



## **NSF Grant Puts Endangered Languages on the Web**

### **EMELD Project Makes Language Material Accessible**

The National Science Foundation of the United States awarded a grant last year to help put endangered language material on the web. This five year, xx million dollar project is led by Helen Aristar Dry (Eastern Michigan University) and Anthony Aristar (Wayne State University). Most linguists are quite familiar with these names, since Helen and Anthony also created and edit the Linguist List, an email distribution list serving xx linguists worldwide. They are leading a group of research units, including the Endangered Language Fund, the University of Pennsylvania, the Summer Institute of Linguistics and the University of Arizona, in a massive project that will have far-reaching implications for linguistics and for the maintenance of threatened languages. In this issue of the newsletter, we wanted to bring you up to date on the grant, especially the ELF part of it.

The project is entitled EMELD, Electronic Metadata for Endangered Languages Data. (You can see its website at <http://linguistlist.org/xxxxxxx>.) The ultimate goal is not just to put some sound files up on the web, but to have linguistic texts marked up in such a way that they can really be made use of by people with all sorts of different ideas about what they can mean. Just putting a text up does not allow someone to “query” it and see whether it contains any third person plural verbs, for example. If you are trying to see how

*See EMELD on page 2*

### **ELF and Rosetta Team Up for Language Query Room**

One component of the ELF’s contribution to EMELD (see accompanying story) is the Language Query Room (LQR). This is an area designed to allow linguists and second-language-learners to ask questions about how to say something in an endangered language. Experts in these languages, typically native speakers, can then post responses. While many endangered language communities do not have internet access, many do, and the LQR is designed to make their languages more prominent on the web.

The implementation of the LQR has been done largely by the Rosetta Project, headed by Jim Mason. (Visit their website at [www.rosettaproject.org](http://www.rosettaproject.org). A recent “Wired” article about the project can be found at <http://xxxxxxxxxxxxxxxxxxxxxxxxxxxx>.) The Rosetta Project’s main goal is to compile a thousand language archive which it will etch onto nickel

*See LQR on page 3*

---

---

### **ELF Annual Meeting to be held in Atlanta, January 2003**

The annual meeting of the Endangered Language Fund will be held in conjunction with the Linguistic Society of America meeting in Atlanta. The date is Friday, 3 January 2003. The meeting will run from 8 until 9 in the morning in the Cherokee Room of the Whatever Hotel.

Anyone interested in the workings of the Fund is invited to attend. We will have updates on our activities and some reports from previous grantees. See you there!

---

---

## Our Board of Directors:

**Douglas H. Whalen**, President/Founder  
Linguist, Haskins Laboratories, New Haven, CT

**Stephen R. Anderson**  
Linguist, Yale University, New Haven, CT

**Rebecca Bending**  
Yakama/Nez Perce, Toppenish, WA

**C. J. Cherryh**  
Author and Classicist, Spokane, WA

**Noam Chomsky**  
Linguist, Massachusetts Institute of Technology,  
Cambridge, MA

**Durbin Feeling**  
Linguist, Oklahoma University, Norman, OK

**Gay Story Hamilton**  
Mohegan Council of Elders, Uncasville, CT

**Dennis Holt**, Secretary/Treasurer  
Linguist, University of Montana, Missoula, MT

**Peter Ladefoged**  
Linguist, University of California, Los Angeles, CA

**Johanna Nichols**  
Linguist, University of California, Berkeley, CA

**Karl V. Teeter**  
Linguist, Harvard University, Cambridge, MA

---

---

ELFN:

Matthew Richardson, Editor  
Staff Researcher  
Haskins Laboratories, New Haven, CT



---

ELFN: Volume 6, No. 3, December 2002  
Printed in New Haven, CT  
©2002, Endangered Language Fund, Inc.  
***The Endangered Language Fund***  
Department of Linguistics  
Yale University  
P.O. Box 208366  
New Haven, CT 06520-8366 USA

e-mail: [elf@haskins.yale.edu](mailto:elf@haskins.yale.edu)  
www: <http://www.ling.yale.edu/~elf>

*EMELD - Continued from page 1*

different languages treat the same linguistic category, it would be very useful to be able to ask questions like this and get a rich set of results,

Finding material about a language begins with the language name, and that has been a large project for EMELD. Thanks to the generosity of SIL International (the Summer Institute of Linguistics), especially through the agency of Gary Simons, the massive amount of work that went into generating the language codes for the Ethnologue has been made available to EMELD, the Rosetta Project and, to some extent, the ISO standards committee on language names. Having no equal in world-wide coverage, the Ethnologue has become the de facto standard for the field. Now it is well on its way to becoming an official standard as well. This will save a great deal of duplication of effort in the naming department, and it should also allow for a more systematic way of addressing theoretical disputes about the status of languages, dialects and genetic relationships. This implementation alone will make the EMELD project a success.

The next aspect is standardizing the mark-up language for linguistic material. This is nothing short of creating an ontology of linguistics--defining what the terms are and what their relationship to each other is. While this kind of work is difficult and requires a great deal of interaction among specialists, the pay-off is huge. Having a consistent way of treating language categories will allow the fullest possible comparison of language material across languages and families.

*EMELD - Continued on page 3*

---

---

## **New ELF Office**

The Yale Linguistics Department has moved to a newly renovated building, and the Endangered Language Fund has moved with it. Our address has changed slightly--the new P.O. Box number is 208366, and the zip-plus-four is 06520-8366. Our street address is now 370 Temple Street, New Haven, CT 06511. Please use these addresses from now on.

We are in the process of listing our library collection, and we hope to have that list out soon. The library is open to all interested people by appointment. Please drop us a line (at the new address or the usual email address, [elf@haskins.yale.edu](mailto:elf@haskins.yale.edu)) if you would like to see the collection.

---

---

disks which will last for 2,000 years or more. The language data there will potentially serve as a key to otherwise unknown languages in the future; in the present, it has crystalized our thinking about the longevity of the efforts we are making in archiving language material.

In the course of performing their main task, the Rosetta team has concentrated a great deal of computer expertise in the domain of language implementation on the web. Rather than try to duplicate that expertise elsewhere, the ELF contracted with Rosetta to do the programming for the LQR. Significant design features and usage considerations were contributed by Mason, and the result is promising to be a tool that is truly useful for linguists, native speakers, and heritage language learners.

One of the most tiresome problems for using minority languages on the web is the lack of a good native orthography interface. Although the Unicode initiative has brought us a long way toward having these scripts usable, they currently rely on the use of particular fonts which not everyone has access to. (See their site at [www.unicode.org](http://www.unicode.org).) What Rosetta has implemented is a pop-up keyboard that will allow relatively easy input of a wide range of orthographies. These include Cherokee, xxx, xxxx and xxx, but also include such prominent interlanguages as Russian, Arabic and Thai (?). (Chinese is currently not supported, even though there is a large Chinese component to Unicode--it is just too large of a problem for our current solution.) Even with languages like Maori or Navajo that have only a couple of unusual characters (the macron over vowels for Maori and the nasalization diacritic for Navajo, for example), having an easy way to type in just those characters can make all the difference between the "right" look and something that's just not quite there.

Another feature that we have implemented is the posting of sound files. This opens up much broader horizons for allowing the serious learning of endangered languages. Every writing system (including the International Phonetic Alphabet, or IPA) is just a sketch of what the sound is like. Ac-

tually hearing the utterance as produced is like the difference between reading the score of a symphony and hearing it performed. The nuances are all there, and the outright misinterpretations that writing is subject to are gone. This also gives an easy way for different speakers to say how a sentence would be in their dialect, allowing for quick comparisons of the variation that we find in any language.

A panel of native speakers has agreed to lend their insights via the Language Query Room Advisory Board. This group is headed by Durbin Feeling of the University of Oklahoma. Durbin has been on the ELF Board of directors since the creation of the ELF. The Advisory Board is spread across the world, and so it meets only via email. The target audience for the LQR is those groups that can make use of the internet, so this is an appropriate way of meeting. Our hope is that we will be able to change the LQR to meet the needs of the native communities and the linguistic community, furthering language survival in a variety of forms.

---

---

*EMELD - Continued from page 2*

Making the file format for text and sound consistent and interpretable will be another great advance. This will take many years, because the file format that makes sense will depend on developments in the more general audio and computing worlds. The datasets need to be predictable, though, so that anyone with a reasonably modern browser throughout the world can make use of the material available. Inevitably, this format will also have to be easily changed into a different format, since the likelihood of any format surviving for a hundred years (or even five) is rather small. What is important is to be able to easily migrate material as the need arises.

Finally, the EMELD project is dedicated to establishing the Showroom of Best Practice. Here, anyone interested in archiving language material can see programs and techniques for any of the inputting, editing and mark-up issues that will arise with any project. The hope is to have open source programs that will be available as well, so that the actual work can begin at the Showroom. Even having references to programs available elsewhere will be a benefit.

The EMELD project is constantly expanding. We recommend that you check the site often, to see additions and improvements. There are papers from EMELD conferences there, and links to other relevant sites. Please feel free to drop a line to EMELD to give your reactions ([EMELD@LinguistList.org](mailto:EMELD@LinguistList.org)).

*The Endangered Language Fund*  
Department of Linguistics  
Yale University  
P.O. Box 208366  
New Haven, CT 06520-8366  
USA  
e-mail: elf@haskins.yale.edu  
www: <http://www.ling.yale.edu/~elf>

## Updates on some ELF grant recipients

**Delphine Red Shirt** (Lakota), who received an ELF grant in 2000, finished the book that the grant supported. *Turtle Lung Woman's Granddaughter* was published by the University of Nebraska Press, and has already gone into its second printing. Delphine has had readings and book signing at R. J. Julia's (Madison, CT) and the Yale Bookstore in New Haven. Some of the recordings of Delphine's mother (which formed the basis of the book) will be made accessible on the internet as a demonstration project under the EMELD grant (see page 1). Delphine is currently working on her third book, a novel about her great great grandfather's experiences in the Wild West shows of the 1800's.

**Valerio Luciani Ascensio** (Kawki) has already begun producing the books that this year's ELF grant supported (see the previous newsletter).

By virtue of receiving this grant (and the external world's recognition that it represents), the Peruvian government is now implementing the bilingual program directly. Thanks go to M. J. Hardman for her sponsoring this grant and helping it have this large impact.

**Carolyn MacKay and Frank R. Trechsel** (Ball State University, in Indiana) received a grant in 1999 to further their work on the documentation of Pisaflores Tepehua in Mexico. They were able to take this experience as a basis for a grant from the Foundation for the Advancement of Mesoamerican Studies, which they are now working on.

We at the ELF are extremely proud that our small grants have been leveraged so successfully. We are striving to increase our funding sources so that even more such grants can be made.