

Global mission races to save dying languages

Media Release, Tuesday 14 January 2003

The Yemba people live in the western province of Cameroon, Africa, near the town of Dschang. Associate Professor Bird conducted extensive research on the tone system of this language, and proposed a new way to write the language which was simpler for people to read and write. He also helped produce the first ever dictionary for the language. . . .

A University of Melbourne computer scientist has embarked on a global mission to save the thousands of endangered languages in the world from disappearing forever.

Linguists estimate there are over 6000 languages spoken worldwide. They also say that 90 percent of these are likely to die before the end of this century, and 440 are within only a generation or two of extinction. Most of these languages are not written down and when they disappear, any record of these languages will also disappear along with the unique oral histories and cultural heritage.

On a mission to avert this tragedy is Associate Professor of Computer Science and Software Engineering, Steven Bird and a team of international colleagues. Together they have launched a worldwide project to design and build an open language archives community (OLAC). OLAC is an international network of 25 digital archives in six countries that already holds over 30,000 items in the form of dictionaries, grammars, field notes, text collections and recordings of the world's languages.

"We are sitting between the onset of the digital era and the mass extinction of the world's languages. The window of opportunity to record these endangered languages is small and shutting fast," says Bird.

"Unfortunately, digital preservation faces huge obstacles. Most digital media and storage formats are obsolete within five years and data must be continually transferred over to new formats," he says.

"A lot of linguistic data is stored in Microsoft formats and secret binary formats that cannot be opened (even by Microsoft software) just five years later.

"Linguists also use a range of tools including word processing documents, spreadsheets, different forms of digital sound and video capture technology and end up in a terrible mess when they try to collate their data with other linguists' research because of the incompatibility or

The Voice

Cover Story

- [Not a drop to spare](#) - /news/5766/?menuLink=/uninews.php
- [Building smarter](#) - /news/5767/?menuLink=/uninews.php
- [Wonder wall](#) - /news/5782/?menuLink=/uninews.php

News

- [The world is an amazing place](#) - /news/5771/?menuLink=/uninews.php
- [The virtual vet](#) - /news/5769/?menuLink=/uninews.php
- [Physio Challenge](#) - /news/5768/?menuLink=/uninews.php
- [The getting of wisdom](#) - /news/5770/?menuLink=/uninews.php
- [Teaching Good Value](#) - /news/5772/?menuLink=/uninews.php
- [From the Vice-Chancellor: A very Australian model](#) - /news/5774/?menuLink=/uninews.php

Reviews and Previews

- [Performance Art/Video](#) - /news/5777/?menuLink=/uninews.php
- [Medieval digital age](#) - /news/5773/?menuLink=/uninews.php
- [Art](#) - /news/5775/?menuLink=/uninews.php

Knowledge Transfer

- [A guiding light](#) - /news/5779/?menuLink=/uninews.php
- [Hidden histories](#) - /news/5780/?menuLink=/uninews.php
- [Learning on the job](#) - /news/5778/?menuLink=/uninews.php
- [Analysing Freud](#) - /news/5781/?menuLink=/uninews.php

Profiles

- [Future food](#) - /news/5783/?menuLink=/uninews.php

Alumni

unsuitability of the various software used."

One of Bird's main tasks with OLAC is to develop a standardised template for researchers to collect linguistic information, and a management system that will ensure the data is accessible to future technologies.

"Without these tools we are simply turning endangered languages into endangered data," he says.

It is not just cultural heritage that dies with a language. Thousands of years of history and a unique point of view on the world are lost along with knowledge of plants, animals and their uses as food and medicines, and their role in a healthy environment.

There are 2000 languages, or a third of the world's linguistic heritage, in the Australasian region that includes SE Asia and the South Pacific. Most are virtually unknown to science.

"It is a living museum of language and culture and Australia is the best placed to preserve this heritage," says Bird.

Australia itself has more critically endangered languages than all other countries combined. There are estimated to be about 80 indigenous Australian languages with only 5 or fewer speakers.

Bird and his colleagues are racing the clock, and while still desperate for funding, Bird has managed to secure \$17 million in grants from the US and UK in the last decade. Late last year, the Universities of Sydney and Melbourne received a \$275,000 Australian Research Council grant to purchase hardware to digitise hundreds of hours of archived recordings of indigenous languages.

"The ARC grant will help us rescue deteriorating tape archives of many endangered and extinct languages from our region making them accessible in perpetuity," says Bird.

The Victorian Partnership for Advanced Computing gave \$30,000 for high-performance computing involving large linguistic databases. The grant will be applied to language technologies of commercial interest including spoken dialogue systems, networked archive systems such as OLAC and the analysis of data from endangered languages.

The Rosetta Project

It was Napoleon, in his march across the Nile region that stumbled across the key to the mysteries of ancient Egypt's hieroglyphic language. The Rosetta stone, a 1,000 year old slab of basalt, had etched into its face three copies of the same text: one in demotic, one in Greek and one in hieroglyphic Egyptian. The stone enabled scholars to at last make sense of the hieroglyphs and unlock the many mysteries of the Egypt's

- [Hope in Liberia](/news/5784) - /news/5784
/?menuLink=/uninews.php

- [One of the Clintons](/news/5785) - /news/5785
/?menuLink=/uninews.php

Sport

- [Rowing 150th](/news/5786) - /news/5786
/?menuLink=/uninews.php

- [Having a ball](/news/5787) - /news/5787
/?menuLink=/uninews.php

What's On

- <http://events.unimelb.edu.au/>
- <http://events.unimelb.edu.au/>

past.

The Rosetta stone has now inspired a separate OLAC project to fashion a modern Rosetta's Stone that will preserve some of the basic knowledge of the world's languages for anthropologists of the future. The team plan to etch into a three inch nickel disk parallel text of the first chapter of Genesis in 1,000 different languages. The technology to read it will require nothing more sophisticated than a 1000_ microscope.

A copy of the disk has been placed on the Ariane space mission departing later this month for the comet P/Wirtanen.

**Images, graphics available

The final version of this dictionary online at: <http://www ldc.upenn.edu/sb/home/publications.html#dictionary>

More information about this article:

Associate Professor Steven Bird,
Dept Computer Science,
University of Melbourne
Ph: 03 8344 1324(W) 03 9894-0111(H)

Jason Major
Media Officer
University of Melbourne
Telephone +(61 3) 8344 0181
Mobile 0421 641 506

See also [Online Experts Guide](http://www.research.unimelb.edu.au/mediacontact/) - <http://www.research.unimelb.edu.au/mediacontact/>