

## EXPANDING THE LEXICAL RECORD: CREATING MULTIMEDIA DICTIONARY DATABASES FOR NORTHERN CADDOAN AND SIOUAN LANGUAGES

Raymond J. DeMallie, Wallace E. Hooper, and Douglas R. Parks  
Indiana University

Slightly over ten years ago I began to migrate my Pawnee and Arikara lexical data from typed file slips to a computer database. At that time, there was no **standard** database software program designed for three user purposes: (1) to support a lexical archive, (2) to compile a dictionary, and (3) to prepare camera-ready copy of it for publication. The program that we chose was Notebook II by Protem, one that is no longer supported, but which at the time met my needs: it allowed up to fifty (50) fields, up to 32,000 characters within a field, and provided for the generation of camera-ready copy for publication in various formats. No less importantly, Notebook II was simple to use. There were eighteen (18) fields that I created for the Pawnee and Arikara databases. In addition to entry form, phonetic form, grammatical classification, and glosses, I wanted to provide other kinds of information for each stem: its derivation; paradigmatic forms of verbs; grammatical forms of stems; examples of the entry in context--both phrases and sentences--to illustrate usage; and citations of the stem or word in historical sources. In addition, within an entry I wanted to cross reference synonyms and other semantically related entries, as well as to provide variant dialectal forms and cognates in related languages. (So, for Pawnee entries, I entered Arikara cognates, and Pawnee cognates for Arikara entries. Eventually, the cognate references will include Kitsai and Wichita, all aimed at a comparative dictionary of Northern Caddoan.)

As the Pawnee and Arikara databases began to expand, various fundamental problems

with the Notebook database program soon became apparent. For example, when an entry form has two or more different glosses—a common occurrence—I was unable to relate specific examples to a specific gloss. Moreover, for inflected forms of stems it was awkward to provide complete information for more than one of them; that is, to give both the surface and underlying forms, gloss, source, and notes on usage or other matters. In short, by 1992 I had outgrown the Notebook database.

About that time multimedia technology began to come into its own, and the desire for a “talking dictionary” took hold among our team at Indiana. The desire for sound grew out of our work with written texts and sound recordings from the early decades of this century--the George Sword material being edited by Raymond DeMallie and the Roaming Scout narratives being edited by me--as well as the urging of many Arikara people. One Arikara friend, for example, a woman in her 70s, has politely nagged me for over a decade to make tape recordings of the words that would be in my dictionary and put them on tape for people like her--for people, she would repeat, who cannot read Arikara. Over the past few years I came to realize that sound recordings are no less valuable for scholars as well, including linguists.

In the early 90s, digital recording technology was also beginning to come into its own, providing sound recordings significantly better than the analog recordings that most of us had been getting from Uher and Sony tape recorders. The issue, then, became one of finding a database program that could incorporate **digitized** sound recordings of entry forms—to which we added the requirement that it accept images and video clips. By accepting images, photographs, drawings, and other useful visuals could be incorporated into entries when appropriate for illustration of cultural objects, cultural acts, historical personages, geographical landmarks, flora,

fauna, and so forth. This feature is important for identifying objects and cultural acts for which there are no standard designations in English.

### **Selection of Visual FoxPro**

In 1994 we looked at database software choices and selected Microsoft's Visual FoxPro, an off-the-shelf program that met all of our needs, including the important consideration that, because of its industry-wide use, the database can be easily migrated to new formats as technology develops, rather than be left behind in some obscure format. Specifically, the basic features that went into the choice of FoxPro are the following:

- **Structure.** Rather than flat field, its structure is relational and allows one-to-many relationships, so that one higher level record may be associated with many lower level records. For example, one entry form may have many glosses, and each of those glosses may have many examples specific to it.
- **Printing.** FoxPro allows for a full set of printing functions that enables users to select for printing any and all of the fields on its different tables and to determine the order of presentation of various elements in each dictionary entry. It also allows choice of fonts and print styles for each field. It has allowed staff programmers to design output routines for four basic file formats: Rich Text Format, HTML, WordPerfect 5.1, and ASCII text.
- **Multimedia functions.** FoxPro will store and play sound files and video clips in a variety of formats as well as display graphic images in various formats.
- **Versatility of use.** The FoxPro database can serve simultaneously as a linguist's archive, in which an exceptionally large amount of data is organized and stored, and as a user's reference dictionary. Thus, the dictionary can be disseminated in two forms (on CDs or

DVDs): one a linguist's version that allows the user to enter the archive, or storage area; and the other a user's version that keeps the user out of the archive, giving access only to the data normally presented in a dictionary.

- **User Friendliness.** A final issue was user friendliness. Although FoxPro requires a skilled programmer to create the database with the features we wanted, the program as designed is notable for its clear organization and navigational simplicity. Data are organized on separate screens or tables.

Programming on the Indiana Dictionary Database (IDD) program began in 1995. In fall 1996 four linguists, including Wally Chafe and Laurel Watkins, and several graduate students, came to Bloomington for a one-week workshop to try out the program as it existed then and to suggest ways to make it generalizable to the needs of a wider range of linguists—or, actually, to a wider range of languages. Subsequently, at IU we began compilation of dictionary databases for Assiniboine and Yanktonai Sioux, both Siouan languages, and more recently Phil LeSourd has also begun two Algonquian dictionary databases, one for Maliseet and one for Arapaho.

In 1995 we purchased our first digital tape recorders and began making sound recordings for the databases. Initially, the goal was to include a recording of the entry form, and for verbs to include one or two inflected forms. Those numbers of recordings for each entry were dictated by space limitations on hard disks and on compact disks at the time. Today, however, as hard disk space increases and as DVD and other larger formats are looming on the horizon, we have moved to the goal of making sound recordings of **all** the forms in the database and including sentences illustrating usage. Moreover, we are digitizing older analog recordings, both of vocabulary and textual material, and plan to incorporate those materials into the database, enabling us to include

recordings of various speakers and even multiple recordings of a given word from different individuals to illustrate speaker variation.

In addition to the recordings for databases, our goal now is to develop sound archives for each of the languages we are documenting, preserving as much material in digital format as possible. The purposes are by now obvious:

- One is to provide the very best sound record of each language that current technology will allow. Such records better serve future linguists, who will not have to rely on printed forms alone but can listen to and otherwise utilize the recordings after the languages have become extinct and after analytic techniques have developed beyond those of today.
- Another is to provide user-friendly dictionaries that will be more useful to tribal members, now and in the future, as in fact to scholars as well. These dictionaries free the user from having to interpret phonetic and phonemic renditions, unlike printed dictionaries that leave the reader wondering if he has actually pronounced a word correctly.
- Finally, the sound archives that are created will also serve as a resource for native community needs: (1) for current and future efforts to create pedagogical materials for language programs, and (2) for such recurring needs as accessing personal names and the meanings of song lyrics.

Now we would like to give a short demonstration of the program.