NATIONAL SCIENCE FOUNDATION

NATIONAL ENDOWMENT FOR THE HUMANITIES

Documenting Endangered Languages

SAMPLE APPLICATION NARRATIVE

This sample narrative conforms to a past set of guidelines.

Please consult the guidelines for the current competition on NSF's website at:
http://www.nsf.gov/pubs/2005/nsf05590/nsf05590.htm

Colville-Okanagan Dictionary, Reference Grammar, Texts

PROJECT SUMMARY

GOALS. For the past thirty-six years I have studied Colville-Okanagan (Cv-Ok), a Salishan language of northeast Washington and south central British Columbia, and its literary forms, with increasing comprehension and control of the data. I now propose to dedicate ten and one half full-time months to the project to

- store in digital format (.wav files) all Cv-Ok audio tapes in my possession;
- transform the data currently digitally stored in band format into dictionary entries to be added to those of the (published) first edition of the dictionary;
- integrate into a reference grammar my many papers and drafts on grammatical topics;
- integrate into anthologies the many texts that I have been editing, analyzing and interlinearizing over the span of my professional career;
- produce more pedagogical materials to promote literacy in the language.

The products of my work will be

- a new dictionary and reference grammar of the language, in printed book format, the former also on CD-ROM, Unicode-compliant;
- two volumes of texts, in printed book format, also Unicode compliant, and on CD-ROM;
- audio CDs as companions to the volumes of texts.
- a web-based set of Cv-Ok language lessons
- a web-based set of Cv-Ok readings
- a web-based Cv-Ok dictionary

INTELLECTUAL MERIT AND BROADER IMPACTS. The documents projected to result from this effort include a dictionary and reference grammar of Cv-Ok, and universally accessible Cv-Ok texts of all genres. The intellectual merit of the project, therefore, is the deeper knowledge and better understanding, not only of Cv-Ok, but of Salish languages in general, that would follow from this comprehensive documentation.

The fact that Cv-Ok is already relatively well-documented and that the interest in language-retention in the community is high[1] suggest that the project would have the broader impact of increasing the likelihood of revitalization of the language. In this sense, then, the project is as valuable as, or perhaps even more valuable than, the documentation of a language with only a handful of elderly speakers, and no realistic chance of being revitalized. Because the quality, sophistication and usefulness of the documentation is directly proportional to the investigator's knowledge of the language and descriptive skills, and because a comprehensive documentation is the best tool to revitalize the language, I think that this project meets the criteria spelled out in the Program Solicitation. In the narrative that follows I will try to explain as clearly as possible that this is so, and how I propose to conduct my work.

---

[1] A few elders know the language well, and several hundred individuals understand it. Czaykowska and Kinkade (1998) estimate the number of speakers at "500?", Ethnologue at 800, and I at "perhaps as many as one thousand" (Mattina 1999)

My study of Colville-Okanagan language and literature dates back to 1968, and has always remained the focus of my scholarly interests.

WORKS PUBLISHED AND IN PRODUCTION. In 1977 I received NSF and NEH support to compile a dictionary of the language. That project produced the *Colville-Okanagan Dictionary*, released as Volume 5 of the *University of Montana Occasional Papers in Linguistics*, a series co-founded by Timothy Montler and me. The *Colville-Okanagan Dictionary* received excellent reviews in *IJAL* (Vol 57, pp. 402-405, 1991) and in *International Journal of Lexicography* (Vol 9, pp. 359-364, 1996), and an equally positive brief mention in *Language* (Vol 65, pp. 433-434, 1989). Equally important is the fact that, not only scholars, but also many Okanagan speakers/learners make regular use of the dictionary and find it useful.

Valued though it is by professional linguists and students of the language, the *Dictionary* remains a preliminary edition based on data from some texts. The reviewer of *IJAL* also wrote that "if there is another edition ..., I would recommend to the publishers that they include a fulsome grammar section" (p. 361). It is my intention to produce a dictionary that includes, or is accompanied by, a reference grammar of the language. Because currently the database that contains the material to be added to the dictionary is the size of the *Dictionary* itself, inserting these data into the *Dictionary* would double the latter's size.

I have produced other works of documentation while teaching full-time at the University of Montana. One such is the edition I, together with co-translator Madeline DeSautel, prepared of Peter J. Seymour's long narrative "The Golden Woman", published by the University of Arizona Press.[2] This bears on the present proposal insofar as it represents the sort of textual analysis I have been conducting, and that eventually feeds the dictionary. Reviewer Eric Hamp referred to it as "a model specimen of published text responsibility" where "the strict formal linguistic accounting is not shirked in the slightest" (*IJAL* 59, p. 96, 1989). Dozens of other texts, by Seymour and others, of all styles, from oratory to ordinary conversation, and from myth telling to joke telling, are at various stages of preparation, many already interlinearized and with glossaries, all others transcribed and translated but not yet parsed, form the huge database that feeds the dictionary and the grammar.

Academic year 2000-2001 I completed another volume, a collection of myths narrated by Dora DeSautel. Endorsed by the Colville Tribal Council, and consisting of a biographical note by Adrian Holm, granddaughter of Dora DeSautel, an introduction that contextualizes the narratives and outlines the grammatical features of the language, eight interlinearized texts, a glossary, and a CD (8 tracks, 66:39 m:s), this volume has been published as the 15th volume in the *UMOPL* series. The interlinearization schema I use in this volume is technically superior to that of *The Golden Woman*, and it is the schema I propose to adopt in the forthcoming anthologies.

A collection consisting of eight texts narrated by Madeline DeSautel, interlinearized, and accompanied by an introduction and glossary, is currently in the hands of a grandson of Mrs. DeSautel, who is writing the biographical sketch of his grandmother. Its accompanying CD has 8

---

[2] Czaykowska and Kinkade (1998) refer to this work as "in a category by itself" (p. 47).

tracks, or about 68 minutes of spoken texts. The volume should be ready for submission to the Colville Tribal Council for its *imprimatur* soon.

LANGUAGE PROGRAMS AND PEDAGOGICAL MATERIALS. I have helped establish and maintain language programs in Cv-Ok speaking communities on both sides of the border. Two years (summer 1992 through summer 1994) I was in residence at the En'owkin Centre, Penticton, BC, where I trained elderly speakers and younger apprentices in Cv-Ok literacy, and I have been a consultant to the Language Program of the Colville Confederated Tribes since its inception in 1994. We have produced sets of language lessons, and a variety of readers, computer language games, and other teaching materials.

Collaboration between the applicant and both the En'okin Centre and the Language Program of the Colville Tribes has been and continues to be active and productive. All materials thus far produced, by me individually, or by us collectively, have been approved for general use by the governing bodies of the two organizations. We follow established practices when deciding, for example, what reading materials to publish. All autobiographical texts, for example, are first approved by the author('s representatives), and then by a Culture Committee. We follow similar procedures to select and publish mythological and other traditional texts. Colville and Okanagan language apprentices seek my instruction, and continue to make steady progress in the production of more language materials.

As I have studied the language, I have witnessed important sociolinguistic changes. For example, while thirty five years ago the reciprocal term of address between members of the opposite sex was *sǝntxʷús*, and it was appropriate only for males to address other males with the term *sĺax̌t*, nowadays the latter term has replaced the former wholesale. Similarly, dozens of lexical items routinely used by speakers in the late 60s, are no longer in the active vocabulary of contemporary speakers, and are not recognized by many. Synthetic verb constructions with incorporated nouns (so-called "lexical affixes") are being replaced by analytical near-equivalent calques. These and other related sociolinguistic facts and phenomena are being and will be properly documented and reported.

CORPUS

Having devoted nearly all of my research time to the documentation of the Colville-Okanagan language and its literature, I have in my possession a large, significant corpus. The so-far published texts, for example, correspond to no more than a fifth of what I have collected.

The data I have collected include: thirty hand-written notebooks (about 4,000 pages of transcriptions and notes), about 150 45-minute tapes gathered during my field trips 1968-1975, and two dozen tapes gathered between 1976 and now.

DATA ON NOTEBOOKS. My notebooks contain transcriptions of texts, paradigmatic and elicited material, and grammatical, cultural, and contextual notes, most done with the help of Madeline DeSautel. The texts are of various genres: myths, historical and autobiographical accounts, ethnographic texts, acculturated and borrowed texts, conversations, and jokes and anecdotes. They comprise more than half of my field notes.

I have stored on disk and checked for accuracy approximately one half of all the texts I recorded, but I need to type and parse another half. The parsing produces texts that are ready for publication in interlinear format.

DIGITIZED DATA.   I estimate that five megabites of files were typed directly on disk during my two year assignment at the En'owkin Centre; and three megabites of .txt files nave been typed directly on disk during my numerous visits to the Colville Language Preservation Program (Omak and Nespelem, Wa). The earliest of these data are stored as DOS text, the more recent data as Word documents.[3] These data consist of transcriptions of traditional narratives, contemporary re-translations of traditional narratives, Colville Tribe-commissioned works (such as a health book that discusses traditional and modern remedies), word-lists, and materials for children's books.

WORK TO BE DONE

The work divides in these categories:
- the digitizing of tape-recoded texts;[4]
- the typing of the texts on computer;
- the parsing and interlinearization of the texts with morpheme analysis, translations, and glossary for inclusion in the anthologies;
- the processing of texts for the extraction of lexical and exemplificatory materials;
- the integration of the new and additional data into the *Dictionary*;
- the write-up of the reference grammar;
- the selection and adaptation of data for use in pedagogical materials.

TYPING TEXTS. I plan to have two volumes of texts completed by August 2006. One volume will include all the myths I have collected. The other will include historical accounts, ethnographic texts, acculturated and borrowed texts (for example, various versions of the *Grateful Dead* (Aarne-Thompson's type 505-508), conversations, and jokes and anecdotes. Each volume will follow the pattern I set in the recently-published collection of texts *Dora Noyes DeSautel ɬaʔ kɬcaptíkʷɬ*, and include biographical notes about each author; relevant introductory, literary, linguistic, and comparative materials; abbreviations, symbols and explanations thereof; English translations; interlinearized original texts; and a glossary. In this model, texts are broken up into sequentially numbered discourse-cohesive units that match, roughly, sentences, interlinearized as described below. Unfinished utterances, reconstructions of unintelligible utterances, interjections and responses of members of an audience, are all specially marked, as are editorial insertions. The glossary contains a list of all the morphemes found in the texts, their glosses, the skeletal make-up of the lexical roots, and the sentence number of their first occurrence in each of the texts.

About 1,200 notebook pages remain to be typed. This typing is an ongoing activity, and I hope to have a fourth of these typed by the end of May, 2005.

---

[3] See below for a discussion of the related technical and methodological issues.
[4] For this purpose I use the versatile and powerful audio editing and restoration software **Diamond Cut DC Six**.

PARSING (INTERLINEARIZING) TEXTS. To parse a text means to analyze each word of the text into morphemes. I output the parsed text in the following format:[5]

ixíʔ	nixʷ	taʔlíʔ	ckʷúʃstsəlx				kaʔ	cx̌ast	
ixíʔ	nixʷ	taʔlíʔ	c	–kʷuʃ	–st	–s	–lx	kiʔ	c	–x̌as+t
that	also	very_much	cust-	work	-st	-3e	-pl	rel	dur	-good

That too, they work wery hard at it before it's any good;

The first line is the surface phonetic representation, as I have written it. The second line is the segmentation of each word into its (underlying) morphemes. I delimit inflectional material with hyphens, and derivational material and word-formation boundaries with plus signs—I do not segment derivational material from its stem. The third line is the gloss of each stem and each inflectional morpheme. The fourth line is the free translation.

The parsing works as follows: the first time I encounter a word, I segment it into its stem and inflectional morphemes. I store that information for future use, and the byproduct of this function is the creation of a database that includes the stems and inflectional morphemes found in the texts. The larger the text database the higher the probability that the list of morphemes (and words) approaches completeness.

IMPORTING TEXT MATERIAL INTO THE DICTIONARY. Other morphological material, not only stems and inflectional affixes, needs to be identified and processed. The stem *kʷəlcəncút* 'cook', for example, consists of the three morphs √*kʷʃ* 'make', =*cn* 'mouth', and –*cut* 'reflexive', and I want each of these morphs, not only the entire stem, to be an entry in my database. I have written user-programmed procedures that allow me to extract any stretch of data from the text material, insert it into an appropriate entry, along with the relevant context of each. In sum, the dictionary will have an entry *kʷəlcəncút* 'cook', with its gloss, definition, grammatical description and exemplification, and it will also have entries for =*cn* and –*cut*, each with its own gloss, definition, grammatical description, and list of exemplificatory forms.

ORGANIZATION OF THE DICTIONARY. I plan to do some reorganization of the lexical entries in the new dictionary, as follows. The preliminary dictionary contains five kinds of entries: roots, particle skeletons, affix skeletons, full affixes, and words.[6] Words listed as main entries are morphophonemic analyses, and the reader is referred to the root (or the affixal skeleton) for the full entry. In the new version of the dictionary I provide definitions, grammatical identifications, and exemplification under each stem, while retaining the cross-reference to the root or skeleton. By way of example, here is the root entry √x̌ẅ:

√x̌ẅ₂ *dry.* •Itr x̌ẅ+x̌aẅ+t *dry* (OkB). x̌ẅ+aẅ *get dry (inch).* x̌ẅ+x̌əẅ+aẅ *pl get dry* <x̌ẅ+aẅ iʔ sɬiqʷ *The meat has dried.* EN> **i-ks+x̌ẅ+áẅ** *i-fut dry* <x̌lap lut ks-x̌ẅ+aẅ-s iʔ sks kʷalts iʔ l sqiltks *In the morning the sweat on his body won't be dry.* PS> •Tr x̌ẅ-nt-ím *dry st* <x̌ẅ-nt-is-lx x̌ẅẅalqʷ, x̌ẅntislx məɬ ʔipslx *They dry the sticks, then they whittle and wipe them.* EN> **x̌w–ɬt–im** *dry st of sb's* <x̌ẅ+x̌ẅ–ɬt–is

i? tkəɬmílxʷ i? sɬiqʷs *She dried the woman's meat.* DD> **c-x̌ẇ-st-im** *cust dry.*
<**c-x̌ẇ-st-is i? l sɬ kʷ kʷaʕst** *She always dries them in the morning.* MD>
- Cplx **x̌w=cn** *dry mouth*; **x̌ẇ=ilt** *dry berries*, **x̌ẇ=ilt-m** *to dry berries*;
  s+x̌ẇ=ilt+[t]n *berry drying rack*; **x̌ẇ+íɬċa?, s+x̌ẇ=íɬċa?** *dry meat*,
  s+x̌ẇ+íɬċa?+m *to dry meat*; **x̌ẇ+ẇ+alqʷ** *dry sticks*; **n+x̌ẇ=ink** *dry inside*;
  n+x̌ẇ=úla?xʷ, n+x̌ẇ+ẇ=úla?xʷ *dry pool, land*; **s+n+x̌ẇ+min** *drying shed or
  rack*; **s+n+x̌ẇ=ilt+[t]n** *drying place*; **s+n+x̌ẇ=íɬċa?+tn** *drying meat/fish rack*;
  t+x̌ẇ+x̌ẇ+us *dry eyes*; **t+x̌ẇ=ilt+[t]n+m** *to dry meat*.
- Idiom **x̌ẇ+x̌ẇ-ilx** *be in a bind*; **x̌ẇ+x̌ẇ=us** *want an excuse*.
- xr √**x̌ẇ₂, x̌iẇ** *raw*.

Here we see
- the root listed by its consonantal skeleton under the root sign.
- a gloss, followed by
- a list of intransitive forms based on that root, each with its grammatical label (most are
  not given in the preliminary edition), and exemplified in context;
- transitive forms, organized by the form of the transitivizer, and again exemplified. The
  second part of the entry, starting a new indented line, lists
- complex forms (root plus lexical affix) and their glosses. The third part lists
- other forms related to the root. In this case the relationship is identified as *idiomatic*.
  Another part of the entry, where appropriate, lists
- compounds;
- cross-reference is made to other entries to suggest possible connections, and
- sources, which are identified when not default.

Other conventions are spelled out in the introduction. Here, for instance, the form **x̌ẇ+x̌aẇ+t**
dry is identified as coming from Randy Bouchard. It is not exemplified in context because I have
not yet encountered such an instance, but the dictionary will have a main entry for the word,
which simply refers the reader to √**x̌ẇ₂**: **x̌ẇ+x̌aẇ+t** *dry*. See √**x̌ẇ₂**. This practice will extend
to all the words listed in the first part of the entry, intransitive and transitive.

 All the complex forms listed in the second part of the entry are main entries themselves,
each with examples. Thus, for example, **x̌ẇ+íɬċa?** will have the following entry:

x̌ẇ+íɬċa? *dry meat*. Itr **s-x̌ẇ+íɬċa?** *dry meat*. <**x̌s?iɬn axá? i? s+x̌ẇ+íɬċa?** *This dry meat
is good.*> **x̌ẇ+íɬċa?-m** *to dry meat*. <**ťi? xi?míx ca? kʷ swit ksx̌ẇíɬċa?x uɬ
x̌ẇ+íɬċa?-m** *Anybody who wants to dry meat dries meat.* CV> icptv **ks-x̌ẇ+íɬċa?-[a?]x**
<**kn ks-x̌ẇ+íɬċa?-[a?]x** *I'm going to dry meat.* perf **i-sc-x̌ẇ+íɬċa?** <**waẏ x̌ast
a-sc-x̌ẇ+íɬċa?** *The meat you dried is good.* CV>

In sum, the structural difference between the preliminary edition and the one proposed here
is that while in the former complex forms are listed and exemplified under the root,[7] in the new
edition I leave the complex form in the root entry, but without examples, and I list complex
forms with their examples as separate main entries. I will update the grammatical information,
and add new forms, around ten thousand such, each with examples. I will delete only redundant
or otherwise flawed data. The preliminary dictionary contains an average of 20 entries per page,

---

[7] This practice makes some root entries extremely long.

a total of about 6,000 entries. An additional 5,600 entries are currently in the database to which I add new material on a regular basis, and I expect this number will double when the data processing is finished. The dictionary will have more than 15,000 entries.

Write-Up of Grammar. The grammar will be a reference work with an extensive index to allow readers to find easily the discussion of topics. It will have as much information as it is known to me about the distribution and features of the various dialects; it will have chapters that cover phonology, morpho-phonology, morphology, and syntax.

I will integrate into the grammar, with appropriate emendations, the many articles, papers, drafts, and lessons I have written. I will also integrate into the grammar about seven hundred notes on grammatical topics that I store in a dedicated database. These include lists of *hapax legomena*, lexicalizations, amalgams, idiosyncrasies or exceptional cases, and a variety of pertinent information. As I parse and interlinearize texts, and as a certain form or construction catches my attention as an appropriate example for a given sandhi phenomenon, or a particular morphologic or syntactic construction, I annotate and then file such examples under appropriate rubrics such as "a-insertion", "headless relative clauses", or even "unclear". Much of this work of integration is editorial work, some of it is still a matter of (re)interpretation of the data, and, of course, there still is, and there will always be, more analytical work to be done or redone.

Phonology and Dialects. Because Cv-Ok family networks range over large portions of the Cv-Ok territory, and because individuals divide their residence between two or more family groups in different locations, not only do individuals incorporate in their speech features of several dialects, but they are also likely to code-switch in response to circumstances. I can document such code-switching, for example, for the late Martin Louie, a consummate speaker and storyteller whose speeches and narratives I have heard in person and on tape. With a primary residence in Inchelium, at the eastern end of the Cv-Ok territory, and another in Penticton, two hundred miles north and west, he was apt to speak "Inchelium" or "Penticton" as place and company moved him to do. The same applies to many speakers I have known.

A different case is a trait that, long ago, Dale Kinkade had suggested I might find, specifically, a dialect of Okanagan with retracted consonants and lowered vowels. Such a dialect would exhibit features that parallel those of Moses-Columbian and of the northern interior Salish languages. The most prominent of these features is that in these languages [š] is the pronunciation norm for /s/ alongside a contrasting /ṣ/ (with tongue root retraction). In the dialects of Cv-Ok with which I am acquainted, however, only slight palatalization may accompany /s/, and I have found no contrast between /s/ and a retracted counterpart.[8] In 1999 I wrote that "if there is widespread consonant retraction in the language then it has eluded me" (Mattina 1999).

At a gathering in Penticton in 2002, I overheard someone say [šámaʔ] for the expected [sámaʔ]. I tried to elicit some forms to establish a retracted ṣ, but, as the speaker claimed limited fluency in the language, my findings were inconclusive. Presumably a group of families that had moved from the Methow valley to the Similkameen valley several generations ago, speak a variety of Cv-Ok that regularly has [š] for /s/. Does this dialect have a contrasting /ṣ/? This is

---

[8] The affricate /c/, however, does have palatalization associated with it, especially in the northern dialects.

something that needs to be explored, and I will travel to Enderby to determine what there is to describe and add to my papers and drafts on the phonology of Cv-Ok.[9]

Morphology. The morphology section of the grammar will include inflection, and derivation and word-formation. The heart of this section is inflection, longer sections will treat the other two topics. I have outlined the inflectional categories of the language in the article "Okanagan and Salish Languages"[10], in the "Grammatical Notes" section of the Introduction to *Dora Noyes DeSautel ɬaʔ kɬcaptíkʷɬ*,[11] and in other such papers as "Imperative Constructions in Colville-Okanagan and in the Other Interior Salishan Languages." (1980), "Okanagan Aspect: A Working Paper" (1993), "Okanagan ks- and kɬ-" (1995; co-authored with N. Mattina), and others.

Syntax. This section of the grammar will include a comprehensive survey of sentence types. I have written many papers and drafts on transitive sentences, the latest of which is "The Colville-Okanagan transitive prototype" (2004). I have written much about transitive, ditransitive and applicative sentences,[12] but an integrated, comprehensive account is outstanding and forthcoming. An account of all intransitive sentence types is also forthcoming.[13]

PEDAGOGICAL MATERIALS. I have an excellent working relationship with the tribally-managed Cv-Ok language programs. I expand, refine, and add exercises to the forty lessons of the Cv-Ok primer *ta nsílxcən* as I discover the need for or as I receive requests for such (samples are posted on a web page I maintain--see below). Staff members of the Language Preservation Program in Omak, Wa, make regular use of my expertise and adopt much of what I produce for their specific teaching needs.

TIME FRAME

I plan to work full-time on the project from June 1, 2005 to Jan 15, 2006; and from May 15, 2006, to August 15, 2006. The narrative of my proposal should have made it clear that I will be working on several components of the project simultaneously. Because in the last two years I have produced two collections of Cv-Ok narratives, and single-handedly attended to nearly every detail of the production of two other *UMOPL* volumes, including the complete transliteration and formatting of the *Cowlitz Dictionary and Grammatical Sketch*,[14] while at the same time teaching three or four classes and seminars each semester, directing several MA theses, and directing the Linguistics Program, I expect that, dedicating full-time to the project, I will complete two volumes of texts and the dictionary before the term of the fellowship expires. I

---

[9] My paper "Okanagan sandhi & morphophonemics" (2000), is a recent example of information I want to incorporate in the reference grammar. But see also Mattina 1999 and Mattina 1979.

[10] Mattina. In Press.

[11] See V. Golla's comment in SSILA *Newsletter* XXI:2, July 2002, p. 12 about the "brief (but, for a Salishan language, remarkably lucid) grammatical sketch" included in this book.

[12] See, for example, Mattina 1982, and Mattina 1994.

[13] It seems to me that the discussion of intransitive sentence types in the various Salish languages has not received detailed attention, possibly because descriptions have focused on the marked category "transitive". See, for example, the statement that "A clause is transitive ... if its predicate contains a transitive ... marker, ... Otherwise, a clause is intransitive." Kroeber (1999), p. 36.

[14] See footnote 16.

expect to have a draft of the grammar ready to share with readers, if not by August 2006, then shortly after. I also expect to continue to update and test the pedagogical materials throughout the period of the fellowship.

<h2 style="text-align:center">DOCUMENTATION, STANDARDS, FORMATS, ACCESS</h2>

The work I propose to do addresses directly three of the activities listed in the DEL solicitation, specifically, to "carry out later stages of documentation including the preparation of lexicons, grammars, text samples, and databases" (2); to "provide wider access to such documents" (3); and to "develop standards ... to make this documentation ... available in consistent, archivable, interoperable, and Web-based formats" (4). The project also involves field work to confirm some phonological features of a dialect (1); and, with its pedagogical materials, trains language apprentices, and thus encourages them to study descriptive linguistics (6).

ASCII, ANSI, AND UNICODE. As a function of the history and development of digital encodings, my digitized Cv-Ok materials are of three types, and need to be made to conform to a consistent standard.

1. Legacy text files. The dictionary, produced in the mid-80s, and several same-vintage narratives, were printed with a long-extinct Sanders media 12-7, and were prepared in tagged format, so that the ASCII text includes formatting instructions enclosed in tags. The text of these materials has to be translated into Unicode, and the formatting has to be translated into current desktop publishing instructions. I accomplish the first task with Jonathan Kew's **TECkit version 2.1**, "A Text Encoding Conversion toolkit" (March 29, 2004). This efficient tool consists of two components, both fully familiar to me: one maps byte encodings to Unicode, another performs the conversion. I accomplish the task of translating the tagged instructions to current formatting instructions in two steps: first, with a simple **Spitbol** program I translate the long and cumbersome tags to simple ones; then I turn the latter into actual formatting code with Visual Basic script language instructions.

2. Word files. I have processed many Cv-Ok texts with *Shoebox*[15] (a powerful parser that does **not** support Unicode), and then imported them into Microsoft Word and formatted them. These documents are ANSI-encoded with fonts that I have customized to represent the needed Salish characters. These need to be transliterated to Unicode fonts, something I do routinely with macros that use a form of Visual Basic script language.[16]

3. Unicode files. I have started to use SIL's *Toolbox*, which does support Unicode, and which, essentially, replaces *Shoebox.* Currently I use Unicode fonts to type all texts. These involve no further processing.

---

[15] *Shoebox* has many features that aid in the lexicographic process, also in conjunction with Coward and Grimes' *Multi-Dictionary Formatter*.

[16] I have tested and used these procedures to convert M. Dale Kinkade's *Cowlitz Dictionary and Grammatical Sketch* (2004), a work given to me in WordPerfect 5.0 format.

HTML AND XML. For several years I have maintained web pages that include prototypes of Cv-Ok pedagogical lessons and readings.[17] These have slowly evolved from pages that display .pdf files, to pages that make use of special fonts, to pages that display universally available unicode fonts.[18] These pages now make use of Unicode fonts, but are still simple HTML files. It is clear that I have to advance to XML and its needed accessories, such as XSL and Javascript. I plan to do so in the near future, I hope by the middle of 2005.

CONCLUSION

Conducted as outlined, my work will produce important documents in "consistent, archivable, interoperable, and Web-based formats" that will enrich our knowledge of Colville-Okanagan and of Salish. Because they include pedagogical materials, these documents will promote the retention and revitalization of the language. I also hope that they will continue to inspire the apprentices currently working for the Language Preservation Programs to persevere in their efforts to gain fluency in the language and in the required analytical and descriptive skills.

REFERENCES CITED

Coward, David F. and Charles E. Grimes 1995. *Making Dictionaries. A Guide to Lexicography and the Multi-Dictionary Formatter.* Summer Institute of Linguistics, Waxhaw.

Czaykowska-Higgins, Ewa, and M. Dale Kinkade, eds. 1998. *Salish Languages and Linguistics. Theoretical and Descriptive Perspectives.* Mouton de Gruyter.

Kew, Jonathan 2004. *TECkit version 2.1. A toolking for Text Encoding Conversion.* http://scripts.sil.org/cms/scripts/TECkitIntro.

Kinkade, M. Dale. In press. *Cowlitz Dictionary and Grammatical Sketch. University of Montana Occasional Papers in Linguistics* 18. Missoula, Montana.

Kroeber, Paul D. 1999. *The Salish Language Family. Reconstructing Syntax.* University of Nebraska Press.

Mattina, Anthony 1979. The morphophonotactics of Colville full vowels. [Paper presented at the 14th ICSNL, Bellingham, Washington.]

Mattina, Anthony 1980. Imperative Constructions in Colville-Okanagan and in the Other Interior Salishan Languages. *Glossa* 14, pp. 212-32.

Mattina, Anthony 1982. The Colville-Okanagan transitive system. *International Journal of American Linguistics* 48: 421-435.

Mattina, Anthony, ed. 1985. *The Golden Woman. The Colville Narrative of Peter J. Seymour.* University of Arizona Press.

Mattina, Anthony, comp. 1987. *Colville-Okanagan Dictionary. University of Montana Occasional Papers in Linguistics* 5. Missoula, Montana.

Mattina, Anthony 1993. Okanagan Aspect: A Working Paper. XXVIII International Conference

---

[17] See http://www.umt.edu/ling/resources/reading1.htm and http://www.umt.edu/ling/resources/lesson1.htm.

[18] A year ago, I posted on one of the web pages I maintain, a table that includes the characters needed for encoding Salish languages, asking for comments aimed at establishing a standard encoding. This page can be viewed at: http://www.umt.edu/ling/resources/USal/UnicodeSalish.htm.

on Salishan Languages, University of Washington, August 19-21.

Mattina, Anthony 1994. ‑tuɬt, and More on Okanagan Transitives. XXIX International Conference on Salish Languages, Salish Kootenai College, Pablo, MT., August 11-13.

Mattina, Anthony 1999. Anticipatory and progressive vowel lowering in Interior Salish, with notes on consonant retraction. In *Papers for the 34th International Conference on Salish and Neighboring Languages*, pp. 185-214. 34th ICSL, Kamloops, BC.

Mattina, Anthony 2000. Okanagan sandhi & morphophonemics. In ICSNL XXXV. The Thirty-Fifth International Conference on Salish and Neighbouring Languages. University of British Columbia Working Papers in Linguistics, pp. 147-158.

Mattina, Anthony. In Press. Okanagan and Salishan Languages. In *Encyclopedia of Linguistics*. ed by Philipp Strazny. Fitzroy Dearborn Publishers.

Mattina, Anthony, and Madeline DeSautel, eds. 2002. Dora Noyes DeSautel ɬaʔ kɬcaptíkʷɬ. University of Montana Occasional Papers in Linguistics 15. Missoula, Montana.

Mattina, Anthony and Nancy Mattina 1995. Okanagan ks‑ and kɬ‑. 30th International Conference on Salish and Neighbouring Languages, Victoria, B.C.

Sample Page, Dictionary Entries

The dictionary will be printed on 8.5 x 11" paper, possibly in smaller font size. This mock-up shows:

- entries are listed in two columns, with vertical line between.
- new header starts each page, listing the first entry of the page on the left and the last entry of the page on the right. Lemmas are boldface, glosses in italics.
- root entries are preceded ty the root sign √.
- full word entries show morpheme breaks with the plus sign (+).
- subscript numbers differentiate homophonous entries.
- sources other than the compilers (Mattina and Colville Language Preservation Groups, Nespelem, Omak, Penticton) are indicated in parentheses, e.g. (OkB).
- example sentences are enclosed in angled brackets <>.
- sources of examples are identified following the quoted material, e.g. PS (Peter J. Seymour); CV (Colville Group); EN (En'owkin Group); etc.
- lexical affixes are indicated with the boundary =.
- idioms, complex forms (**cplx**), compounds (**cpd**) are so identified.
- cross references are marked **xr**.

√**x̌ẃ₂** *dry*. •Itr **x̌ẃ+x̌aẃ+t** *dry* (OkB). **x̌ẃ+aẃ** *get dry (inch)*. **x̌ẃ+x̌ẃ+aẃ** *pl get dry* <x̌ẃ+aẃ iʔ sɬiqʷ *The meat has dried.* EN> **i–ks+x̌ẃ+áẃ** *i-fut dry* <x̌lap lut ks–x̌ẃ+aẃ–s iʔ sksk̓ʷalts iʔ l sqiltks *In the morning the sweat on his body won't be dry.* PS> •Tr **x̌ẃ–nt–ím** *dry st* <x̌ẃ–nt–is–lx x̌ẃẃalqʷ, x̌ẃntislx məɬ ʔipslx *They dry the sticks, then they whittle and wipe them.* EN> **x̌w–ɬt–im** *dry st of sb's* <x̌ẃ+x̌ẃ–ɬt–is iʔ tkəɬmílʷ iʔ sɬiqʷs *She dried the woman's meat.* DD> **c–x̌ẃ–st–im** *cust dry.* <c–x̌ẃ–st–is iʔ l sɬ kʷ kʷaʕst *She always dries them in the morning.* MD>

- Cplx **x̌w=cn** *dry mouth*; **x̌ẃ=ilt** *dry berries*, **x̌ẃ=ilt–m** *to dry berries*; **s+x̌ẃ=ilt+[t]n** *berry drying rack*; **x̌ẃ+íɬċaʔ**, **s+x̌ẃ+íɬċaʔ** *dry meat*, **s+x̌ẃ+íɬċaʔ+m** *to dry meat*; **x̌ẃ+ẃ+alqʷ** *dry sticks*; **n+x̌ẃ=ink** *dry inside*; **n+x̌ẃ=úlaʔxʷ**,

n+x̌ẃ+ẃ=úlaʔxʷ *dry pool, land*; **s+n+x̌ẃ+min** *drying shed or rack*; **s+n+x̌ẃ=ilt+[t]n** *drying place*; **s+n+x̌ẃ=íɬċaʔ+tn** *drying meat/fish rack*; **t+x̌ẃ+x̌ẃ+us** *dry eyes*; **t+x̌ẃ=ilt+[t]n+m** *to dry meat.*

- Idiom **x̌ẃ+x̌ẃ–ilx** *be in a bind*; **x̌ẃ+x̌ẃ=us** *want an excuse.*
- xr √**x̌ẃ₁**, **x̌iẃ** *raw.*

x̌ẃ+íɬċaʔ *dry meat.* Itr **s–x̌ẃ+íɬċaʔ** *dry meat.* <x̌sʔiɬn axáʔ iʔ s+x̌ẃ+íɬċaʔ *This dry meat is good.*> **x̌ẃ+íɬċaʔ–m** *to dry meat.* <ṫiʔ xiʔmíx caʔ kʷ swit ksx̌ẃíɬċaʔx uɬ x̌ẃ+íɬċaʔ–m *Anybody who wants to dry meat dries meat.* CV> icptv **ks–x̌ẃ+íɬċaʔ–[aʔ]x** <kn ks–x̌ẃ+íɬċaʔ–[aʔ]x *I'm going to dry meat.* perf **i–sc–x̌ẃ+íɬċaʔ** <waẏ x̌ast a–sc–x̌ẃ+íɬċaʔ *The meat you dried is good.* CV>

x̌ẃ+x̌aẃ+t *dry.* See √**x̌ẃ₂**.

Sample Page, Interlinearized Text

The odd page header contains text title, page number, and English translation of the title. The even page header has the title of the volume.

- the first line of the interlinearization has the utterance number, and the phonetic representation of the line.
- the second line has the morpheme analysis of each word in the utterance. Inflectional affixes are marked with the hyphen, and derivational and word-formation affixes are marked with the plus sign.
- the third line glosses each stem and inflectional morpheme.
- the fourth line is the English translation of the utterance.
- material in curly brackets { } is a false start or an unfinished utterance
- the underscore keeps together glosses and/or multi-word items, e.g. kill_one.

caʔsláqs   Page #   *Mosquito*

---

24. ixíʔ mat ʔitx uɬ cyaʕp iʔ sẏaɬwánk
    ixíʔ mat ʔit+x uɬ c+yaʕ+p iʔ s+ẏaɬw+ánk
    then maybe sleep and arrive_here art Cricket
    He slept and the Crickets came.

25. cyaʕp uɬ ixíʔ {pul} ƛ̓əxʷəntíməlx
    c+yaʕ+p uɬ ixíʔ ƛ̓xʷ –nt –im –lx
    arrive_here and then kill_many -nt -psv -pl
    They got there and they killed them all.

26. púlstməlx uɬ ċsápəlx
    pul –st –m –lx uɬ ċsa+p –lx
    kill_one -st -psv -pl and gone -pl
    They killed them, all gone.

27. kmax_k̓im iʔ tətw̓ít ixíʔ {iʔ l} iʔ l kʷílstən aláʔ mut ixíʔ ilíʔ
    kmax_k̓m iʔ t+tw̓it ixíʔ iʔ l kʷils+tn aláʔ mut ixíʔ ilíʔ
    only art boy that art in sweat_house here sit_sg that there

    cxʷəlxʷált
    c –xʷl+xʷal+t
    hab -alive
    Only the little boy that was sitting in the sweat house is left alive.

28. ixíʔ qiɬt nt̓a·· ċsápəlx ƛ̓əxʷəntíməlx
    ixíʔ qiɬ+t nt̓a ċsa+p –lx ƛ̓xʷ –nt –im –lx
    that awaken intj gone -pl kill_many -nt -psv -pl
    He woke up, gee, they're all gone, killed.

29. ixíʔ sxʷuys iʔ k̓əl t̓ík̓ʷət uɬ ilíʔ iʔ ċíʔstən
    ixíʔ s –xʷuy –s iʔ k̓l t̓ik̓ʷt uɬ ilíʔ iʔ ċiʔs+tn
    that dur -go -3i art to lake and there art weeds_sp
    He went to the lake, and there there were those jointed weeds.

30. ixíʔ kʷis iʔ ċíʔstən
    ixíʔ kʷi –s iʔ ċiʔs+tn
    that take -3erg art weeds_sp
    He took the jointed weeds.

The glossary that accompanies each volume of texts is formatted as follows:
- the page header lists the first and last entry in the page.
- inflectional affixes are marked with the hyphen and glossed with an abbreviation, e.g. *dur* = 'durative'.
- circumfixes are marked with ^, so that, e.g. **s-** *cust*^ glosses the prefixal part of the *customary* circumfix.
- allomorphs are listed in parentheses, e.g. **s-** (**c-**)**.**
- consonantal make up of the root is in parentheses with the symbol √**.**
- the capital letters that follow each entry abbreviate the text title, e.g. GU 'Two Girls and their Uncle,' FC 'Frog and Crane,' etc., and the numbers mark the first occurrence of the item in each text.

---

s– *inten*. In GU 134.
s– (**c–**). *cust*^. In GU 295.
s– *ipftv*^. In FC 4, CMB 86, CE 86, CR 31, GU 62, M 1.
s– *dur*. In SF 6, CN 7, FC 2, CMB 11, CE 34, CR 2, M 22.
s– *?* In CN 78.
–s *2obj*. In CR 40, GU 151.
–s *3erg*. In SF 3, CN 11, FC 10, CMB 8, CE 2, CR 1, GU 9, M 8.
–s *3i*. In SF 6, CN 7, FC 2, CMB 11, CE 34, CR 2, GU 35, M 15.
-s *3in*. In SF 3, CN 2, CMB 1, CE 2, CR 1, GU 1, M 20.
sax̌ʷ+t (√sx̌ʷ). *go downhill*. In CR 19.
sa?+stám (√stm). *brother-in-law, sister-in-law*. In GU 78.
s+c– *ipftvp*^. In FC 40, CMB 83, CR 79, GU 175, M 20.
s+c+caʕsnt (√cʕsnt). *little stars*. In CE 27.
sc+q̇ʷĺ+iẇs (√q̇ʷĺ). *berry picking*. In CR 9.
s+ċ+ċẇ+xan' (√ċẇ). *leg dim*. In CR 54.
s+ċi (√ċm). *bone*. In GU 3.
s+ċẇ+xan (√ċẇ). *animal hind leg*. In SF 82, CR 50.
sic (√sc). *new, then*. In FC 19, CMB 77, CR 88.
sílxʷa? (√slxʷ?). *big*. In SF 126, CR 89.
síṗi? (√sṗ?). *hide*. In CN 53, CMB 47, GU 114.
siwɫ kʷ (√swɫ). *water, liquor*. In SF 56, CN 34, CE 19, GU 4, M 10.
siw+st (√sw). *drink*. In CE 24.
siwst+a?x (√swst). *dinking water*. In CE 9.
síya? (√sy?). *saskatoons, service berries*. In CN 20.
s+kic+c (√kc). *visitor*. In CN 12.
s+k+kʕáka? (√kʕ?). *bird, animal, chicken*. In CE 25, CR 106.
s+k+l kʷt+ilt (√l kʷ). *distant child*. In CMB 3.
s+kɫ+ikṅ (√kɫ). *rump fat*. In GU 142.
s+kmxist (√kmxst). *bear*. In CN 8, GU 116.
s+kṁxist (√kṁxst). *bear*. In GU 101.
s+ḱawíla?x (√ḱwl?x). *Addy*. In SF 96.
s+ḱl+cniɫc (√cnɫc). *to the side*. In SF 21.
–skʷ *tsimptv*. In CN 56, CMB 36, M 13.
s+kʷan+x̌n (√ kʷn). *prisoner*. In GU 227.
s+kʷ+kʷr+ína? (√ kʷr). *clam shell*. In CR 32.
s+ kʷ+ kʷusnt (√ kʷsnt). *stars*. In CE 27.
s+ kʷ+ kʷ?+iɫp (√ kʷ?). *rose bush*. In CE 18.
s+ kʷlis (√ kʷls). *kinnickinnick*. In CR 3.
s+ kʷls+iɫmlx (√ kʷls). *kinnickinnick*. In CE 37.
s+ kʷr+xan (√ kʷr). *Crane*. In FC 6.
s+ kʷs+ kʷist (√ kʷst). *name pl*. In CN 4.
–skʷy *tpimptv*. In M 13.
s+kʷ+kʷiṁ+ɫt (√kʷy). *child*. In GU 98.
s+kʷúma+lt (√kʷm). *virgin*. In FC 4.
skʷut (√kʷt). *half, other side*. In GU 64.
s+kʷuy (√kʷy). *man's mother*. In CE 44, GU 22.
s+liṗ (√lṗ). *fire wood*. In M 42