

## **The California Athabascan Website: [home.att.net/~wranders](http://home.att.net/~wranders)**

[now at <http://www.turtlenodes.com/calath/>]

Sally R. Anderson

Indiana University

[Published in: Tuttle, Siri. 2003. Proceedings of the 2003 Athabascan Languages Conference: Arcata, California, June 5 - 7, 2003. (Alaska Native Language Center Working Papers, 3.) Fairbanks: Alaska Native Language Center, Univ. of Alaska. viii+93pp. Updated slightly, 2015]

### **Introduction**

In this paper I will be discussing, in brief, the processes involved in creating usable materials for web dissemination from the various early 20th century linguistic and cultural data recorded for the Cahto language in specific and, to a lesser extent, Eel River languages as a whole. Special attention will be paid to the process of working with these original source materials and evaluating them. In the process, I will be demonstrating some of the web resources created from the Cahto and Eel River data.

The general goals of this presentation are 1) to describe the use of the original source materials, and 2) to impress upon the audience (community leaders, language program leaders, and linguists) that disseminating linguistic and language pedagogy materials on the web lies within anyone's means.

### **Outline**

Over the course of the last seven years the author has been engaged avocationally in working with historical materials on the California Athabascan languages in general, and the Cahto language in particular, with the general goals of 1) repatriating heritage language materials to the native peoples, especially the children; and 2) disseminating these materials to wider audiences of scholars and interested lay people. Such dissemination of materials from an under-documented and nearly-extinct language entails several considerations: interpreting the often poorly recorded and organized information, converting it into a format that is usable by language learners and scholars of all ages and linguistic sophistication, and distributing it in as inexpensive and easily accessible a manner as possible.

Interpreting the original information has involved determining the approximate phonetics and phonology of the language as a whole, and the phonemes of the individual words, based primarily on transcriptions in the original sources that are generally fairly poor. In many cases determining the precise meaning of words has also been difficult, especially with regard to names of plants, animals, and culture items which are often given very rudimentary glosses in the original sources.

Converting the Cahto data into a more usable format has most notably involved the creation of a (provisional) practical orthography and attempting to implement some reasonably consistent conventions for representing and delimiting words and boundaries. Once the phonemics were understood this became a relatively simple endeavor, largely informed by common practices used for other Athabaskan languages and the desire to make the language easily usable without the need for diacritics or special characters.

The primary medium chosen for dissemination has been the World Wide Web, as it is by far the least expensive and most readily accessible means of "publication" available to reach a wide audience. Many of the online materials have, however, been designed to be printer-friendly so that the Cahto leaders involved in education can easily produce hard-copies to share with their students.

### **Interpreting the original materials**

As the Cahto language is nearly-extinct,<sup>1</sup> almost all research is necessarily of a philological nature, working with source materials dating almost exclusively from first half of the twentieth century. For the purpose of this paper, only the interpretation of phonetics/phonology and semantics will be discussed.

#### *Interpreting the phonetics/phonology*

There are essentially four types of sources for the interpretation of Cahto phonetics and phonology: descriptions of the sounds, transcriptions, an instrumental study conducted by Goddard, and evidence from comparison with other Athabaskan languages. Goddard (1909 and 1912) gives quite detailed descriptions of each of the sounds that he could distinguish, and gives clear indications when he is unsure about distinctions that he may not be able to hear.<sup>2</sup> Other descriptions of the sounds are of an extremely cursory nature, and consist more of simple keys to the particular author's orthography. The largest body of transcriptions is the 111-page set of texts recorded by Goddard (1909) from Bill Ray. Other transcriptions are of individual vocabulary items or, especially in Loeb (1932), short phrases. Of the transcriptions, only those of Harrington (1942), Sapir (1907-8), and Goddard (1909, 1912) are reasonably accurate, with Harrington's and Sapir's being far and away the best, especially with regard to the apical vs.

---

<sup>1</sup> There is one woman in the community who still knows a great many individual words and some phrases of the language, as taught to her by her great grandfather Bill Ray and her grandfather Gil Ray, both of whom spoke fluent Cahto. She has an authentic pronunciation, as Victor Golla (pers. comm.) informed me several years ago and I was able to confirm a few days after this conference when I was finally able to meet her, and an ear for the correct sounds of the language which nicely confirms such difficult points as the laminal vs. apical contrast between the two sets of "alveo-palatal" region sounds, and that the reduced vowels are not, in fact, neutralized but merely happen to fall into the very wide range of the English shwa sound and were therefore generally heard by English speakers as if they had been neutralized.

<sup>2</sup> In this regard particularly note his descriptions of "c" (1912, p.8), the "palatals" (1912, p.11), "g" (1912, p.11), and "affricatives" (1912, p.13).

laminal<sup>3</sup> contrast and glottalization. The only known instrumental study of Cahto phonetics was done by Goddard using the Rousselot kymograph device, which basically recorded oral air pressure but also had attachments for measuring nasal air pressure and the movement of the larynx. The kymograph tracings, which are published in Goddard (1912), are excellent for demonstrating duration, laryngeal states, amplitude/power, and degree of closure. Unfortunately, they are not useful for resolving questions about place of articulation or laminal vs. apical distinctions. Comparative evidence has been most useful for separating sounds that have been neutralized in the transcriptions by the English speakers transcribing Cahto, once it has been established from either the better transcriptions or the kymograph tracings that such a contrast exists. This comparative evidence has been especially helpful in distinguishing the reduced vowels (/a/ [ɐ], /o/ [ə, ʊ, wə], /i/ [ɨ, ə, ɪ]) and the laminal and apical series of alveo-palatal sounds. For certain words, it may now be impossible to determine the exact phonemic make-up, due to a combination of poor transcription and a lack of clear cognates in other languages. This is especially the case with words that are recorded only in Essene (1942), Loeb (1932), or Merriam (1979).

### *Interpreting the semantics*

A number of words given in the various sources for Cahto have exceedingly sketchy definitions. Many of the most difficult words are names of unidentified plants (e.g. naa'aalee [in a list of edible bulbs]), animals (e.g. t'ooshook, "a bird"), or culture items (e.g. kaa', "headdress"). In many cases, it has been possible to find better definitions in other sources that have far inferior, but still recognizable, transcriptions of the same word. For example, some animal and plant terms can be identified based on Merriam (1979). In other cases, the word has cognates in the neighboring languages. In particular, several of the plants have been identified from cognates in Wailaki and Sinkyone. For culture items, Loeb (1932) has been particularly useful. Cognation has not usually been of much help in the case of culture items, as Cahto culture is quite different from that of the other Eel River Athabascan peoples.

### **Converting to a usable format**

The next major step was to try to create a more usable format for presenting the language materials. This has largely entailed developing a provisional practical orthography and some reasonably consistent methods of marking boundaries.

The development of this practical orthography was guided by the following general principles:

1. It must be reasonably intuitive to English speakers (especially those with some exposure to Spanish or other languages with more phonemic orthographies than English).
2. It must use only keyboard characters, in order to be easily usable in electronic media.

---

<sup>3</sup> This contrast is possibly essentially postalveolar vs. palatalized alveolar, as perhaps implied by Sapir's (1914, p.288) use of the character ś.

3. There should be a degree of consistency in the representation of sounds based on place and manner series.<sup>4</sup>

4. It should encode all and only phonemes.<sup>5</sup>

5. It should use, whenever possible, conventions that are common in the practical orthographies developed for other languages, especially fellow Athabaskan languages.

With the above principles in mind the orthography was then developed taking inspiration from the orthographies of other Athabaskan language (e.g. the use of di- and tri-graphs for fricatives, glottals, and affricates; double-vowels for long and full vowels) and, to a lesser extent, Goddard's transcriptions (e.g. tc and dj, uu, and l'). The original orthography development was done in consultation with Victor Golla and Ruth Bennett of Humboldt State University and Douglas Parks of Indiana University, with some involvement since 1996 by various members of the Cahto community.

## **Dissemination**

In the earliest stages of this project, the World Wide Web was chosen as the primary means of dissemination of materials. This decision was based primarily on considerations of cost<sup>6</sup> and accessibility to non-linguists<sup>7</sup>. In this section will be discussed both general considerations for putting language materials online (e.g. image and sound formats, special characters, programming languages) and some of the particular materials made available on the Cahto Language Homepage.

### *General considerations*

#### Image formats

There are two main formats used for images on the web: Graphical Image Format (GIF) and Joint Photographic Experts Group format (JPEG). Many of the graphics programs used on computers will by default produce other formats (e.g. bitmap, PhotoShop) but these programs usually have the option to save as a JPEG or GIF instead. In general, JPEG's are larger files, but higher quality images, and are therefore best for photographs and important images. As highly

---

<sup>4</sup> For example, all glottalized sounds are marked with apostrophe; the laminal postalveolar series is written in the same way as English (ch, j) while the apical series is written with the more unusual symbols from Goddard (tc, dj).

<sup>5</sup> This has been relaxed in the case of /n/ (with allophones written as n and ng) and /tɬ/ (written initially as tɬ' but finally as l', following in Goddard's practice).

<sup>6</sup> The only costs incurred for web services to date are the regular monthly charges from Internet Service Provider AT&T Worldnet, which would occur regardless for personal Internet access and email. All of the other web services used, including the webspace on Crosswinds.net, the webspace and perl programming environment on ProHosting.com, the Tiger Mapping service from the U.S. Geological Survey, and the original website on GeoCities.com, have been completely free of charge.

<sup>7</sup> The tribal center and the local schools all have computers for the use of students and teachers. On an informal note: I wish you could all visit the tribal center after school. It's really wonderful to see all the kids come in and just hang out there having educational programs, working on the computer, and just hanging around and playing. There's a really nice community atmosphere there.

compressed images, GIF's are very small files, but very poor quality images, best used for thumbnails and icons.

### Audio formats

The sound files on this website are all in Real Audio (RA or RM) format, which is a highly compressed format. This has the advantage of producing very small files that can be downloaded quickly even on a modem connection, but the disadvantage of being somewhat lower quality.<sup>8</sup> The other main formats for audio files online are WAV and MP3. WAV files are much larger than Real Audio files and require considerable storage space and download time. The size differences between WAV and Real Audio files are much greater and the quality differences somewhat less than the differences between the image formats mentioned above. MP3 format has been developed since many of the sound files for the site were created, so it has not yet been considered. Both Real Audio and MP3 formats require free players that the user can download and install easily. The ubiquity of these two compressed formats means that most users will already have the players installed anyway.

### Special characters

In the case of Cahto, there is no need for special characters except in the representation of historical citations of words, as the practical orthography has been deliberately designed to use only basic "typewriter" characters. Many other Athabascan languages require special characters, particularly slash-l, haček, and markings for tone and/or nasalization on vowels. Thanks to the spread of the Unicode standard it is becoming easier and easier to include special characters on web pages. The special characters are encoded as "entities" in the html code for the web page and an appropriate font is selected. For the benefit of fellow Athabascanists the following list of common entities is included.<sup>9</sup>

---

<sup>8</sup> The free utility from Real.com used to convert wav files into Real Audio files allows one to select the desired quality of the final audio file by selecting the connection speed (28Kbps modem through 512Kbps DSL) of the target audience. Even at the highest quality setting the file will be considerably smaller than the original wav file.

<sup>9</sup> For other characters please see the Code Charts online at [www.unicode.org](http://www.unicode.org).

acute accent ´ &#X0301;	macron ¯ &#X0304;	slash-l ł &#X0142;
grave accent ` &#X0300;	ogonek ˛ &#X0328;	eng ŋ &#X014B;
circumflex ^ &#X0302;	haček ˇ &#X030C;	inverted-R Ɔ &#X0281;

The accent characters simply follow the character on which they should appear, for example "a#X0301;" will display as an acute accented a.

## Programming languages

Two main programming languages are used in this website: perl and JavaScript. When this project was first developed, many people still had browsers that could not run programs in JavaScript or other client-side<sup>10</sup> programming, for which reason the author was reluctant to make use of such programs. This situation has thankfully reversed itself over the past few years, so some of the more recent pages now make use of some JavaScript bells and whistles. Unfortunately, the main browsers (Netscape Navigator and Internet Explorer) have developed mutual incompatibilities in how they interpret certain aspects of these programs, so it is often difficult or impossible to program the same effect for both browsers. In the case of a project with actual funding and a JavaScript programmer, it would probably be necessary to invest the extra time and money to program for both environments.

Aside from the compatibility issues, there are major differences in the strengths of each language. The perl language was developed by a linguist and works extremely well for almost any form of linguistic data processing, thanks largely to the extremely flexible regular expression<sup>11</sup> search and replace features built into the language. As an example, one can define classes for consonants and vowels in a particular language, and use them in regular expressions to search through an entire text or dictionary for certain syllable structures. Regular expressions are frequently used on this site to convert from one orthography to another and to search through the the lexical and text databases.

The disadvantages of perl are that it is not able to do the kinds of real-time modifications of web pages of which JavaScript is capable, and that the host server must support perl and cgi.<sup>12</sup> Another excellent server-side programming language for such websites is PHP, which is very similar to perl, but somewhat more user friendly. Many web-hosting services that support cgi and perl also support PHP. JavaScript allows one to modify certain aspects of the display of a web page instantaneously. This is very useful for such things as checking answers to test

---

<sup>10</sup> "Client-side" refers to programming that is interpreted by the user's own internet browser and therefore is dependent on the quality of each individual user's browser, whereas "server-side" programming is interpreted on the computer hosting the website and thus works consistently regardless of what browser the user has.

<sup>11</sup> "Regular expression" refers to what is essentially a very advanced version of wildcards (e.g. the old DOS \* and ?). But instead of just the simple wildcards, one can create basically any class of characters that one can think of and then use those classes to search for particular patterns.

<sup>12</sup> Unfortunately, very few free web hosting services support cgi and perl as doing so can entail some security risk to the host computer. The site currently uses Prohosting.com's free cgi hosting and in the past has used Virtual Avenue as well.

questions instantly and providing clues for incorrect or incomplete answers, changing images, and providing pop-up definitions for words.<sup>13</sup> There are now modules for JavaScript that emulate the regular expression capabilities of perl, so it is a much stronger candidate for use on language websites than it once was.

### *Particular materials on the Cahto website*

#### Searchable databases

Two main database searches are currently provided on the website, a lexical and textual database search and a text display. Both searches are written in the perl language. The data are contained in simple delimited text files that any database program can easily create, each with separate configuration files identifying the field names and characteristics for the program to be able to interpret and display each field correctly. In principle, it is possible to search and display any flat-file database<sup>14</sup> using the database search. Most customization simply involves setting up new field types with the appropriate HTML code for the display style (e.g. fonts, formatting for hyperlinks, images, audio files) and whatever perl programming is necessary to convert the data as required (e.g. constructing hyperlinks, converting orthographies). For now, the lexical database contains a separate listing for more-or-less every token given in the various sources for the languages, although (for Cahto at least) headwords have been created and will soon be used as main entries in a more dictionary-like search. Clicking on the Cahto "Practorthog" essentially produces a concordance page of all occurrences of that exact word (letter for letter) in the text database. The author is working on creating a Cahto morphological processor (also written in perl) that may eventually make it possible to search for any form of a given verb, rather than the exact form given.<sup>15</sup>

#### Custom maps

Probably the most useful free web service the author has encountered during the course of this project has been the Tiger Map Server from the U.S. Geological Survey. This service allows one to create a text file of place names, types of markers (e.g. red arrows), and latitude and longitude coordinates<sup>16</sup>; and to use those data to generate custom maps online. From the URL in the hyperlink the programmer can then control the default values of what map layers (e.g. streets,

---

<sup>13</sup> See the "Experimental Learning Materials" section of the Cahto website for examples of JavaScript effects.

<sup>14</sup> "Flat-file" refers to a database that is contained within a single table, as opposed to multiple related tables. It is actually fairly easy to modify the program to use multiple tables with external links to another search. This is how the concordancing links from the lexical database to the texts and the dictionary look-up links from the text display work.

<sup>15</sup> "May" because the search may become so complex as to slow down the search to speeds unacceptable to most web users. It may require running the morphological processor on the text to invisibly tag each word with its Cahto headword.

<sup>16</sup> Due to the potential for vandalism it is probably best to leave off actual village sites, or at least those that are not already well recognized and officially protected. For this reason the Cahto and Wailaki maps only show the names of land features (creeks, hills, valleys, etc.) and the name of the town of Laytonville. There will soon be a more detailed password-protected version including the village sites that will only be available to the Cahto tribe.

parks, waterways) appear and what location and size of area to display. Then, from Tiger's online interface the user can choose different layers, zoom in or out, and move to a different center point.

### Alphabet book

The Cahto Alphabet<sup>17</sup> Book represents the author's most recent attempt at presenting the sounds and orthography of the language in a simple, user-friendly format. This site makes use of HTML frames<sup>18</sup> to display the whole alphabet as a clickable table of contents on the left side of the page, the current page on the right, and a space for footnotes on the bottom of the page.

### Coloring book

Currently the Coloring Book Dictionary is more of a proof-of-concept than a real work. It illustrates several potentially useful web-programming techniques, however. It makes use of fairly simple JavaScript routines to do several things: 1) change from one image (a black-and-white outline) to another (an example of the same image already colored); 2) display a small pop-up definition and pronunciation guide for the word over which the mouse is currently positioned<sup>19</sup>; 3) hide the English translation of the text so that the page appears in Cahto alone<sup>20</sup>; and 4) display a pop-up glossary of all the words used in the text. The information for the pop-up definition and glossary is all contained in a data array in the HTML for the individual page, so the JavaScript program does not need to access any other files.

### Coast Trip Vocabulary

The Coast Trip Vocabulary was originally created for use on the tribe's annual walking trip to the coast, reenacting their traditional journey to gather various seafoods. This was designed to be printed out and taken on the trip, but it also has a simple JavaScript effect designed to be fun (hopefully) for kids: when the mouse is positioned over the word for a plant or animal in the vocabulary list, the picture on the page changes to illustrate that item.

## Afterword

---

<sup>17</sup> Note the nonstandard use of the term "alphabet." This is used for the sake of familiarity to non-linguists. To the author much of this site seems like a fine line between dumbing down and too much information (TMI), which is probably inherent in a project with such a varied target audience.

<sup>18</sup> Note that the World Wide Web Consortium (W3C) now recommends using the division element with Cascading Style Sheets (CSS) instead of frames, but the author has found that some older browsers do not properly display the divisions. All even marginally recent browsers should display frames correctly, so for the near term it is probably easier to continue using frames. Both are about equally simple to program.

<sup>19</sup> This effect only works properly in Internet Explorer, so boxes at the top and bottom of the page in which the same information appears for the benefit of users of other browsers. It would be possible to program this so that the boxes only appear for Netscape Navigator, in other words, when they are actually needed.

<sup>20</sup> Unfortunately this also only works on Internet Explorer.



This website is obviously in a relatively primitive state in terms of presentation design. However, it serves as a demonstration of what can be done fairly easily with little or no funding, and in one's spare time.<sup>21</sup> This is the main point that the author wishes for linguists and community leaders to take away from this too brief presentation. Please feel free to contact the author (sally@turtlenodes.com) if you are interested in pursuing similar projects or making use of the technologies mentioned herein.

[Afterword, 2015: With changes in web technology, many of the particulars have changed since this was written. For example, the tiger mapping service has gone defunct, the crosswords with Word Splash reference an old version of JavaScript and fail, and I've replaced the realaudio sound files a couple of times, first with wav's and later with mp3's. The whole site needs a good working over, and to be made compliant with Responsive Web standards to make it better viewable on the smaller devices of today. However, the essential fact that the creation and dissemination of useful language support and revitalization materials is doable within any budget.]

## **Bibliography**

Essene, Frank J. Round Valley. 1942. Anthropological Records 8(1).

Goddard, Pliny Earle. 1912. Elements of the Kato Language. UCPAAE 11(1).

\_\_\_\_\_. [1902-03; 1922]. Field notes in California Athapascan languages. MS, APS. [18 notebooks. Pomo, Hupa, Kato, Wailaki, Sinkyone, Tolowa, and Nongatl texts; ethnographic and ethnohistoric materials.]

\_\_\_\_\_. 1909. Kato Texts. UCPAAE 5(3).65-238.

Harrington, J.P. 1942. Notes on Kato. Harrington collection microfilms at the National Anthropological Archives. Volume II, reels 003, 004.

Loeb, Edwin M. 1932. The Western Kuksu Cult. UCPAAE 33(1).

Merriam, C. Hart. 1979. Indian Names for Plants and Animals among Californian and other Western North American Tribes. Socorro, NM: Ballena Press.

Sapir, Edward. 1914. "Notes on Chasta Costa Phonology and Morphology. Anthropological Publications of the Univ. of Pennsylvania Museum 2(2).271-340.

\_\_\_\_\_. 1907-8. 13 pp. of Cahto transcribed by Sapir in winter or spring of 1907-8. [Included in Goddard, Pliny Earle. Athapaskan Field Note and Manuscript Collections, 11. Kato Linguistic

---

<sup>21</sup> Please note also that relatively little of the author's time on this project has been spent on the actual programming.

Miscellany (with Edward Sapir). Item #12.11 in the UC-Berkeley Anthropology Department Archives (housed in the Bancroft Library).]

Image 1:

[Now: <http://www.turtlenodes.com/calath/cgi-bin/dbsearch.cgi?databasename=eelriver>]

Image 2: [Now: <http://www.turtlenodes.com/calath/castoryindex.html>]

Image 3: [Now: <http://www.turtlenodes.com/calath/caband.html>]

Image 4: [Now: <http://www.turtlenodes.com/calath/cahtoalphabetbook/AlphabetBook.html>]

Image 5: page from coloring book: [Now:

<http://www.turtlenodes.com/calath/coloringbook/blackoak.html>]

Image 6: words for coast trip page: [Now:

<http://www.turtlenodes.com/calath/coasttrip/coasttripvocabx.html>]