Forthcoming in Language Documentation. Practices and values, by Grenoble, Lenore A. and Louanna Furbee (eds.), scheduled to be published in September 2010.

Constructing adequate language documentation for multifaceted cross-linguistic data

A case study from the virtual center for study of language acquisition*

Barbara Lust, Suzanne Flynn, María Blume, Elaine Westbrooks, and Theresa Tobin

This paper confronts the challenge of constructing language documentation and data management in the face of continually expanding sets of cross-linguistic multi-media data arising in collaborative language acquisition research. It describes the development of an infrastructure and methods for creating and managing such shared language data across a Virtual Center for Language Acquisition(VCLA) by fostering collaborative scientific research in the language sciences across multiple institutions. The infrastructure reflects a research lab/academic library collaboration that integrates metadata organization in research methods. This paper describes both the research and educational components involved in the development of the VCLA

In this paper we describe collaborative work in which we seek to establish best practices for documentation of large, continually expanding amounts of language data of various types. Existing multimedia data in one lab alone (the Cornell

* The authors wish to express their thanks to the co-editors of this book for important feedback and discussion concerning the contents of this paper, as well as to Janet McCue, associate university librarian for teaching, research outreach, and learning services at the Cornell Library for her insight and guidance regarding the project described in this chapter. In addition, the authors thank Alex Immerman, Brian Lowe, and Gail Steinhart for their helpful comments and suggestions throughout the revision process. This chapter was prepared with the partial support of National Science Foundation (NSF) Small Grant for Exploratory Research NSF-0437603 to Janet McCue and Barbara Lust and with NSF Office of Cyberinfrastructure grant OCI-0753415 to María Blume and Barbara Lust.

Language Acquisition Lab [CLAL]) currently involve thousands of samples of language at various periods of language acquisition (child and adult), in various situations (naturalistic and experimental), and across more than 20 different languages from no fewer than 20 countries. Through a Virtual Center for Language Acquisition (VCLA; http://www.clal.cornell.edu/vcla), this language-acquisition lab can link both nationally and internationally to many others who are interested or involved in language-acquisition research. Thus, we must now prepare for exponentially increasing cross-linguistic data to accumulate and enable continual collaborative work with these data across distance and time. By linking researchers with academic librarians, we seek to develop a documentation system for present and future data that at once (1) links the data to domain-specific linguistic analyses that are necessary for research; (2) attempts to calibrate data across various languages in so doing; (3) links to current fieldwide standards for language description, such as those being developed by Electronic Metadata for Endangered Languages Data (E-MELD); (4) links to fieldwide resources according to standards such as those currently under development by the Open Language Archives Community (OLAC); and (5) links to the crucial upper-level documentation system of an academic library and interlibrary domain, which, through metadata systems and general Web-based ontologies, situates language data in a general knowledge domain and renders it accessible to library users worldwide. In this paper, we report on our program and its progress and challenges in this endeavor.

1. Theoretical issues

We now work in an age during which developments in cyberinfrastructure offer new possibilities for research questions and methods (Atkins 2003; Atkins et al. 2003; National Science Foundation Cyberinfrastructure Council 2007; Borgman 2007). Linguists have begun to investigate how and where the power of cyberinfrastructure can be brought to bear on the scientific study of language and the language sciences. The documentation of endangered languages provides one example of the opportunities cyberinfrastructure affords. Other possibilities involve the development of new scientific methods in the language sciences. For example, new possibilities now exist for interdisciplinary collaborative research and for empowering cross-linguistic and cross-cultural research in a global perspective.

Realizing these recent possibilities, however, requires development in the field of linguistics and the language sciences. For example, such research developments require (1) an infrastructure of collaboration; (2) standardized tools of *best*

practices that can be shared while at the same time allow unique methods by individual researchers; (3) infrastructure for data storage, management, dissemination, and access, including means for interfacing diverse databases that differ in both type and format; and (4) protection and "portability" of data and related materials into the future. In this paper, we will focus on (2) and (3), the development of standardized tools of *best practices* and the development of infrastructure, exemplifying developments in these areas that have emerged in constructing the recent VCLA.[1] The development of a collaborative culture is currently under study (e.g., Pfirman et al. 2005; and Science of Collaboratories http://www.scienceofcollaboratories.org/), and portability issues have been extensively elaborated upon in Bird and Simons (2003) and Simons (2004).

All these developments in turn require the establishment of rigorous and shared methods of data creation and data documentation. For example, unless data provenance is well recorded and continually linked to language data, language samples are of limited scientific use. Without such documentation, language data cannot survive the extensive process of scientific data creation; data storage for shared use, access, and dissemination; or data calibration for comparative and/or collaborative research.

1.1 Data creation

In the case of language, these requirements are challenging because the conversion of language samples to scientific data is not straightforward; sound waves in the air do not instantly constitute data. Rather, data must be created. (See Appendix 1 for a sketch of initial steps in data creation in one component of the Virtual Linguistics Laboratory [VLL] that is being developed in the VCLA.)

Language data arise in multimedia formats (audio and video, analog and digital). Various linguistic theories are invoked across the field for data description and analyses, creating a need to interface theoretical vocabularies. Varied languages have their unique needs for description in language typology. The search for language universals requires uniform formats for cross-linguistic comparisons. This last challenge is being confronted by the General Ontology for Linguistic Description (GOLD) project in the E-MELD enterprise. Audio or audiovisual

1. The founding members of the VCLA include the following – Cornell University: Professors B. Lust, E. Temple (now at Dartmouth), Q. Wang, M. Casasola, J. Gair, and C. Cardie; California State University, San Bernardino: Y-C Chien; Massachusetts Institute of Technology: S. Flynn; UTexas at El Paso: M. Blume; Southern Illinois University: U. Lakshmanan; Rutgers-Newark: J. Austin; Rutgers-New Brunswick: L. Sanchez; and MIT and Boston College: Claire Foley. Description of founding-member research interests can be found at www.clal.cornell.edu/vcla.

samples (and video samples, in the case of sign language) provide the authoritative archival form of language data, creating technical challenges (e.g., Grotke 2004). Generating transcriptions of language requires a time consuming, cognitive and analytic process with variation expected across individual transcribers (Edwards 1992a, b). At every moment, different points of data creation must be linked, and sound methods of data documentation must be applied.

Finally, language data arises from human subjects. This in turn requires procedures ensuring human subjects' protection and confidentiality both at the stage of data collection and at subsequent stages of data storage, archiving, and dissemination.

1.2 Language-acquisition data

In the case of research on language acquisition, language samples at various periods of language development, arising from various experimental and naturalistic methods, must be accessible in a way that allows comparability (either across samples from one child or adult language learner studied longitudinally or across samples from different children or adult learners studied cross-sectionally). Moreover, this comparability must be ensured across diverse languages. Often studies require analyses of large data sets, with numerous and continually expanding data points related to each set (e.g., all analyses performed on that data).

2. Values and practices

Unless high-quality metadata on the language source is available, the scientific worth of language studies is questionable. Since studies of language acquisition generally seek explanation of the source of language development, rather than merely description, the theoretically based methods for linguistic analyses as well as provenance records are critical. The study of language acquisition thus provokes all the basic issues of scientific methodology required for language data, and these issues are often intensified.

Additional ethical issues arise. Procedures for work with human subjects to ensure confidentiality and informed consent are set by individual institutional review boards (e.g., the required training programs such as the University Committee on Human Subjects education and training programs instituted at Cornell) in conjunction with new mandates by federal funding agencies (e.g., the National Institutes of Health; http://grants2.nih.gov/grants/policy/

data_sharing). Work with children as participants in these studies necessitates further steps, since children cannot be expected to give informed consent. All records regarding human subjects must become part of the complete language-documentation process.

Finally, intellectual-property rights must be addressed in the case of language data as for research data in general. Language data painstakingly collected and created by individual language scientists belong primarily to the researcher and to the institution in which they work. Principles for sharing data or scientific materials must be developed in a manner that respects this premise. Such agreements must also become part of comprehensive language documentation where language is to constitute scientific data.

3. Training

With today's growing concern for the need to share data across diverse repositories, and with new technical means for wide dissemination of data through cyberinfrastructure (Atkins 2003), researchers and teachers are struggling to find ways of managing data. Since usability of data requires structure for data access and comparative description of data, many fields of science are now only beginning to implement such structures. At this time, separate databanks have typically been created by individual researchers using different procedures for collecting, labeling, and storing data; methods now must be developed post hoc so that these diverse data sources can be linked, calibrated, and subjected to reliability standards. Often critical facts regarding data provenance are not known. Researchers must strive to constitute a post hoc structure for accessing and studying various preexisting data sets of various types in various formats and for letting data "speak to" data (Williams 1997; see Pearson 2004 on biobanks, for example; Nature 2005; Pennisi 2005).[2] In order to ensure that future language research is not similarly hindered, the primary research process must now be transformed. The rising generation of researchers needs to be trained in new methods to ensure that language data are henceforth created in such a way as to allow future use and reuse, collaborative analyses, and wide access. Researchers need tools to ensure language data that are reliable and authentic, archived and preserved long-term, confidential and private, and accessible in a variety of formats (e.g., AIFF [Audio Interchange File Format], WAVE [Waveform Audio File

2. We set aside the massive challenge of digitization and long-term (to perpetuity) storage of original archival data, such audio- and videotapes (National Science Board 2005), in order to concentrate here on the data-management problem (see also Nature 2009).

Format], MP3, transcript/.txt file, etc.). The data also must be described and preserved with systematic and significant metadata, which are in turn expressed in terms of both general concepts recognized across fields and specific concepts relevant to particular linguistic inquiries.

These training challenges exist in addition to the need to develop a culture of collaboration beyond what is now supported or encouraged in most academic environments (see Borgman 2007).

4. Case study

In order to meet the challenges we have summarized in sections 1 through 3, we are currently constructing an infrastructure that involves merging research labs with academic libraries (Figure 1) and developing the technology, systems, and human resources to support this merger in the area of the language sciences.

Libraries have traditionally been stewards of intellectual content, responsible for collecting expanding amounts of information, storing it over time, and developing systematic means for its widespread dissemination and access. In this role, they have developed the metadata structures necessary for the description and exchange of materials as well as systems and methods for preservation. They provide technical infrastructure for information storage and retrieval as well as consulting and outreach services.

In the information age, academic libraries are transforming themselves. With new vision, they are now becoming stewards, trustees, and custodians of research data, as exemplified by various digital initiatives (Cornell University 2007; also, see http://dcaps.library.cornell.edu for example). In this role, they are expanding their expertise to the preservation and management of various forms of research data. At Cornell, we are combining the developments of the VCLA with new vision and new initiatives at the Cornell University Albert R. Mann Library in order to explore the possibilities for integrating academic-library expertise with research needs such as those we articulated in sections 1 through 3.[3]

This collaboration promises not only to empower the VCLA but also to exercise and exemplify the developing strength of the academic library to meet the challenges of the expanding digital universe of research in new contexts provided by cyberinfrastructure.

3. Janet McCue and Barbara Lust, "Small Grant for Exploratory Research: Planning Information Infrastructure Through a New Library-Research Partnership," NSF-0437603 (unofficial project name, 'LiLaC'; http://metadata.mannlib.cornell.edu/lilac/). <<kept this in endnote but removed from ref list per your advice.>>

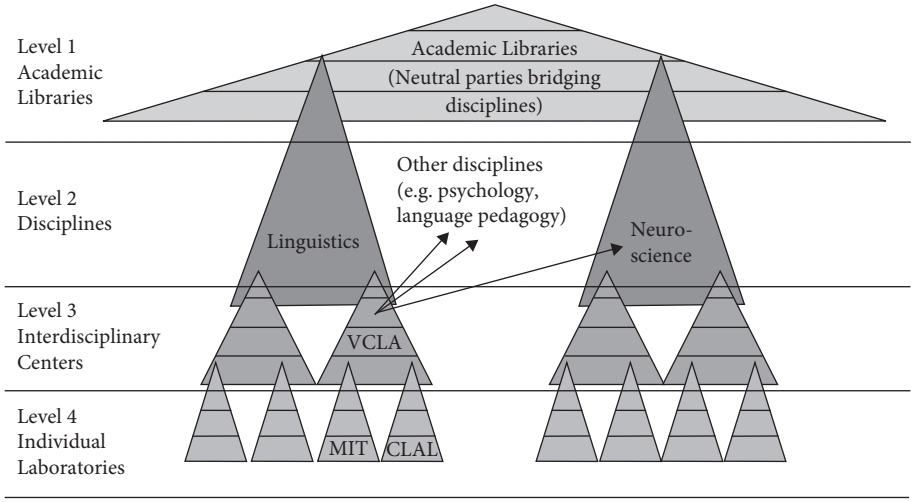


Figure 1. Multiple Levels of Discourse

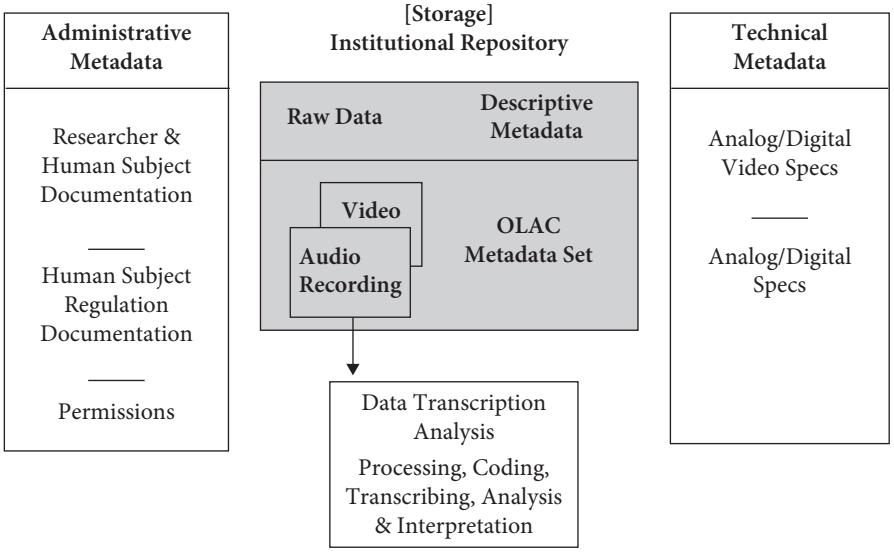


Figure 2. Metadata Infrastructure

The Mann Library has begun to advise the VCLA in the following areas: (1) metadata configuration (Figure 2); (2) automatic conversion of language materials descriptions to OLAC (Simons 2009); (3) formulation of best practices for audio and video archiving (Westbrooks, Pantle, and Lowe 2005); (4) audio digitizing and preservation, and (5) development of infrastructure for linking

lower- and higher-level ontologies for language data description and dissemination (e.g., Lust et al. 2005).

4.1 Interlibrary collaboration

Finally, as libraries transform themselves into digital management resources, so must interlibrary structures (for example, the concept of an interlibrary loan must now be extended to include "data grids" and other data and materials exchange structures). Thus, the infrastructure we build involves cultivating interlibrary collaboration necessitated by our project. As a first step, the Mann Library and the MIT Humanities Library are investigating bridge-building across these institutions. For example, this component of our project will identify metadata schemas that would be necessary for effective and efficient research data and materials exchange between Cornell and MIT research labs, where the academic libraries form a systematic conduit. This phase of the project will also explore resolving intellectual-property rights issues in cross-institution research exchange.

4.2 Institutional repository

In addition, the libraries are developing the load of materials in an online digital archive, DSpace (http://www.dspace.org/, http://dspace.library.cornell.edu/community-list), and other current alternatives such as Cornell's institutional repository, eCommons (http://ecommons.library.cornell.edu/), in order to assist in making academic scholarship freely available and in effectively utilizing metadata within shared environments (see also the current Albert R. Mann Library DataStaR project, http://datastar.mannlib.cornell.edu, as in Lowe 2009.

Since 2000, institutional repositories (IRs) have become a commonly used tool to help institutions manage a wide variety of materials generated by faculty and staff (e.g., publications, images, multimedia, preprints, literature). Cornell's eCommons repository makes use of the open-source DSpace software; other open-source and commercial solutions, such as Fedora (http://fedora-commons.org) are also used by libraries.

The instantiation of such software platforms empowers faculty and staff by providing a set of open-source tools that enable the collaborative storage, submission, and organization of any type of material. Such shared digital space, however, must be combined with metadata and data infrastructure such as the ones we have outlined herein so as to make data accessible and usable in teaching and in collaborative scientific endeavors. The burgeoning Cornell-MIT academic library infrastructure can be tested against other academic libraries, opening up the

potential for wide dissemination of and collaboration on language data and related documentation through interfaced databases.

5. Technology: The DTA tool

In order to cultivate a research lab–library infrastructure, as well as lab–lab exchanges, it is necessary for the research lab to develop means by which to create and describe its data and materials in a form that will integrate with academic library metadata structures, ontologies. and data-preservation techniques.

To this end, the CLAL has developed the VLL consisting of materials to ensure *best practices* in the area of the language sciences, particularly the area of language acquisition (http://vcla.clal.cornell.edu/en/vll/). One component of the VLL includes a Web-based Data Transcription and Analysis (DTA) tool.[4] This tool guides the researcher or student in data creation so as to meet the outlined challenges. Through a system of Web forms (a point-and-click interface with menu-driven operations), it guides the user through completion of a series of data and metadata fields for situating the data and establishing data provenance. It then guides the user through transcription and analysis of the (potentially cross-linguistic) audio or video data. Sample Web forms pertaining to subject and session metadata entry appear in Appendices 2a and 2b.

The DTA tool then leads the user through basic forms of linguistic description and coding. Eventually user-defined forms can be adapted to additional topics of relevance specific to any individual research study. Annotation fields in the tool record transcription and analysis histories from multiple transcribers and users over time.

The DTA tool's structure provides a framework for producing comparable, calibrated, scientifically valid and high-quality data, thus establishing grounds for collaborative and comparative data analyses across individuals and institutions. It integrates a primary research tool with the potential for permanent archiving in the form of a cross-linguistic relational database. By integrating its metadata structure with the academic library metadata system, it provides a primary mechanism for the transfer of research data from the research lab to the academic library and interlibrary infrastructure, where wide national and international outreach can be achieved.

4. We are indebted to Cliff Crawford, a former Cornell graduate student in linguistics, for development of the Web version of this tool.

6. Conclusions

Figure 3 summarizes the infrastructure being developed in this case study.

This infrastructure can be viewed in general as an attempt to enhance "scholarship in the digital age" such as discussed by Borgman (2007) and to do so specifically with regard to the study of language. The materials and cybertools developed in the VLL reflect an attempt to provide data-management principles and tools necessary for this scholarship. More current developments, made possible through the National Science Foundation (NSF), have now allowed us further to develop a pedagogical component. In particular, the NSF has permitted us to begin to address the recent charge to scientists in the face of the current explosion of data: "data management should be woven into every course in science, as one of the foundations of knowledge" (Nature 2009). Through a current NSF grant (Blume and Lust 2008), we are working with other founding members of the VCLA to develop a series of courses intended to educate a new generation of researchers and scholars in the use of cybertools, methods, and principles provided by the VLL. These courses are coordinated across diverse institutions, either synchronously or asynchronously. We together seek to teach the rising generations to conceive of data and metadata

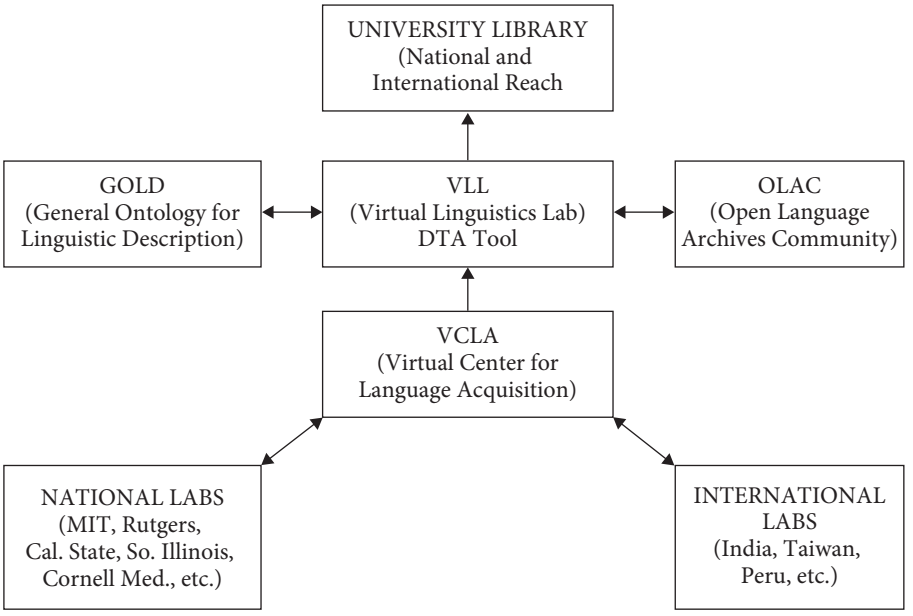


Figure 3. Representing Language Data: Linguistic Ontologies[5] (Lust et al. 2005)

5. Searching Interoperability Between Linguistic Coding and Ontologies for Language Description: Language Acquisition Data <http://www.emeld.org/workshop/2005/papers/lust-paper.html>

organization and management as fundamental components of the primary research process and of scientific knowledge. In addition, we wish to encourage and facilitate a collaborative community equipped to take advantage of all these components.

The project we describe in this chapter does share some properties of other initiatives in linguistics and the language sciences. However, our project is unique in its comprehensive attempt to develop an infrastructure and methods for language documentation that allow active access of data and related scientific materials. This in turn provides a foundation for continual, endlessly expanding collaborative research and teaching across diverse geographical and theoretical domains. The project is also unique in its attempt to invoke the academic library structure as a long-term, lab-independent component in research-data management, preservation, dissemination, and access. Lastly, the project uniquely situates the VLL in an educational environment in order to facilitate training in and dissemination of its products.[6]

Appendix 1

Virtual Center

Cornell University Virtual Linguistics Laboratory

Data-Creation Steps

Capturing natural language so that it can become reliable scientific data requires a multistep process. These steps provide one component of the Virtual Linguistics Laboratory's (VLL's) methods for the study of language acquisition. Note that while the creation steps follow a sequential order, many stages can and will be performed simultaneously. The full process involving these steps is designed to provide reliable data for reliable collaborative and interactive research through

6. For example, in linguistics, the Universals Archive at Universität Konstanz (http://typo.uni-konstanz.de/archive/intro/; or DELAMAN (Digital Endangered Languages and Music Archives Network (www.delaman.org). In child language acquisition, several different initiatives for data management, access and data sharing exist, each using different formats (e.g., MacWhinney and Snow 1985; Miller and Chapman 1983; Long and Fey 1993; Wimbish 1989; Lum et al. 1999; and others); most researchers develop individual systems for data management, which may not exist beyond individual research grants.

the Virtual Center. The Virtual Center is responsible for the infrastructure by which the multistep process of data creation is orchestrated and integrated.

These steps presume the prior establishment of scientific methods for the generation of natural language data (B. Lust, M. Blume, and T. Ogden, "Research Methods Manual: Scientific Methods for the Study of Language Acquisition" [Cornell University Virtual Linguistics Laboratory, in preparation; http://www.clal.cornell.edu/vcla]).

1. An audio (and/or video) recording is made of language behaviors. Speech so recorded provides the foundation for the following steps of data creation. This first recording is the primary, authoritative step in data creation.
2. Basic metadata surrounding the item is entered in the VLL Data Transcription and Analysis (DTA) tool first inventory forms. The metadata provides the basis for labeling along all further data-creation steps. Each subject is entered into the first forms of the DTA tool (Lust et al., "Cornell University Virtual Linguistics Laboratory Data Transcription and Analysis Tool Manual" [in preparation; http://www.clal.cornell.edu/vcla], to be used in conjunction with the VLL Research Methods Manual.[7] These first screens contain metadata regarding the subject and regarding the session(s) of recording.
3. Recording labeling is checked in accord with the system established in the VLL Research Methods Manual and entered into a recording database according to procedures established by the *Mann Library Digital Archiving Manual* (Westbrooks, Pantle, Lowe 2005).
4. A copy is made of the original audio (and/or video) recording.
5. A Stage I digitization is made from the audio recording and saved in a specified format (e.g., AIFF [Audio Interchange File Format] or WAVE [Waveform Audio File Format]), if the original is not itself in digital form. This first-stage digital recording is burned or exported to a hard-copy format (e.g., CD, DVD, solid-state drive [SSD]) and also saved on a CLAL/VCLA server. Its purpose is simply to copy the original recording, with minimal editing. It provides the authoritative archive copy in digital form.
6. A backup copy is made of this stage I digitization. Stage I copies are simply copies of the original recording from which the data came (possibly involving more than one subject/session per recording).
7. A Stage II digital file is created for each individual subject and exported to a hard copy and saved to the server. This provides the original digital audio record that will become the basis for research. The stage II digitization involves separating data that may have been combined on the original recording,

7. In the case where both a video and an audio recording exist, transcriptions across these need to be calibrated, and comments on "context" entered accordingly into the DTA tool fields specified.

such as separate subjects on a single tape/disc/SSD and separate tasks for a single subject (e.g., natural speech as well as experimentally derived language). A stage II record contains all sessions for a single subject, regardless of type of data elicitation (experimental or naturalistic). Each stage II record contains audio (or video) data for only one subject. General links to inventory metadata are made.

8. A backup copy is made of this stage II digital file.
9. A stage III digital file is then created from each stage II record. Here it is edited and formatted to assure the highest audio quality possible. Precise links to metadata for each task and each subject are made.
10. Three types of backup copies are made of this stage III digital file: for example, server, hard-drive backup, hard-copy backup in duplicate.
11. An initial transcription is made of the recorded speech. If this transcript is done by hand, the first transcript is then digitized and saved as digitization transcript #1.[8] Ideally, transcription is done on the basis of a digitized form of the original data, preferably a stage III form.
12. A second independent transcription is made of the recorded speech and saved as digitization transcript #2.[9]
13. A reliability check is conducted by comparing and contrasting transcriptions, noting discrepancies, and resolving these to provide an accepted working transcript. This reliability check includes listening to the whole record in the presence of the digital edited audio file (stage II). Annotations are added to the accepted working transcript to reflect where discrepancies occurred.
14. A phonetic edit is added to the accepted working transcript. These phonetic edits are also conducted in accord with the digital audio file. Both the audio and written data are precisely integrated. Editing especially includes cases where speech has been in some way deformed, for example, if the child (or speaker) has made an error in pronunciation. (In these cases, the standard

8. The VLL Research Methods Manual provides guidelines and methods for the transcription process. Note that if transcription has been initially done outside the Virtual Center, it may not have the benefit of these structured guidelines for transcription. However, subsequent re-transcriptions within the Virtual Center will be able to add this value in final reliability checking.

9. Transcription of speech from audio (and/or video) data is a critical step in the creation of natural language data, as discussed in the VLL Research Methods Manual. Given the nature of spoken language, transcription is in fact a form of linguistic analysis; it provides a cognitive transformation of heard speech into a linguistic representation. It thus varies naturally from hearer to hearer and speaker to speaker (see Edwards 1992a, b; and Edwards and Lampert 1993, for example). Hours of time may be required for completing reliable transcriptions on a small number of utterances. Transcripts vary widely in reliability accordingly.

spelling system cannot be used.) The phonetic edit provides a final reliability check on the data.[10]

15. The accepted working transcript is then entered into the next screens of the DTA tool, and a sequence of structured analyses and annotations begin through that tool (following the VLL Data Transcription and Analysis Tool Manual, to be used in conjunction with the VLL Research Methods Manual).
16. If the speech data involve a language other than English, then literal and general glosses are entered into the DTA tool screens accordingly.
17. At each stage in this process, the data involves an ID or signature, indicating the full set of steps that have been completed to date and allowing the researcher to indicate which stage of data they are using. Individual researchers who participate in various stages of data creation at various times are recorded in the database accordingly.
18. Human-subjects criteria for anonymity of records are maintained throughout (VLL Research Methods Manual). Data ID procedures involve an anonymous ID: subject-name initials plus birth date.
19. At this point, scientifically sound data have been created for possible collaborative research and for ultimate deposit in an institutional repository for ultimate, wider dissemination.

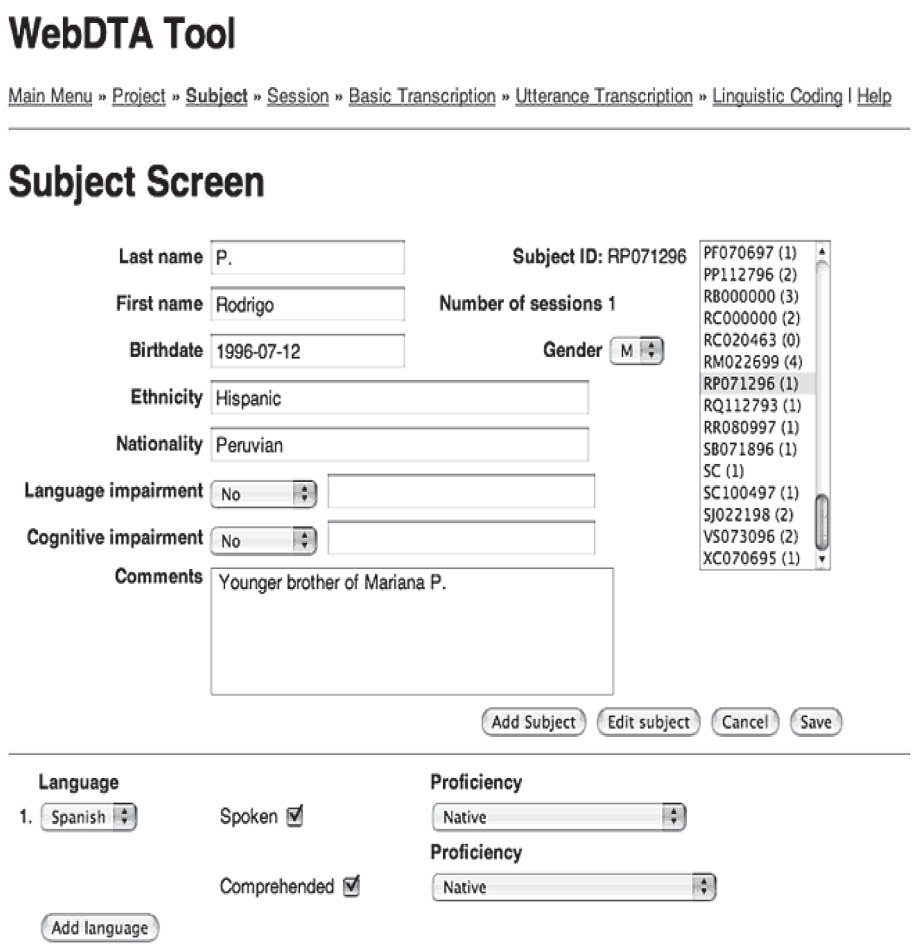
The full process of data creation is not a linear one. In fact, each time the created data are used and reused by researchers, further value is added to the data; transcriptions are newly amended and/or added to. The infrastructure designed by the VCLA VLL allows for this nonlinear process of data handling.

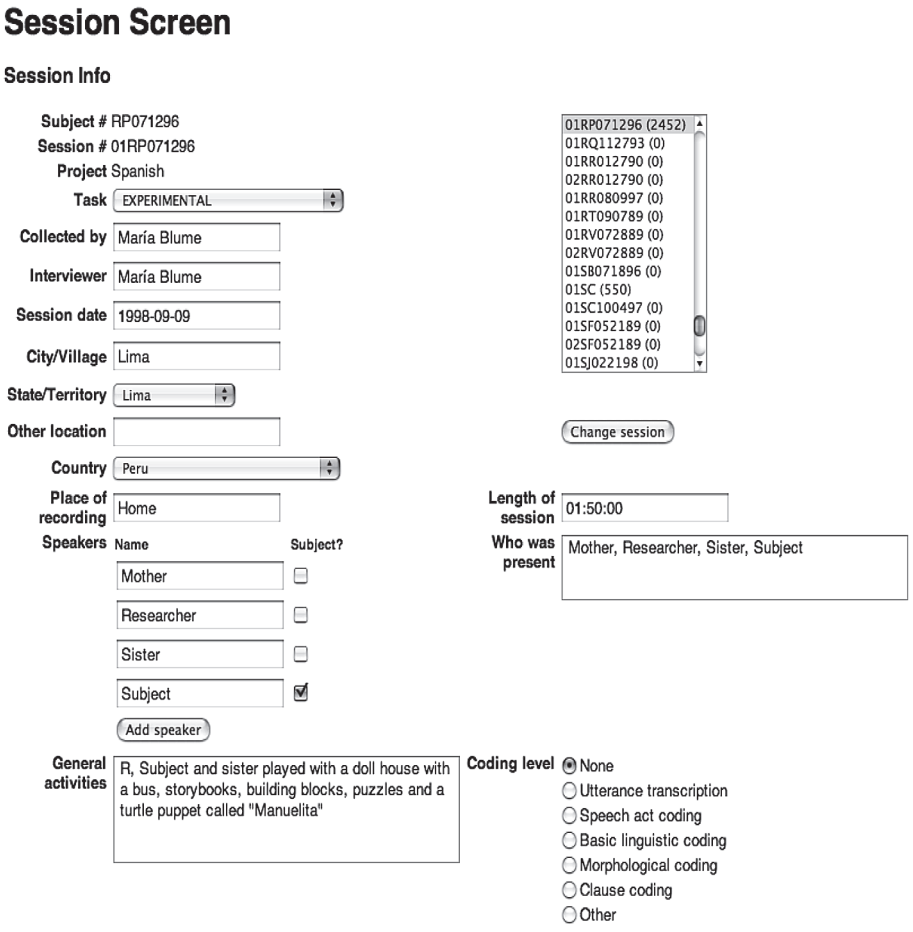
Audio and video data may require different formats for preservation (e.g., CD, DVD) and need to be adapted to ever-changing technological innovations.

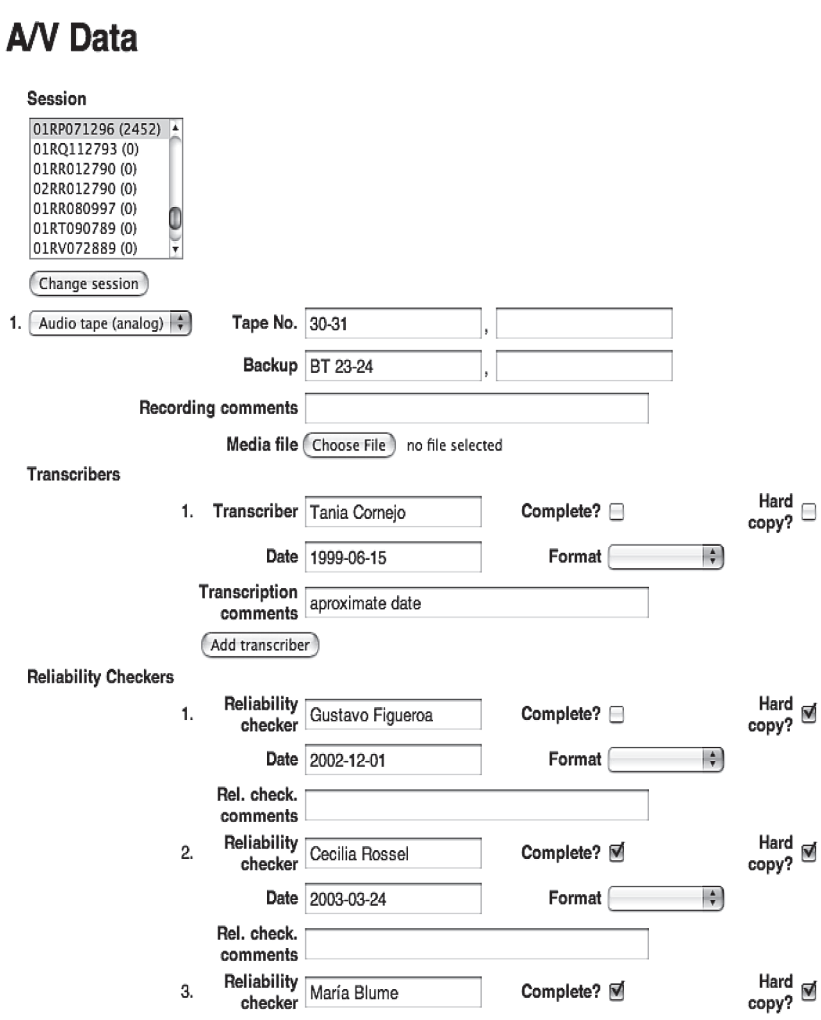
10. Phonetic edits may be partial (emphasizing the child's deformed forms only); or "full" (where a transcription is made completely in a phonetic alphabet). The latter would be required for a study concerned with the phonology of the language. Partial edits may suffice where the research questions concern the syntax or semantics of the language. Standard data creation in the CLAL/VLL assumes partial phonetic edits, unless specified otherwise.

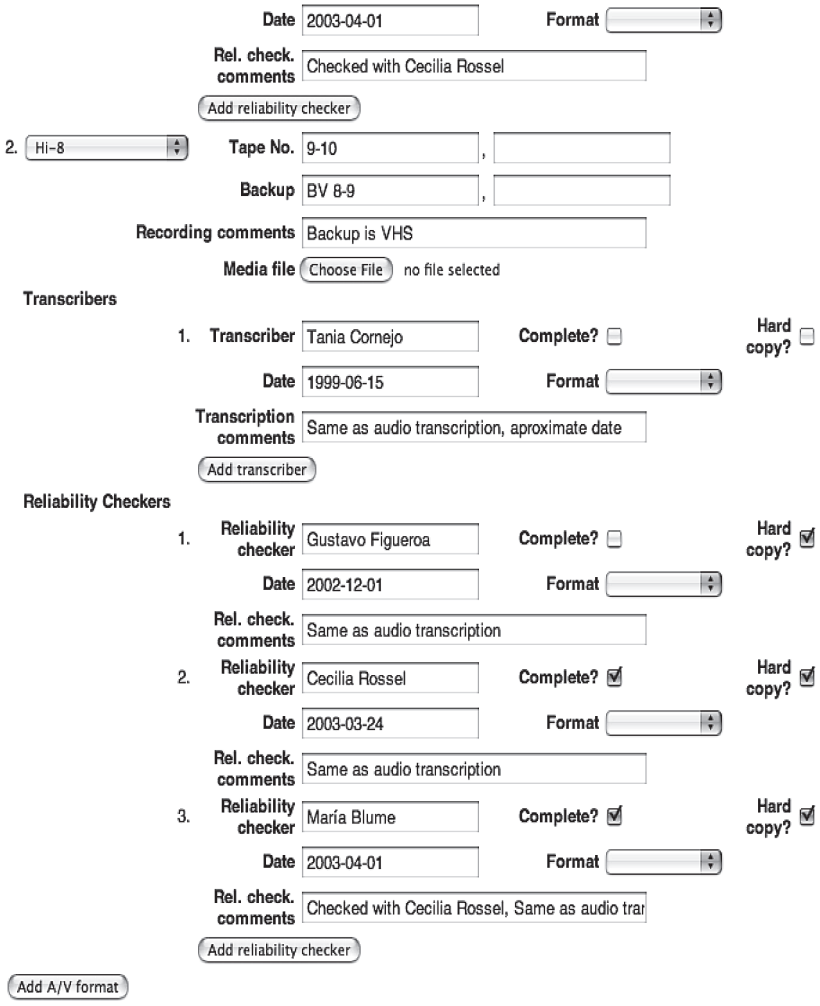
Appendix 2

Data Transcription and Analysis (DTA) Tool Sample Screens




Session Screen
Session Info
Subject # RP071296
Session # 01RP071296
Project Spanish
Task EXPERIMENTAL
Collected by María Blume
Interviewer María Blume
Session date 1998-09-09
City/Village Lima
State/Territory Lima
Other location
Country Peru
Place of recording Home
Speakers Name Subject?
Mother
Researcher
Sister
Subject
Add speaker
General activities
R, Subject and sister played with a doll house with a bus, storybooks, building blocks, puzzles and a turtle puppet called "Manuelita"
01RP071296 (2452)
01RQ112793 (0)
01RR012790 (0)
02RR012790 (0)
01RR080997 (0)
01RT090789 (0)
01RV072889 (0)
02RV072889 (0)
01SB071896 (0)
01SC (550)
01SC100497 (0)
01SF052189 (0)
02SF052189 (0)
01SJ022198 (0)
Change session
Length of session 01:50:00
Who was present Mother, Researcher, Sister, Subject
Coding level
None
Utterance transcription
Speech act coding
Basic linguistic coding
Morphological coding
Clause coding
Other


A/V Data
Session
01RP071296 (2452)
01RQ112793 (0)
01RR012790 (0)
02RR012790 (0)
01RR080997 (0)
01RT090789 (0)
01RV072889 (0)
Change session
1. Audio tape (analog)
Tape No. 30-31
Backup BT 23-24
Recording comments
Media file Choose File no file selected
Transcribers
1. Transcriber Tania Cornejo
Complete?
Hard copy?
Date 1999-06-15
Format
Transcription comments aproximate date
Add transcriber
Reliability Checkers
1. Reliability checker Gustavo Figueroa
Complete?
Hard copy?
Date 2002-12-01
Format
Rel. check. comments
2. Reliability checker Cecilia Rossel
Complete?
Hard copy?
Date 2003-03-24
Format
Rel. check. comments
3. Reliability checker María Blume
Complete?
Hard copy?


Date 2003-04-01
Format
Rel. check. comments Checked with Cecilia Rossel
Add reliability checker
2. Hi-8
Tape No. 9-10
Backup BV 8-9
Recording comments Backup is VHS
Media file Choose File no file selected
Transcribers
1. Transcriber Tania Cornejo
Complete?
Hard copy?
Date 1999-06-15
Format
Transcription comments Same as audio transcription, aproximate date
Add transcriber
Reliability Checkers
1. Reliability checker Gustavo Figueroa
Complete?
Hard copy?
Date 2002-12-01
Format
Rel. check. comments Same as audio transcription
2. Reliability checker Cecilia Rossel
Complete?
Hard copy?
Date 2003-03-24
Format
Rel. check. comments Same as audio transcription
3. Reliability checker María Blume
Complete?
Hard copy?
Date 2003-04-01
Format
Rel. check. comments Checked with Cecilia Rossel, Same as audio tra
Add reliability checker
Add A/V format

Analyses completed

Publications

Comments Elicited Production data is also transcribed as part of this session due to its closeness to natural speech. All Elicited Production utterances are marked with "EP" in the comments field.

Subject Info (specific to session)

Age 02;01;28

School Nido Melody

Number of siblings 1

Level of Education Pre-K

Position among siblings 2

Occupation

Address

How long has subject lived there? 2 years 1 mon

Language Spanish

Language Dominance 1

Add language

Parent / Primary Caretaker Info

	Caretaker	Occupation	Level of Education
1.	Mother	Elementary school teach	College graduate
2.	Father	Engineer	College graduate

Add caretaker