Web-based Dictionaries for Languages of the South-west USA

Sonya Bird, Michael Hammond, Maria Amarillas, Melody Jeffcoat, Heidi Harley, Mizuki Miyashita, Laura Moll, Mary Ann Willie, and Ofelia Zepeda
Department of Linguistics, University of Arizona, USA

Abstract

This paper outlines a project currently under way in the Linguistics Department at the University of Arizona to create electronic dictionaries of indigenous languages of the south-west USA and make them available over the Web for language instruction as well as for linguistic, psycholinguistic, and anthropological research. Working with three languages—Tohono O'odham, Navajo, and Hiaki—we have created an XML scheme that serves as a general template for structuring and archiving language databases. We describe the process of compiling databases for different languages and converting these databases to XML, which contains all the relevant information in a manner that is easily accessible. We discuss the general programming scheme used for searching, and the interfaces used for presenting the dictionary on the Web, which include several front ends for different user groups. We end with a discussion of how to ensure that special characters are displayed properly on the Web.

Correspondence:
Sonya Bird,
Department of Linguistics,
University of Arizona,
Douglass 200 E,
Tucson, AZ 85721, USA.
E-mail:
sbird@u.arizona.edu

1 Introduction

According to Krauss (1998), of the 210 indigenous languages still extant in North America, only thirty-four (16 per cent) are spoken by speakers of all generations, thirty-five (17 per cent) are spoken by the parental generation and up, eighty-four (40 per cent) are spoken by the grandparental generation and up, and fifty-seven (27 per cent) are spoken only by a few elders. As Krauss points out, all of these languages are severely endangered in that fewer and fewer children are learning them. In response to the threat of extinction facing indigenous languages in North America (and around the world) a push has begun in recent years to create electronic resources for these languages to help with revitalization and maintenance efforts and, should these efforts fail, to at least provide written documentation of these languages.[1,2]

This paper outlines a project that contributes to this push by creating electronic dictionaries for endangered languages of the south-west USA. The goal is to make these dictionaries available over the Web for native

1 Some examples are listed in Buszard-Welcher's (2001) paper 'Can the web help save my language?' Updated links to these and other examples can be found on Buszard-Welcher's website: http://www.potawatomilang.org/Reference/endlgsweb_update.htm.

2 Funding has come from the University of Arizona (UA), Sun Microsystems, the Department of Linguistics at the UA, and the Social and Behavioral Sciences Research Institute at the UA.

speakers and language learners, as well as teachers and researchers in the fields of linguistics, psycholinguistics, and anthropology. We began our project in 1998 by scanning, editing, and uploading to the Web an out-of-print dictionary of Tohono O'odham (Mathiot, 1983). Since then, we have been expanding our work along three lines: (1) we have included other languages in the project—Navajo and Hiaki; (2) we have refined the general structure of the lexicons; (3) we have fine-tuned the presentation of the material on the Web. In this paper, we introduce the languages involved in the project (Section 2), discuss the general programming model used in creating the dictionaries (Section 3), and consider the question of displaying special characters correctly on the Web (Section 4). We conclude by outlining the next steps of the project.

2 Languages Involved and Database Compilation

We have focused on languages of the south-west, to make our project maximally useful to the indigenous populations closest to the University of Arizona: Tohono O'odham, Navajo, and Hiaki. These languages share the fact that they are endangered, and creating resources for them now is crucial to help with preservation and maintenance efforts. However, they differ in the availability of existing written materials on which to base an electronic dictionary. In this section we consider each language, outlining what is being done towards compiling an electronic dictionary based on the availability of existing written material.

2.1 Tohono O'odham

The Tohono O'odham (formerly Papago) Nation is located about 65 miles west of the University of Arizona Campus and the city of Tucson. In the 1990 census, its population was estimated at 20,000, with approximately 11,819 native O'odham speakers, most of whom were over twenty-five (Ethnologue, 2002).

There are two dictionaries of the Tohono O'odham language, one written by Saxton *et al.* (1983) and the other by Mathiot (1983). We chose to work with Mathiot's *A Dictionary of Papago Usage* (Mathiot, 1983)[3] for two reasons: first, the Dictionary is out of print, with no plans to reprint. Second, Mathiot's dictionary is much more comprehensive than the Saxton *et al.* (1983) dictionary: as well as containing many more lexical entries, it includes detailed grammatical information on each entry, example sentences, and information on dialectal variation (Miyashita and Moll, 1999). Database compilation has consisted of scanning and editing this dictionary. The editing process involves two tasks. First, many of the special characters in Tohono O'odham were lost in the scanning process, and had to be corrected. Short vowels in O'odham are spelled ă, ĕ, ĭ, ŏ, and ŭ. The palatalized and velar nasals are spelled ñ and ŋ, respectively. Finally, O'odham has two retroflex sounds spelled ṣ and ḍ. None of these characters were read properly by the optical character recognition (OCR) software available to us during the scanning process. These characters and others therefore had to be corrected by hand. This was done in two stages. In a previous, text-based,

3 With the permission of the author.

version of the Tohono O'odham electronic dictionary (Miyashita and Moll, 1999), special characters that were misread by the OCR software were converted to symbols found in the US character set. For example, 'f' was used for ḍ and 'x' was used for ṣ. The user interface of the dictionary included a legend indicating the correspondences between the characters used in the electronic dictionary and the Tohono O'odham characters (http://w3.arizona.edu/~ling/mh/lmmm/to.html). Currently, as part of the Java-based electronic dictionary we are creating (see Section 3), we are able to display the actual O'odham characters in a platform-independent fashion. This procedure is discussed in more detail in Section 4.

Apart from difficulties with the fonts, the original printed dictionary contained some typos and other errors, which were frustrating for speakers and learners. The second task involved in editing has therefore been the correction of any mistakes in the content of the lexical entries. This work is being done by a group of native speakers of Tohono O'odham.

2.2 Navajo

The homeland of the Navajos is a 26,000 square mile reservation, which stretches into three states: Arizona, New Mexico, and Colorado. The 1990 census estimated the Navajo population to be 218,198, with 148,530 native Navajo speakers. However, only 30 per cent of first graders spoke Navajo as their first language in 1998, compared with 90 per cent in 1968 (Ethnologue, 2002). The most current population count exceeds 250,000, but only approximately 20,000 people still speak Navajo.[4]

There are two monumental dictionaries of the Navajo language: *The Navajo Language* (Young and Morgan, 1987) and *An Analytic Lexicon of Navajo* (Young *et al.*, 1992), both written by Robert Young and William Morgan, scholars of Navajo (the analytic lexicon was also written with Sally Midgette). Although these two voluminous dictionaries are important contributions to Navajo studies, unfortunately the works are inaccessible to most students of the language. This is primarily because of the complexity of verb forms in Navajo. Verbs are built on abstract roots, which combine with appropriate prefixes to become utterable words, corresponding to full sentences in English. For example, the verb *dabidishní* translates to 'I say it to them individually' and is composed of the verb stem *-ní*, and four prefixes: *da-* ('them, individually'), *bi-* ('them'), *di-* (adverbial prefix), and *sh-* ('I', present tense). Given all of the possible prefix and root combinations, the number of utterable words in Navajo is virtually limitless (Bird, 2001). Because of the nature of Navajo verbs, Young and Morgan were forced to take reasonable shortcuts in creating their lexicons, and these shortcuts are what make the dictionaries unusable to most people. To look a word up in *The Navajo Language*, it is necessary to know the first person singular present form of the verb (which is not always obvious). In *An Analytical Lexicon of Navajo* it is necessary to know the abstract root form of the relevant verb stem (also often difficult to identify). These two lexicons are so difficult to make sense of that the Diné (Navajo) Community College offers a course that focuses solely on learning to use these resources.

4 Mary Willie (Navajo language expert), personal communication.

Our goal in creating an electronic dictionary for Navajo is to create a more accessible resource for users, something that will be possible because electronic materials are not subject to the same restrictions in size as printed materials. We have created a relatively small lexicon based on Professor Mary Willie's draft Navajo language teaching materials, so as to have a tool that will be immediately useful to Navajo students at the University of Arizona. We are now in the process of collecting materials with which to expand this lexicon. Navajo also has special characters, which present the same difficulties as the Tohono O'odham ones. Section 4 presents a discussion of how these are being displayed.

2.3 Hiaki

Hiaki (also referred to as Yaqui or Yoeme) is a Uto-Aztecan language, spoken by a population that straddles the US–Mexico border. There are about 9,000 Hiakis living in Arizona, and approximately 30,000 in the Mexican state of Sonora. Approximately 150–250 Arizona Hiakis are fluent speakers of the language, all of whom are fifty or older.[5]

Existing written resources for Hiaki include an English–Hiaki bilingual dictionary (Shaul *et al.*, 1999) and a Hiaki grammar (Dedrick and Casad, 1999). In addition, Maria Amarillas, a fluent speaker, language educator, and student in Linguistics at the University of Arizona, has a preliminary trilingual dictionary of Hiaki under development (Hiaki–Spanish–English), which we are in the process of digitizing. This dictionary will serve as the basis for the Web-based dictionary, because it is trilingual and will therefore be most useful for Hiaki speakers and learners (whose first language is often Spanish). Unlike Navajo and Tohono O'odham, Hiaki has no special characters, and can therefore be displayed using the US character set. This makes Hiaki easier to display on the Web than Navajo and Tohono O'odham.

2.4 Summary

The three languages with which we are working differ in the written materials currently available for them. Tohono O'odham has a printed dictionary, which we were able to scan and use as a basis for an electronic dictionary. Navajo, on the other hand, has no dictionary in a format easily made electronic. Finally, the Hiaki dictionary we are digitizing is still a work in progress. These languages also differ in terms of the elements that are required of an electronic dictionary. Navajo and Tohono O'odham both have special characters, which must be displayed properly on the Web, whereas Hiaki does not. The Hiaki dictionary—unlike the other two—includes a Spanish interface, as many of its users speak Spanish rather than English as their first language and as it is based on Amarillas' dictionary, which is trilingual. Finally, Tohono O'odham, Hiaki, and Navajo differ in morphological complexity. Consequently, they differ in the amount of information that must be included on the internal structure of words. For example, the structure of the Navajo dictionary must allow for detailed descriptions of the morphological breakdown of words, as this breakdown is not always transparent. The level of detail required for

5 Maria Amarillas (Hiaki language expert), personal communication.

encoding information on Navajo morphology is not necessary for Tohono O'odham or Hiaki, as they are much simpler morphologically. Despite the differences among Navajo, Tohono O'odham, and Hiaki, the general programming model for the three dictionaries is the same. In the following section, we turn to a detailed discussion of this model.

3 General Programming Model

Although the differing availability of written materials for Tohono O'odham, Navajo, and Hiaki necessitates differences in how their databases are compiled, they are all being converted to the same format using a generalized XML markup scheme.[6] Based on this markup scheme, the Tohono O'odham, Navajo, and Hiaki dictionaries will all be accessed in the same way on the server side (in conducting searches), and the user interfaces will also be the same across languages. In this section we consider first the general programming model used in the dictionary project (Section 3.1). We then discuss the details of the server side components (Section 3.2) and user side components (Section 3.3) of this model.

3.1 General programming model

The general programming scheme used in this project is Java-based. On the server side, persistent servlets are used to store the databases and search through them. On the user side, applets are used to handle all aspects of user interaction.[7]

3.2 Server side: servlets, XML markup scheme, and search mechanism

3.2.1 *Servlets*

On the server side, we are using a single servlet-based system for dictionary access that works for all our dictionaries. This system has two advantages. First, it is fast: once the dictionary program has been accessed, it stays running (until the server is shut down). Practically, this means that the first time a user enters a query it takes a few seconds, because the dictionary program must be uploaded and stored in memory. However, any subsequent searches are very fast, as the program is already loaded and waiting to be accessed. The second advantage to this system is that it is easily extensible to new languages, which means that including other languages in the project will be simply a matter of formatting the dictionaries in the appropriate manner.

3.2.2 *XML markup scheme*

We use XML as our markup scheme, a text-based system used to organize and archive data, including linguistic data. There are two reasons why XML was chosen as the markup language for structuring the databases. First, XML allows us to use generally available tools for parsing textual data, e.g. XMLspy, Expat, Jaxp, etc. Second, XML provides a means of encoding rich structure in a simple text format. This allows for portable data, which means others can use it more readily. We are working in

6 Similar schemes are used in other projects such as KirrKirr, software designed by Manning *et al.* (2001) and used to display an electronic Warlpiri dictionary in an interactive fashion (http://www-nlp.stanford.edu/kirrkirr

7 Thanks to a grant from Sun Microsystems, we have a server dedicated entirely to the dictionary project, which gives us the necessary processing power. In addition to this, we have several workstations for our development purposes.

coordination with the E-MELD project (http://saussure.linguistlist.org/cfdocs/emeld/), the goal of which is to create a common standard for the digitization of linguistic data (Langendoen *et al.*, 2002). The E-MELD group has recommended the use of XML as a markup language so as to be able to share linguistic data with others easily.

An alternative approach would involve encoding the data using the Text Encoding Initiative (TEI) standard (Sperberg-McQueen and Burnard, 2002). The TEI standard is used in the creation of other electronic dictionaries, such as the Scottish National Dictionary (Rennie, 2001). We chose not to use the TEI standard because its recommendations for dictionary encoding were developed to represent the structure of already existing printed dictionaries for widely spoken languages, and do not address the needs we have outlined with respect to the dictionaries we are working on.

The following paragraphs illustrate the components of the XML markup scheme used.[8] The XML files consist of elements, tags, and attributes. An element is a piece of information associated with a lexical entry. Each element is surrounded by tags, which encode the kind of information provided by the element. Attributes are associated with specific tags, and provide additional information about elements. The following example illustrates: <headword xml lang="i-navajo" id="e392"> ashiik;e</headword>.[9] The components of this partial entry are listed in Fig. 1.

element:	ashiiké ('boys')
tag:	headword (lexical entry)
attribute language:	Navajo
attribute id:	392

Fig. 1 Components of an entry in XML format.

The attribute 'id' gives each entry an identification number, which is used, for example, in cross-referencing.

The form of our XML scheme is given in a 'document type definition' (dtd) file (a dtd file formalizes the hierarchical structure involved in the XML files). The dtd file used for the dictionaries in this project has the structure illustrated in Fig. 2.[10]

dictionary → entry+
entry → headword sense+
sense → grammar definition
grammar → grammaritem*
definition → *text* (breakdown) example*
breakdown → morpheme
morpheme → form meaning+
example → sentence translation
headword → *text*
grammaritem → *text*
form → *text*
meaning → *text*
sentence → *text*
translation → *text*

Fig. 2 Document type definition (dtd) file for the dictionary project.

8 Those familiar with XML may wish to skip this section.

9 Note the use of *;* in the word *ashiik;e*. In the output, the sequence /;e/ is converted to *é* using Java graphics, displaying the word *ashiiké*.

10 The actual dtd file is formatted somewhat differently. It can be found on the website: www.lexicon.arizona.edu/~sbird/dictionaries.html.

In the pseudo-dtd file shown in Fig. 2, arrows can be read as 'consists of'. The first line *dictionary* → *entry*+ specifies that the dictionary consists of one or more entry.[11] The next line *entry* → *headword sense*+ specifies that each entry consists of a headword and one or more senses. Each sense consists of some grammatical information and a definition (*sense* → *grammar definition*). The grammatical information associated with an entry includes any number of pieces of grammatical information (*grammar* → *grammaritem**). For nouns, for example, this may include information on gender, number, case, etc. The definition consists of the English translation, morphological breakdown of the headword (parentheses indicate that this field is optional), and any number of example sentences (*definition* → *text (breakdown) example**). The field *breakdown* provides, for each morpheme, its form and its meaning (*breakdown* → *morpheme*+ and *morpheme* → *form meaning*). Finally, 'examples' consist of sentences and their translations (*example* → *sentence translation*).

Figure 3 illustrates what the lexical entries look like once they have been encoded using the general XML markup scheme. The headword of this entry is *ashiiké*, which only has one sense ('boys'). The grammatical information associated with *ashiiké* includes information on its part of speech (noun), and on its number (plural). The definition includes only the English translation here; neither morphological breakdown nor an example sentence is provided.[12]

```
<entry>
        <headword xml lang="i-navajo" e="392">ashiik;e</headword>
        <sense>
                <grammar>
                        <grammaritem xml gram="POS">noun</grammaritem>
                        <grammaritem xml gram="num">plural</grammaritem>
                </grammar>
                <definition>boys</definition>
        </sense>
</entry>
```

Fig. 3 Sample Navajo entry.

This XML scheme can be manipulated using the Java 'JAXP' package. This allows for search and retrieval operations, printing, and conversion to other encoding schemes. In terms of searching and retrieving information, JAXP is particularly useful for morphologically complex languages such as Navajo, because it searches over the morphological structure encoded in the *grammar* element. Also, because it can allow for retrieving substrings, users can search for particular morphemes, whether or not they recognize them as morphemes. This solves the problem with Young and Morgan's dictionaries, in which finding information requires understanding of morphological structure (see Section 2.2). All of the words in the Young and Morgan dictionaries will eventually be in the electronic dictionary. However, as the search is handled by the software, users need not be troubled by trying to determine what the form is under which entries are organized. In other words, users will simply be able to

11 Parentheses means 'zero or one'. The symbol '+' means 'one or more'. The symbol '*' means 'any number (zero included)'.

12 Some lexical entries are much more complex, including example sentences and references to related words. Ultimately, all of the lexical entries should contain all of this information.

type in a word as they see it, rather than having to search for related forms of the word they are interested in (as is the case with the Young and Morgan dictionaries).

Two features of the markup are worth mentioning here. First, one can include information in the database without necessarily having to display it. For example, if the database is compiled with the help of specific speakers, each entry could include information on the name of the speaker who provided the word. Although this information is not normally displayed, a specialist interested in dialect differences among speakers would be able to access this information. Because not all elements in the database must be displayed, it is possible to include structure in the database that is not necessarily used for each language. For example, it is possible to include a field for the Spanish translation of Tohono O'odham words, even though we do not (at this time) have access to these translations. This flexibility allows us to generate a single dtd file for all languages. The fields that are irrelevant for a specific language are simply left empty of content and not displayed. The general dtd file is also easily manipulated such that fields and hierarchical structure can be added as needed. For example, if sound files and images are associated with lexical entries (something that we hope to do in the future), displaying them will simply require adding a field for them under the definition in the dtd file (resulting in the structure: *definition* → *text (breakdown) example** *sound image)*.

The second feature of this markup worth mentioning is that, using other XML technologies such as XSLT, XML can be easily converted into a format appropriate for a printed dictionary. Thus, the same database object can be used for an electronic and a printed dictionary. For example, we have an XSLT script that reformats the Tohono O'odham dictionary as a LaTeX document. Using such technologies, generating printed dictionaries containing different kinds of information can be greatly simplified.

Converting the dictionaries to XML format is a work in progress. In the case of Tohono O'odham, there is still a lot of work to do. Figure 4 provides an example of what the lexical entries of the scanned dictionary look like.[13]

```
#baab
%to be somebody's mother's father, mother's father's
brother

vu~ Stat lixpr sgs [pls vu] baaba?a]: to be somebody's
mother's father, mother's father's brother ex: Heg go g1(<
vu,d) nb?ab. He is my mother's father (or: my mother's
father's brother).
```

Fig. 4 Scanned Tohono O'odham entries.

Converting this to XML involves pulling out each piece of information and inserting it between the relevant tags (as well as inserting the appropriate special characters). For example, the first two lines of the entry in Fig. 4 are represented in XML in Fig. 5.

13 The symbols # and % were inserted manually as initial informal markup to make the headwords (#) and simple English definitions (%) easy to extract from the dictionary file.

```
<headword xml lang="i-to">baab</headword>
    <sense>
        <definition>be somebody's mother's father, mother's father's
        brother</definition>
    </sense>
```

Fig. 5 XML version of the entry in Fig. 4.

As can be seen from Figs 4 and 5, there is a lot of information associated with each entry that must be carefully separated out and placed into appropriate fields. We have automated this conversion as much as possible with Perl scripts, but much of it must be done by hand. This is part of the work that the O'odham editors are doing.

For Navajo and Hiaki, the conversion to XML format is not an issue because the databases are being created from scratch. In the case of Navajo we have been using XML from the start. As for Hiaki, we are digitizing Amarillas' draft dictionary in Microsoft Word, for word processing ease, but the format we are using is easily convertible to XML.[14]

3.2.3 *Searching*

To run searches on the dictionaries, we currently use a 'Document Object Model' (DOM) parser, which reads the relevant XML file and stores it in memory as a DOM object. It is this object that serves as the basis for any searches. Because iterative searches on the DOM object are relatively slow, we have created a hashtable[15] for each dictionary, which contains all of the headwords. The search mechanism starts by locating the appropriate headword in the hashtable. The headword is associated with an index that points to the complete entry in the DOM object. Using this search method speeds up search time considerably.

3.3 User side: applets and uniform Java interfaces

One of the key components of this project is the development of multiple Web-based interfaces to the same dictionary objects. The idea is that native speakers of a language, learners of that language, and researchers will make use of the same dictionary object, using different front ends tailored to their needs. For example, learners of Tohono O'odham would be presented with an English-based interface to the dictionary that would allow them to search for words easily and would return them in a format familiar from other pedagogical materials. Researchers would have a different interface, one that allowed them to search for words using the variables of interest, and returning lists of words exhibiting the relevant properties. Finally, speakers of the language would be presented with an interface *in the language*.

14 Again, we have Perl scripts to do this.

15 A hashtable is a table containing a list of keys and values associated with each key. In our case, the keys are headwords, and the values associated with them are indices (which point to the full entries found in the XML DOM model).

4 Displaying special characters

Displaying special characters is a common concern in creating Web-based language resources. As mentioned above, Tohono O'odham and Navajo both have special characters that are not found in the standard Unicode character set. If the Web pages were static (i.e. not interactive)

special characters could be displayed simply, using gif files for example. However, our pages are necessarily interactive, as users must be allowed to enter a word, in Tohono O'odham or in Navajo, and have it displayed properly as they are typing. To do this, there are several possibilities.

One possible way of displaying special characters is to use a specific font set, which users must download. This possibility was rejected because in some cases it may not be possible for users to download fonts, either because they are not familiar enough with computers to do this or because they are not permitted to do so.[16] Another possibility is to make use of an extended Unicode character set, which would include the characters necessary for displaying Tohono O'odham and Navajo properly. At this time, most computers do not come with any extended Unicode character sets that include the appropriate special characters (although this is changing). Using extended Unicode was therefore rejected for the same reason as using special fonts: it requires users to download an additional character set, which may not always be possible. An additional reason for rejecting this possibility is that Navajo high-toned nasalized vowels are not found in any existing Unicode set, which means that displaying them requires some combination of the characters available as part of Unicode, e.g. a + ´ + ˛ = á. There are currently no plans to develop a Unicode set that includes Navajo high-toned nasalized vowels as single characters.

To provide users of various dictionaries with accurate representations of the relevant language without requiring them to install any sort of special fonts, we have created Java applets that take as input predetermined symbol combinations and use graphics to display special characters as output. The Navajo special characters are displayed in Fig. 6.

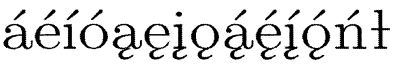

Fig. 6 Navajo special characters displayed using Java graphics.

The Tohono O'odham special characters use a similar typeset but have different diacritics, as shown in Fig. 7.

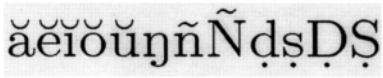

Fig. 7 Tohono O'odham special characters displayed using Java graphics.

Figure 8 illustrates how Java is used to display Tohono O'odham fonts in a text.

The relevant dictionary interfaces include a legend indicating which key strokes to use for which characters. For example, a high-toned vowel in Navajo is written by typing the sequence *; vowel*. In summary, using graphics allows users to see their language in a familiar way, without requiring them to make any changes to their local machine.

5 Conclusion

In this paper we have presented a project to create Web-based dictionaries for languages of the US south-west. The project is very much a work in progress. We are currently working on: (1) augmenting the Navajo and

16 For example if the user is on a machine in the local library.

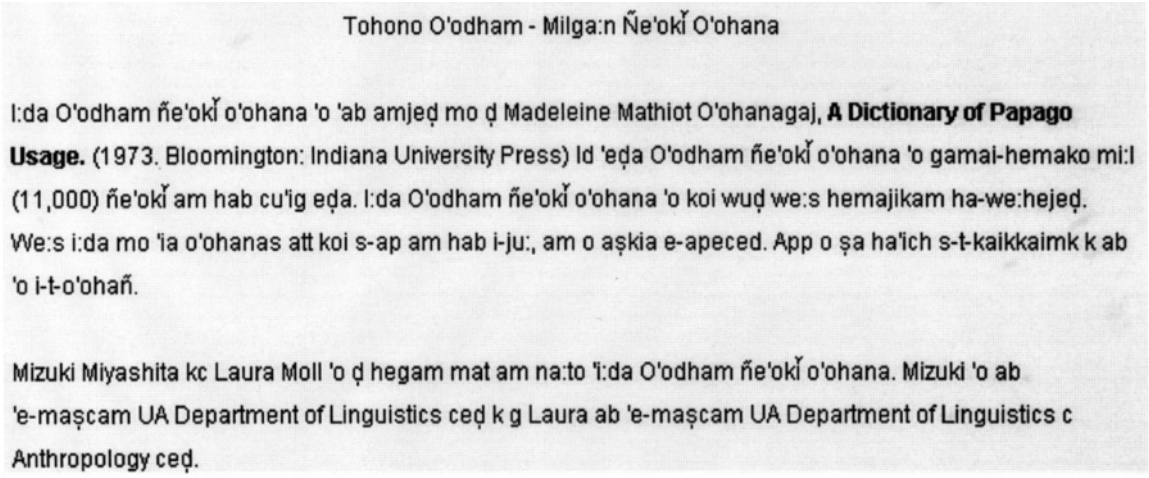

Fig. 8 Using Java to display Tohono O'odham text accurately.

Hiaki databases; (2) converting the various written materials we have to XML format; (3) fine-tuning the markup and lookup schemes; (4) perfecting the user interfaces. Furthermore, although we will maintain XML as an archival and transfer medium, we are exploring a relational database model for online access to our dictionaries.

Despite the work that remains to be done on the individual dictionaries, we now have a general template for creating Web-based dictionaries and making them accessible to native speakers, teachers, language learners, and researchers. This template is particularly important because it is extendible to other languages, and will therefore be a useful tool for continuing documentation and revitalization work on endangered languages. Furthermore, the data is in XML format, which means it is easily accessible to other researchers as well as to projects such as E-MELD. For more information on this project, please consult our website: www.lexicon.arizona.edu/~sbird/dictionaries.html.

References

Bird, S. (2001). What is a word? Evidence from a computational approach to Navajo verbal morphology. *WSCLA 5: Proceedings of the 5th Workshop on Structure and Constituency in the Languages of the Americas*. Vancouver, B.C.: UBC Press, pp. 1–17.

Buszard-Welcher, L. (2001). Can the web help save my language? In Hinton, L. and Hale, K. (eds), *The Green Book of Language Revitalization in Practice*. San Diego, CA: Academic Press, pp. 331–48.

Dedrick, J. M. and Casad, E. H. (1999). *Sonora Yaqui Language Structures*. Tucson, AZ: University of Arizona Press.

Ethnologue (2002). Website: http://www.ethnologue.com. Copyright © 2002 SIL International.

Krauss, M. (1998). The condition of Native North American languages: the need for realistic assessment and action. *International Journal of the Sociology of Language*, **132:** 9–21.

Langendoen, D.T., Farrar, S., and Lewis, W. (2002). Bridging the Markup Gap: Smart Search Engines for Language Researchers. *Proceedings of the Workshop on Resources and Tools for Field Linguistics*, Las Palmas, Canary Islands.

Manning, C., Jansz, K., and Indurkhya, N. (2001). Kirrkirr: software for browsing and visual exploration of a structure Warlpiri dictionary. *Literary and Linguistic Computing*, **16**(2): 135–52.

Mathiot, M. (1983). *A Dictionary of Papago Usage*. Bloomington, IN: Indiana University Press.

Miyashita, M. and Moll, L. (1999). Enhancing language material availability using computers. In Reyhner, J., Cantoni, G., St. Clair, R. N., and Parsons Yazzie, E. (eds), *Revitalizing Indigenous Languages*. Flagstaff, AZ: Northern Arizona University, pp. 113–16.

Rennie, S. (2001). The Electronic Scottish National Dictionary (eSND): work in progress. *Literary and Linguistic Computing*, **16**(2): 153–60.

Saxton, D., Saxton, L., Enos, S., and Chury, R. L. (1983). Dictionary: Papago/Pima–English, O'othham–Mil-gahn; English–Papago/Pima, Mil-gahn–O'ottham. Tucson AZ: University of Arizona Press.

Shaul, D., Molina, F. S., and Valenzuela, H. (1999). *Yoeme–English/English–Yoeme Standard Dictionary*. New York, NY: Hippocrene Books.

Sperberg-McQueen, C. M. and Burnard, L. (eds) (2002). *Guidelines for Electronic Text Encoding and Interchange*. TEI Consortium. Oxford: Humanities Computing Unit, University of Oxford.

Young, R. and Morgan, W. Sr (1987). *The Navajo Language: a Grammar and Colloquial Dictionary*, 2nd edn. Albuquerque, NM: University of New Mexico Press.

Young, R., Morgan, W., and Midgette, S. (1992). *An Analytical Lexicon of Navajo*. Albuquerque, NM: University of New Mexico Press.