

## Language Documentation: Theory, Practice, and Values

Harvard University  
Emerson Hall  
Cambridge, MA

9-11 July 2005

Impelled by concerns about the accelerating loss of languages, language archiving has emerged as a prominent issue for linguists and for heritage language speakers alike. The quality of documentation available for an endangered language can determine the success of its revitalization. This conference builds on the LSA Conversation on Endangered Languages Archiving which met 5-6 January 2005 immediately before the LSA Annual Meeting. Six themes will be treated in conventional paper presentation sessions, in poster sessions, and in panel presentations with open discussion.

- The requirements of field linguistic training
- The concerns and involvement of the heritage language communities
- The question of what is adequate documentation
- The uses of documentation in speaker communities
- Training and careers in field linguistics
- Ethics and archiving best practices

### Acknowledgements

We thank the following organizations and individuals for their support of this conference:

National Science Foundation: Joan Maling and James Herbert

Harvard University: Jay Jasanoff, C-T James Huang, Cheryl Miller, and Kobey Shwayder

MIT: Sabine Iatridou, Mary Grenham, and Suzanne Flynn

This **conference** was funded by an NSF Grant .

## Public Sessions

**Day 1: Saturday, 9 July 2005**

### Registration

Time: 7:30 AM – 12:00 noon

Room: Emerson Hall (hallway in front of Room 210)

Cost: \$25.00 (cash or check only)

### Program

*All sessions will be in Emerson Hall Room 210, unless otherwise indicated.*

## Conference Opening and Welcome

Time: 8:00 AM

Convener: N. Louanna Furbee (U MO-Columbia / LSA Archivist)

### Session I Training for Language Documentation

Chair: Peter Austin (School of Oriental Studies [SOAS], London)

8:15 AM Keynote Speaker: Peter Austin (SOAS)

*Training in Language Documentation: The SOAS Experience*

8:50 AM Carol Genetti (UC-Santa Barbara)

*Components of a Summer Institute in Field Linguistics*

We present a model for a biennial Summer Institute in Field Linguistics to be held at UC-Santa Barbara beginning in 2008. The institute will consist of two independent phases: a six-week phase containing two courses in field methods, and a shorter phase containing workshops on practical aspects of fieldwork. Topics could include recording techniques, discourse transcription, digital archiving, software and fonts, intercultural and inter-personal issues, and maintaining health and well-being during fieldwork. We seek feedback from the broader linguistics community so that we can mold the institute to best meet the needs of the field.

9:15 AM Frances Ajo, Valerie Guerin, Ryoko Hattori, & Laura Robinson, (U HI-Manoa)

*Native Speakers as Documenters: A Student Initiative at U HI-Manoa*

9:40 – 10:00 AM BREAK

### Session II What Is Adequate Documentation?

Chair: K. David Harrison (Swarthmore C / Chair, CELP)

10:00 AM Keynote Speaker: K. David Harrison (Swarthmore C / Chair, CELP)

*Ethnographic Dimensions of Documentation*

10:40 AM Laura Buszard-Welcher (Rosetta Project / UC-Berkeley)

*Necessary and Sufficient Data Collection:  
Lessons from Potawatomi Legacy Documentation*

Our modern concept of necessary, or 'core' language documentation remains much the same as it has for the last 75 years: a grammar, dictionary, and text corpus. While this set is certainly necessary, it is demonstrably not sufficient. The collection of Potawatomi legacy documentation

by Charles Hockett is a good illustration: His discussion of morphosyntax, while extensive, was based entirely on what he found in collected narratives, and it turns out that conversational morphosyntax is quite different. This example raises the question of whether our own concept of language documentation will be viewed as comprehensive by following generations of researchers.

11:00 AM Verónica Vázquez Soto (UNAM, México, DF)

*Documenting Different Genres of Oral Narrative in Cora (Uto-Aztecan)*

We focus on documenting a broad range of oral genres in Cora (Uto-Aztecan language spoken in Northwest Mexico in the State of Nayarit) oral narrative. Adequate documentation of oral narrative in this language should cover different genres which go from the 'word for playing' or 'toy word' called in Cora *niaukari hi'wahkari* to the 'ancient word' called in the language *niukari mime'ekan*. The former genre is usually told for children and covers funny stories about particular animals whereas the latter genre is properly told for adults, although children can be part of the audience for this narrative. The ancient word covers stories about the creation of the Cora universe, the creation of corn and other important entities of Cora culture. We hope that this project would give rise to a more adequate documentation of one Cora dialect from which we can obtain an extensive collection of texts and a more precise and exhaustive dictionary, as well as better studies of the grammar, such as word order, evidentials, and other relevant grammatical and pragmatic subjects. We will show some fragments of texts as well as some fragments of videotapes to illustrate the different styles of the narrators telling a story.

11:20 AM Discussion (led by K. David Harrison)

11:40 AM – 12:40 PM LUNCH BREAK

**Session III Models of Successful Collaboration**

Chairs: Akira Yamamoto (U KS/Past Chair, CELP)  
Arienne Dwyer (U KS/DoBeS, Germany)

12:40 PM Keynote Speaker: Martha Macri (UC-Davis/Terralingua)

*Working with Language Communities in Unarchiving:  
Making J. P. Harrington Notes Accessible*

The J. P. Harrington Database Project consists principally in transcribing and coding the linguistic and ethnographic notes on American Indian languages collected by Harrington during the first half of the 20th century. The men and women he interviewed were often among the last remaining speakers of their language--the notes a treasure of indigenous knowledge that otherwise would have been lost. Perhaps of greatest value to Native American community scholars in cultural and language revitalization, the project has been especially successful in combining efforts of both the academic community and native scholars.

1:10 PM Philip Cash Cash (U AZ/Nez Perce Tribe)  
Susan Penfield (U AZ)

### *Technology-Enhanced Language Revitalization*

We present a progress report on recent efforts to establish meaningful links between technology-based language instruction and the field of language revitalization. Through a collaborative grant between the Colorado River Indian Tribes (CRIT) and the University of Arizona, a project was implemented to train CRIT speakers of Mohave and Chemehuevi to use software which supports preservation and pedagogy for these languages and to offer this training episode as a model, disseminating information about technology-enhanced language revitalization at the University of Arizona's American Indian Language Development Institute (AILDI). The results are a growing roster of multimedia-based courses at AILDI.

1:30 PM      Arienne Dwyer (U KS/DoBeS, Germany)

#### *An Appropriate Collaborative Model*

Since local circumstances vary widely, many different models of successful collaboration exist. This short contribution instead explores the commonalities of productive collaborative research. Ingredients including fully consultative planning, clear goal-setting, sufficient training, a smooth workflow, and developing autonomous indigenous capacities. Illustrative examples from several projects are provided.

1:50 PM      Discussion (led by Arienne Dwyer)

2:20 - 2:40 PM      BREAK

### **Session IV    Uses of Documentation in Speaker Communities: Three Case Studies**

Chair: Nicholas Ostler (Foundation for Endangered Languages)

2:40 PM      Dennis Kartammeru O'Brien (U Adelaide, Australia)

*Waking Up an Ancient Language:  
Reclaiming Kurna of the Adelaide Plains, S Australia*

3:00 PM      Faizi Inayatullah (NWFP, Pakistan) **WITHDRAWN**

*Technical Document for Semiliterate Speakers:  
Success Story of Kalasha Language, Northern Pakistan*

Kalasha is an endangered language of the Indo-European family (Indo-Aryan subfamily) in the Kafir group of Dardic languages. The Kalasha minorities live in the Chitral district of NWFP, in the Hindu Kush mountain range on the border of Afghanistan in Northern Pakistan. The Kalash were once the dominant people in the area. The Kalash culture as well as the Kalasha language is highly endangered at the wake of forceful conversion. In the case of Kalasha, highly technical reports have been helpful for the semiliterate speaker community in revitalizing the language and promoting 'linguistic activism' among the younger generation.

3:20 PM Colleen M. Fitzgerald (TX Tech U)

*Language Documentation in the Tohono O'odham Community*

We examine documentation efforts on the Tohono O'odham (or Papago) Reservation in Arizona, focusing on the legacy recordings held in archives or by individual researchers which are inaccessible to community members. These efforts reflect input from community members about community concerns, educational needs, future goals for the Nation, language revitalization, dialect education, and accessibility. These recordings present ethical considerations dealing with 'ownership' and intellectual property, and the treatment of culturally-sensitive collections. A long-term project to transcribe, translate, and publish these O'odham materials has implications beyond linguistics, raising significant questions about who determines the benefit and impact to indigenous communities.

3:40 PM Discussion (led by Nicholas Ostler)

**Poster Session**

Room: Emerson Hall 2nd Floor Hallway

Time: 4:10 - 5:30 PM

Chair: Lenore A. Grenoble (Dartmouth C/Past Chair, CELP) & Theme Chairs

**Theme I: Training for Language Documentation**

Chair: Peter Austin (SOAS)

Andrew Garrett (UC-Berkeley)

*Fieldwork as Philology, or, the Boasian Revolution in Linguistics"*

We present an example or two, from California, of linguistic fieldwork methodology stressing (1) transparency, in the form of documentation that is fully and immediately accessible to the language community, and (2) integration of both modern and archival sources and of structural-grammatical information and textual-sociocultural information.

David Leedom Shaul (Venito Garcia Lib & Archives, Tohono O'odham Nation)

Danny Lopez

*Living Latins:*

*The Role of Literature (Written Broadly) in Language Revitalization*

We examine the role of literature (oral and written) in different processes of language revitalization. The role of traditional as well as metropolitan genres are examined as well as the role of audience/participant in genres. The role of tribal libraries and other cultural institutions are also examined.

**Theme II: What Is Adequate Documentation?**

Chair: K. David Harrison (Swarthmore C/Chair, CELP)

Susan Meredith Burt (IL SU)

*Documentation of Pragmatics and Metapragmatics:  
Language Shift and Pragmatic Change in Hmong Language in Wisconsin*

The pragmatics of minority languages are vulnerable to change early in language shift and should, therefore, be included in language documentation. To research language shift and pragmatic change, 30 speakers of Hmong were interviewed in Wisconsin, using an oral Discourse Completion Task. Responses show that influence from Anglo-American pragmatics has affected younger speakers' usage in Hmong; in particular, the frequency of *please* has influenced frequency of the Hmong equivalent *thov*, causing younger and older speakers to differ in metapragmatic comments on usage. Pragmatics and metapragmatics matter because ordinary speakers, who construct their social worlds through linguistic interaction, may treasure tools that allow them to do so—pragmatic resources of their language.

Barbara Lust (Cornell U), Elaine Westbrook (Cornell U), Suzanne Flynn (MIT), & Theresa Tobin (MIT)

*Developing Adequate Documentation for Multifaceted Cross-Linguistic Language Acquisition Data*

We describe collaborative work in which we seek to establish best practices for documenting large, continually expanding amounts of language data of various types. Existing multimedia data currently involve thousands of samples of language at various periods of language acquisition (both in child and adult), in various situations (both naturalistic and experimental), and across more than 20 different languages in more than 20 countries. By linking university researchers with university library experts, we seek a documentation system for this and for future data which at once (1) links the data to domain-specific linguistic analyses which are necessary for research; (2) attempts to calibrate across various languages in doing so; (3) links to field-wide standards for this description, e.g. such as those being developed by Emeld (Electronic Metastructure for Endangered Languages Data) now; (4) links to field-wide standards for linking such resources to others in the field, e.g. such as OLAC (Open Language Archiving Community) is developing now; and (5) links to the crucial upper level documentation system of a university library and interlibrary domain, one which, through metadata systems and general web-based ontologies, situates language data in a general knowledge domain and renders the data accessible to general library users. We report on our program, progress, and problems in this endeavor.

Jess Tauber (Oakland, NJ)  
*Enough Is Enough?: The Case for Yahgan*

**Theme IV: Uses of Documentation in Speaker Communities**  
Chair: Nicholas Ostler (Foundation for Endangered Langs)

Hermelindo Aguilar Méndez (CISC, Comitán, México), Teresa López Méndez (CISC, Comitán, México), Juan Méndez Vasquez (CISC, Comitán, México), Maria Bertha Santiz Perez (CISC, Comitán, México), Ramon Jiménez Jiménez (CISC, Comitán, México), Louanna del Socorro Guillén Roveló (U Valle de México, Tuxtla Gutierrez, Chiapas, México), N. Louanna Furbee (U MO-Columbia/CISC, Comitán, Chiapas, México/ECOSUR, San Cristobal, Chiapas, México), Robert A. Benfer (U MO-Columbia), & Rolando Tinoco Ojanguren (CISC, Comitán/ECOSUR, San Cristobal, Chiapas, México)

*Saving Languages, Saving Lives:  
Tojolabal Maya Language Revival within a Health Research NGO*

The Centro de Investigaciones en Salud de Comitán (CISC) engages in epidemiological research and health intervention in southern Chiapas, Mexico, where the predominant indigenous language is Tojolabal Maya. With 50,000 speakers but a shrinking footprint, Tojolabal is in early endangerment. CISC offers a model for other non-linguistic enterprises interested in language renewal. Partnering with the Tojolabal-speaking community, it has conducted collaborative research into bilingual interviewing, published literary works in Tojolabal and Spanish, and sponsored an international gathering of Tojolabal scholars. Its work increases Tojolabal use by adding intellectual discourse as an appropriate context for the language and encourages documentation centered on sets of coherent subject matters.

Tjeerd de Graff (Frysk Acad, Ljouwert)  
Presenter: Nicholas Ostler (Foundation for Endangered Langs)

*Siberia's Sound Archives:  
From the Early 20th Century to Today's Research and Tomorrow's Communities*

**Theme V: Ethics and Archiving Best Practices**

**Chairs:** Alice Harris (U Stony Brook-SUNY) and Martha Macri (UC Davis/Terralingua)

David Golumbia (U VA)

*Representing Minority Languages and Cultures on the World Wide Web*

This poster offers an evaluation of websites describing minority languages from around the world with an eye toward developing best-practices guidelines for their construction. We argue that the most effective websites are not necessarily the most comprehensive in terms of the specifically linguistic information they include (dictionaries, grammatical description, text archives, etc.) since these can give native speakers a feeling of estrangement from their own language, while also creating the impression that the language is a static artifact. The best websites strike a balance between linguistic representation and a sense of social exchange with the larger world.

Comfort Pratt (TX Tech U)

*Death of a Dialect: The Case of Adaeseño Spanish*

Nick Thieberger (U Melbourne, Australia)  
Presenter: David Nash (Australian Ntl U)

*Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC)*

The Pacific And Regional Archive for Digital Sources in Endangered Cultures (PARADISEC) has been operating since mid-2003 based in a consortium of four universities in Australia. In that time, we have located and digitized some 850 hours of field recordings, with a further 3,000 hours in the queue. We have established a methodology for description of this material and are

experimenting with linking fieldnotes and textual material with media. We are particularly involved in training and advocacy for language documentation methods that result in good archival data and in repatriating these recordings once digitized. This poster covers the main issues dealt with by the project, which is the result of a community of interested linguists and musicologists working collaboratively in a cross-institutional partnership.

Helen Aristar-Dry (LINGUIST List/E MI U), Anthony Aristar (LINGUIST List/Wayne SU), Naomi Fox (LINGUIST List/Wayne SU), Susan Hooyenga (LINGUIST List/EMI U), Steve Moran (LINGUIST List/E MI U), Megan Zdrojkowski (LINGUIST List/E MI U)

*Creating Language Resources That Last:  
The E-MELD School of Best Practices in Digital Language Documentation*

A growing number of projects focus on documenting endangered languages before they disappear, but it is also important to be aware of the need to preserve the documentation itself. Paper, audiotapes, videotapes, and computer storage media are all prone to degradation and destruction. Moreover, the proliferation of digital standards and formats inhibits the interoperability and reusability of language resources. The E-MELD School of Best Practices (<http://emeld.org/school/>) is designed to teach linguists how to create language documentation that can be accessed and used by the widest number of current speakers and researchers, as well as by future generations.

*(Refereed posters will be on display from 8:00 AM - 5:30 PM.)*

## **Day 2: Sunday, 10 July 2005**

### **Session V Ethics and Archiving Best Practices**

Chairs: Alice Harris, (U Stony Brook-SUNY)  
Martha Macri (UC-Davis/Terralingua)

Moderator: Alice Harris

8:00 AM Martha Macri (UC-Davis/Terralingua)

*The Linguist's Responsibilities to the Community of Linguists*

8:20 AM Keren Rice (U Toronto)

*The Linguist's Responsibilities to the Community of Speakers*

The past 30 years or so have seen dramatic shifts in what is regarded as the responsibilities of linguists to the communities with which they work. We provide an overview of some of the types of changes that have occurred in what is considered to be responsible fieldwork, focusing on work with Aboriginal communities in Canada. A model has developed that involves taking seriously a number of key points, particularly working with the community to determine what should be researched and how, and carrying out the research in a respectful way.



8:40 AM Michael Krauss (U AK-Fairbanks)

*Can Linguistics Be Ethical?*

Linguistics can be ethical, during the Language Holocaust now under way, by documenting disappearing languages in the field so as also to help them not disappear. Rational coordinated prioritization: along with social factors, e.g. also for languages with fewest speakers, most isolated genetically or typologically, and of course least documented. Responsibility for previous or related work, basic competence, e.g. observational adequacy, sure and accessible archiving. Mutuality of understanding with speakers and community, training of speakers, and strengthening of language status in the process.

9:00 AM Lise M. Dobrin (U VA)

*When Their Values Conflict with Ours:  
Linguists and Community Empowerment in Melanesia*

In Melanesia, where hundreds of languages are increasingly endangered, relationships with outsiders are highly desired, and they are empowering for communities because they reflect positively on their value and enable them to galvanize the cooperation and unity necessary for community-level projects to succeed. So even though a language documentation project might appear to us to have been successfully completed, it may be perceived by the people themselves as having failed when the outside linguists withdraw. To understand our role as merely developing resources like grammars or documentation skills is to risk disappointing and even disempowering the Melanesian people whose languages are at stake.

9:20 AM Discussion (led by Alice Harris)

9:40 - 10:00 AM BREAK

**Session VI: Training and Careers in Field Linguistics**

Chair: Martha Ratliff (Wayne SU/Past Chair, CELP)

10:00 AM Keynote Speaker: Paul Newman (IN U)

*Field Methods Courses at American Universities:  
The Institutional (Non)Response to the Endangered Languages Movement*

This paper presents the results of a recent survey of field methods courses in America. Surprisingly, the situation is little changed from what it was a dozen years ago before language documentation came into vogue. It is suggested that whatever value field methods courses may have, that value does not lie in preparing students for actual field research. Some reasons for this are: (1) Students in the courses are inadequately trained. (2) Such courses have to cater to "casual takers"; (3) their length and intensity are insufficient, (4) the courses bear little connection to the reality of fieldwork.

10:40 AM Spike Gildea (U OR)

*Field Linguistics Training at the University of Oregon*

11:00 AM Mary Laughren (U QLD, Australia)  
Jane Simpson (U Sydney, Australia)

*A Survey of Australian Fieldwork Training Programs*

We discuss fieldwork training in Australia, including internships, field methods courses, and computer skills workshops. We discuss the results of a survey (21 respondents) carried out by the Linguistics Program (Melbourne U) in 2004/2005 for the LSA Committee for Endangered Language Preservation on training students to carry out linguistic fieldwork in Australian and New Zealand universities. (Collated Cathy Bow, Australian Linguistic Society Newsletter, Feb. 2005; <http://www.als.asn.au>). We talk about the role of Aboriginal Language Research Centres in Australia as a focus for fieldwork and intermediaries between university linguistics departments and Aboriginal communities. We discuss career opportunities for field linguists..

11:20 AM Discussion (led by Martha Ratliff)

**Scribes' Summary Report**

Time: 11:45 AM –1:00 PM

Doug Whalen (Haskins Labs/Endangered Lang Fund)  
Jeff Good (Max Planck Inst-Leipzig/OLAC)

## **Closed Session**

**Day 3: Monday, 11 July 2005**

**Session VII: Extending the LSA Conversation on Archiving Endangered Languages**

Chair: N. Louanna Furbee (U M)-Columbia/LSA Archivist)

Participants: LSA Conversation Participants from Oakland Meeting & Other Invitees