), is to foster research in the study of human and animal communication. The community is encouraged to collect its own data and share it via the TalkBank network. Sample IRB-approved protocols and consents forms for TalkBank studies from various institutions (CMU, University of Kansas, University of South Florida) are available from this page, http://talkbank.org/share/irb/. Researchers can design the level of data access for their study by consulting suggested consent options from this page, http://talkbank.org/share/irb/options.html

Stanford University Linguistics Department Experimental Protocol, http://linguistics.stanford.edu/department-resources/experimental_protocol/. The protocol is described as appropriate for "linguistic experiments that involve no risk or personal information (beyond information that we use to select speakers)". The page includes links to the protocol, advertising flyers and various consent forms.

University of California San Diego Experimental Handbook, http://idiom.ucsd.edu/~bpajak/exp.html. This was developed by a linguistics graduate student and includes a sample approved application and consent form.

SIL Materials. Some data collections may include IRB documents and information for the particular study. For example, SIL Language and Culture Documentation and Description 2, Agta Demographic Database: chronicle of a hunter-gatherer community in transition, http://www.sil.org/silepubs/abstract.asp?id=49227 includes the Informed Consent Documentation (including IRB materials) for this data collection, http://www.sil.org/silepubs/Pubs/49227/49227_InformedConsent.pdf

Conclusion

Obtaining an approved protocol from an IRB for linguistic fieldwork can be relatively straightforward provided that researchers understand the process and are prepared to address how the proposed work will meet the requirements of respect, beneficence and justice. Model documents are good starting points that can be adapted to individual Board requirements. In most cases, the goals and methods of linguistic fieldwork are well within the scope of the federal regulations. The biggest challenge remains articulating that fact to the IRB.

References

Bowern, Claire. 2010. Fieldwork and the IRB: A snapshot. *Language* 86.4. 897-905.

Sieber, Joan E. and Reuel M. Baluyot. 1992. A Survey of IRB Concerns about Social and Behavioral Research. *IRB: Ethics and Human Research* 14.2. 9-10.

The Belmont Report – Ethical Principles and Guidelines for the protection of human subjects of research, http://ohsr.od.nih.gov/guidelines/belmont.html.