

# INTRODUCTION TO THE VIRTUAL LINGUISTICS LABORATORY

(VLL)(5/15/05)

---

The Virtual Linguistics Laboratory™ is a unique collection of linguistic data, data collection tools and analysis tools. Dr. Barbara Lust and her students developed the VLL through over 20 dedicated years of studying language acquisition in the context of cognitive science. These tools are now being converted to Internet form. Working with collaborators across the world, they are now being expanded and converted to widely accessible form.

## 1.Primary Research Tools

The foundations for the VLL involve hardware and software, as well as linguistic and conceptual resources. All of these are intended to support primary research.

These materials include:

1. A general Manual, Cornell University Virtual Linguistics Lab (VLL) Research Methods Manual: Scientific Methods for Study of Language Acquisition (Lust, Blume and Ogden, in prep), which describes basic procedures for training in child language acquisition research and includes numerous modules regarding several different experimental tasks and methods used in the lab to assess children's knowledge through their: a) language production and b) language comprehension, as well as guidelines for studying children's speech (natural and experimental). It

provides the researcher with guidance in the actual process of collecting and transcribing data in a scientific manner.

2. A Data Transcription and Analysis (DTA) software tool which is under development for structuring scientific research on cross-linguistic speech (both natural speech and experimentally derived speech) and an accompanying DTA manual. The DTA tool structures the researcher in the process of recording language data and converting language to scientific data. It leads the researcher through a systematic set of metadata fields which identify and authenticate the language data. It automatically generates a powerful relational database that lets users conduct rigorous linguistic studies by breaking languages down into their components.

3. A cross-linguistic relational database involving more than 20 languages.

4. A set of web pages which cohere with the survey course in language acquisition taught by Dr. Barbara Lust at Cornell University (HD/Psych/CogSt/Ling 436).

5. A set of audio and audio-visual materials which:

a. demonstrate various methods and tasks being applied in interviewing children.

b. provide examples of the behavioral phenomena being studied (The 5-minute “Doggies and Apples” video module provides an

example for the “overextension” phenomenon in early word learning).

c. provide interactive multimedia learning modules which teach researchers and students how to conduct research using various methods; and which demonstrate various experiments which can be replicated.

d. can be connected to course web page materials.

6. An Experiment Bank which provides all the components of an experiment allowing replication by researchers or students in conjunction with published papers based on the experiment.

7. A set of tools for teaching a hands-on Research Methods course (CogSt 450/Ling 450/HD 437/Psych 437).

## **2. Location**

The CLAL (Cornell University, MVR, G17) is connected with several national and international labs. For a list of labs that collaborate with CLAL through a Virtual Center for the study of Language Acquisition, see [www.clal.cornell.edu/vcla/](http://www.clal.cornell.edu/vcla/) .

## **3. Hardware**

The hardware at Cornell University Upper Campus is tied into the Cornell Language Acquisition Laboratory and consists of both computers and AV equipment as well as digitizing and CD/DVD processing equipment. A central UNIX-based server connects the CLAL to several other labs across campus

(e.g., Uris #201 in the Cognitive Studies program which allows small group lab courses) and across the world.

Many of the local resources are derived from a NSF ILI grant to Professors Lust, Jongman, and Lantolf for undergraduate education.

#### **4. Software**

Software includes network software for these computers and the proprietary DTA tool.

#### **5. Cross-Linguistic Database**

The VLL contains a unique corpus of language data collected from many languages and many countries over 20 years. The uniqueness of these data is that, unlike most data available for linguistic analysis, they were collected and processed through standardized procedures in order to allow comparability across subjects, sessions, and languages. Much linguistic data cannot support quantitative analyses across data sets because each data source is different in collection and transcription, methods and analysis. In this case, data integrity issues threaten hypothesis testing.

The VLL is also unique in that it has multimedia forms of language data so a researcher can at least partially participate in the primary data collection experiences through calling up these data. Subject data are captured on audio, and in some cases, video; there are paper records supporting and describing the

data collection, transcription, and analysis processes. Precise records for transcription and scoring are available.

The following is a list of languages that have multimedia information collected and catalogued by Dr. Barbara C. Lust, and/or her students, of Cornell University, as well as by collaborators across the world, to support analysis and hypothesis testing for natural language studies. They are the foundation for numerous Ph.D. dissertations, Master's Theses, Honors Theses, as well as independent research (See "Publications" in the Cornell Language Acquisition Lab Manual and in the Cornell Language Acquisition Lab web page: [www.clal.cornell.edu](http://www.clal.cornell.edu)).

Through the Virtual Center, numerous languages are under study. They are being analyzed for their linguistic structure, as well as for their first and second (or multi) language acquisition.

**Table I-1: Cornell Language Acquisition Lab Cross-Linguistic Database**

<u>CLAL Cross-Linguistic Database</u>	
• Arabic	• Japanese
• Basque	• Javanese
• Chinese	• Korean
• Dutch	• Malayalam
	• Oriya
• English	• Sinhalese
• French	• Spanish
• German	• Swedish
• Hindi	• Taiwanese
• Hebrew	
• Inuktitut	• Tulu

