Tools of Dual Utility: Multimedia Applications for Native American Language Preservation and Teaching

Arienne Dwyer, Sue-Ellen Jacobs, and Charles Hiestand

CARTAH, Box 353680, University of Washington, Seattle, WA 98195-3680, USA

KEYWORDS: multimedia, Native-American languages

AFFILIATION: University of Washington

E-MAIL: adwyer@u.washington.edu
sueellen@u.washington.edu
chh@u.washington.edu
FAX NUMBER: (206) 543-9285 (Attn.:CARTAH)
PHONE NUMBER: (206) 543-4218

1. Introduction

1.1 Text, Multimedia, or Both?

As hyperliterate academics, it is easy for us to forget that the basis of language is speech and gesture. We are accustomed to viewing language as written text, yet textual conventions only approximate language; they are not language itself.

A more accurate representation of a language would include audio (speech) and video (gesture). The analysis and manipulation of language data, however, require it to be in an abstract form, and rightly so. This requirement has lead to a bias in linguistic computing towards texts and against multimedia. We argue here for a multimedia application which includes speech and gesture and which is fundamentally based in a fully abstractable text.

This has been a project to create a multimedia CD-ROM of a Native American language, Rio Grande Téwa. It is fully cross-platform, with a user interface and an abstracted expandable language database. It contains videos of Tribal storytellers with accompanying text (transcribed and translated), a pronouncing dictionary, language-learning activities, and many images. The aim has been to start with speech and gesture (video, audio), but to still maintain primary focus on the texts (story transcription, dictionary, language games). These texts, representing organic language, need to be in a format flexible enough to allow manipulation of the data by future users. Yet multimedia applications available to date treat text as an irritating vestigial appendage, usually as graphical images. The problem that we have addressed, then, is how to use the advantages of multimedia yet still have the text in an abstract form.

1.2 The Social Conscience of Computing Technology

Those of us working with this expensive and (to most) inaccessible technology have at times feared it elitist, reinforcing inequities in society. Yet many of us have justified the use of such technology by citing its potential to resolve societal ills: it is supposed to enhance the collecting, archiving, processing and teaching of language data.

Recently there has been a wave of interest in multimedia applications among native peoples of the Americas as a means of archiving and teaching indigenous languages and traditions. This is part of a larger movement that of cultural and linguistic renewal worldwide.

The goal which drives this project is local empowerment. In the 1990s and beyond, notes linguist Kenneth Hale, all work done on Native American languages will be under the auspices, direction, and control of Native Americans themselves. Native American language rights in schools and government are again being challenged in the United States today by a small but vocal group of English-only advocates. Those Native American juveniles and adults who were forbidden to speak their language when growing up now have a keen desire to learn their own language with the computing tools they have become accustomed to in the workplace. With these tools, the speech community has not only a stand-alone CD for language-learning, but it also has the tools to append information to the language database. Other language-teaching modules can easily be created.

2. Project Background

Téwa is a Kiowa-Tanoan language spoken in six Native American communities in New Mexico and one in Arizona. Through the years these communities, known as Pueblos, have maintained and transmitted their language orally, often in the face of active language suppression. It is only since the 1960's that orthographies were devised for these Pueblo languages. Only with an orthography is wide-scale language renewal possible.

This project provides language renewal tools and curriculum materials for one of these communities, San Juan Pueblo. Once a leader in Native American bilingual education, the San Juan Pueblo Day School was forced to give up its language training staff and curriculum during the 1980's due to pressure from the Bureau of Indian Affairs. In 1995, Jacobs, who has worked with the community for 25 years, obtained permission from the San Juan Pueblo Tribal Council to work with community members and school staff to develop

computer-based tools for teaching Téwa as a second language. Within months, the Tribe was awarded funds from the Chamiza Foundation to develop these pedagogical tools for both children and adults at three key sites: the Pueblo school, the library, and the Ohkay Ówîngeh Cooperative. This language renewal program is *The Téwa Language Project*.

At the University of Washington's Center for Advanced Research Technology in the Arts and Humanities (CARTAH), we have been working on two aspects of this program. First, we created PC and Macintosh fonts in the Téwa practical orthography. These are already being used both by children at the Pueblo school and by adults for Tribal business. Having Téwa fonts, said one Pueblo member, is "a liberating experience.... At last I can write in my own language!"

Secondly, we created an interactive CD-ROM for pedagogical and archival purposes. We are using multimedia authoring software (*ToolBook* and *Director*) to display video, concurrent text, and play sound files; the CD also includes a queryable dictionary and numerous language-learning activities.

3. Practical Considerations

3.1 Fonts

The dialect differences between the Rio Grande Pueblos are reflected in their orthographies. Téwa uses a practical (as opposed to phonetic) orthographic system introduced in the 1960s by Summer Institute of Linguistics researchers Randell and Anna Speirs. Since Téwa distinguishes both tone and nasalization phonemically, a Latin-script based orthography necessarily includes a large character set. Using *Fontographer*, we designed and created Téwa practical orthography fonts for both Macintosh and PC-Windows platforms. A simple keyboard-reassigning program allows Pueblo users mnemonic access to characters not found on the keyboard.

The text material can be represented in the database as a font specification plus an ASCII code. While not as elegant as a 16-bit (Unicode) representation, the installation of a font in Windows or on the Macintosh is simple, and ASCII text works with any application, while Unicode is still not available in most applications. As this is a practical orthography to be used by novice and experienced computer users alike, we were compelled to use schemes that could easily be applied to a wide variety of applications.

3.2 Media Objects in a Database

While storing ASCII text in any database is easy, the storage of audio, video, and pictorial material presents some complex problems. First, these objects tend to be very large and therefore cannot be stored in normal fields. As a result, there is no indexing scheme available for the data; we cannot, for example, mark video frames in any standard database. Also, formats for different media types tend to vary on different platforms. It is fairly simple to convert between formats. But if the objects are stored in the database, we must decide whether (1) to store all possible formats, or (2) to devise a scheme that converts binary objects on the fly. Another problem is the unavailability of cross-platform database file formats that support multimedia objects. This problem alone is enough to make both of the above approaches unfeasible. One solution has been to store only references to multimedia files in the database. This presents some new problems and doesn't solve many of the problems of an embedded system, but it is certainly feasible. Neither does this approach address any scheme of indexing in the files themselves, nor does it solve problems of differing media formats across platforms. Referencing files presents new problems, including the management of the data files, the verification of the data files, and the construction of naming schemes the are independent of any operating system. Because it is only ASCII text that is being stored, from the perspective of the database itself implementation is not problematic. Its advantages include: the database stays a small size; cross-platform compatibility of the database is proven; it is somewhat easier to manage files as opposed to managing large binary objects in a database; and any large or universal formatting changes can be done independent of the database (e.g. converting stereo to mono audio files).

We've chosen to use an XBASE file format, *dBase*, because we've found suitable readers for the XBASE format for both Mac and PC. Since the file format is very common, it is a simple matter to build reader applications on any platform. The files originate as data entered into a specially-designed *ToolBook* application, which are then exported to a *dBase* file format. Once we have the database well defined, we will move to a *ToolBook* front end to the *dBase* file and enter data directly to the database. *ToolBook*'s facility for this is provided in a DLL that is part of the *ToolBook* product. On the Macintosh platform, we are using a product called *FileFlex* that is made available as an XLIB that, when incorporated into *Director*, allows *Director* to act as a front end to the *dBase* file. We chose *ToolBook* as the primary editing tool because its relationship to *dBase* is well known by members of the team. We have been able to edit the database from *ToolBook* on a PC and immediately view the results in *Director* on a Macintosh.

3.3 The CD-ROM

We will demonstrate a CD-ROM with the database information organized and optimized to be cross-platform. Along with the database, we will present supporting applications and examples of language study curricula. The dictionary is independent of any application or platform, being abstracted into a *dBase* file with suitable references to audio and visual media. The organization of the media files is contingent on compatibility within any one medium. Some file formats will play on both platforms and so will be held in a common directory. Others may need different formats; the original format will be converted to one suitable for the remaining platform and both files will be placed on the CD. The reader applications on the CD will need to be aware of all these quirks and discrepancies so that they can interpret the media references in the database in a meaningful manner.

These considerations limit the amount of information we can place on the CD. Video consumes large amounts of space. It becomes imperative that any video we use not be duplicated, so we must convert our videos to QuickTime files, as this is the most stable cross-platform standard for video. Still, the size of the videos makes it impossible to include more than five or six extended (six minute or more) videos. This is unfortunate, as this type of material, typically story-telling, conveys best of all the gesture aspects of the language. Obviously, much of the material that is currently on video cassettes held in private collections will only be available for general distribution once higher-density media becomes available.

Audio files must be dealt with in a similar manner, but it is easier to control the size of the files. Mostly, audio files will be used for pronunciation of the words in the dictionary, and while these will be short, there will be a large number of them. In order to avoid duplication, we will need to use a single format that both platforms can read. Illustrations and texts can be treated in a similar manner.

The main database itself will not be particularly large. The final version of this database will probably not be larger then two megabytes. The fields for each record in the database are: a Téwa headword: a unique identifier; English gloss; example sentence; English gloss of the examples sentence; reference to the audio file; semantic category; and variant lexemes.

Certain possible fields we have deemed unessential to the purpose of the primary database, but as having potential interest to some users. We will build secondary, related databases as time allows and include them on the CD-ROM. Examples of these databases include an index of video frames that illustrate words and an index of where words are used in the included texts.

References

Hale, Kenneth. 1972. A New Perspective on American Indian Linguistics. In Alfonso Ortiz, ed., *New Perspectives on the Pueblos*, Albuquerque N.M.: University of New Mexico Press: 87-133.

Leap, William L. 1988. Indian Language Renewal. *Human Organization*, Vol. 47, No. 4: 283-291.

Martinez, Esther. 1982. *San Juan Pueblo Tewa Dictionary*. San Juan Pueblo, New Mexico: SJP Bilingual Program.

Speirs, Randall. 1968. *Linguistic Aspects of Rio Grande Tewa (Ph.D. dissertation). Ann Arbor, MI: University Microfilms.*