

# FINDING THE LOCUS OF BEST PRACTICE: TECHNOLOGY TRAINING IN AN ALASKAN LANGUAGE COMMUNITY<sup>1</sup>

**Andrea Berez**

*University of California, Santa Barbara*

**Gary Holton**

*Alaska Native Language Center*

*University of Alaska Fairbanks*

## Dena'ina Language and the DATA Project

At the time of writing, a three-year language technology project focused on the Dena'ina Athabascan language of Southcentral Alaska is coming to its end. The Dena'ina Archiving, Training, and Access (DATA) project was developed to address three primary concerns: (1) requests by the Dena'ina community for greater access to existing language materials; (2) the need to implement enduring standards of digital language archiving; and (3) the need for intensive training of community members and graduate students in language technology and fieldwork techniques. It is the last of these goals, training, that is the focus of this paper. We discuss our experiences bringing language technology into a small speaker community setting, and reflect on why the results of our training efforts differed from what we expected. It seems that we, as linguists, may initially have been blinded by our 'best practice ideology' to the unique and personal interests of Dena'ina heritage speakers regarding the place of technology in language revitalisation efforts. Ultimately we found that a true division of labour between linguists and community members may be the best solution in situations of severe language endangerment.

Dena'ina is spoken by fewer than 75 people in the Cook Inlet region of Alaska (Figure 1). While several nascent maintenance efforts are underway, English remains the dominant language of the Dena'ina community, and children are no longer acquiring Dena'ina as a first language. However, the current endangerment of the language is in sharp contrast with the depth and breadth of existing documentation: while there is no comprehensive dictionary or grammatical description for the language, Dena'ina is among the better documented of the Northern Athabascan languages. The Alaska

Native Language Center (ANLC) Archive contains documentation dating back to a wordlist collected in 1778 by William Anderson during Captain Cook's voyage to Alaska, and Dena'ina has received considerable attention since the early 1970s, particularly through the work of linguists James Kari and Joan Tenenbaum and speakers Peter Kalifornsky and Albert Wassillie (Berez & Holton 2005). Among the materials at the ANLC Archive are extensive fieldnotes, grammatical descriptions, narratives, ethnographic information, pedagogical materials, and hundreds of hours of audio recordings.

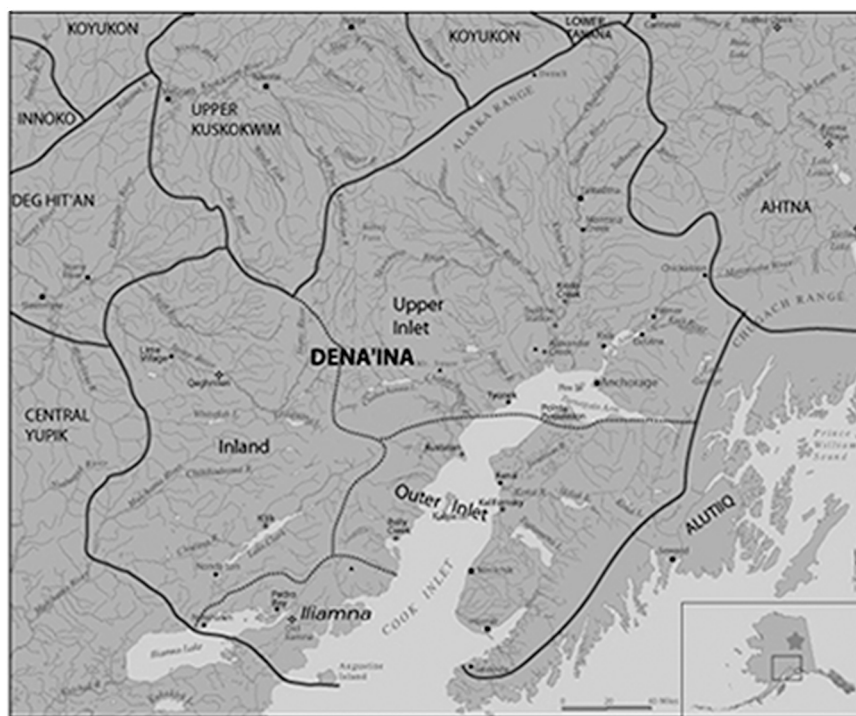


Figure 4: Dena'ina language area (Fairbanks is indicated with a star in the inset). Base map courtesy Lake Clark National Park and Preserve and Barbara Bundy; text by James Kari.

Unfortunately, the ANLC Archive is located in Fairbanks, far from the Dena'ina region (see Figure 1 inset), so access to these items remains difficult for Dena'ina heritage speakers and linguists worldwide. Prior to the beginning of the DATA project, Dena'ina resources in the ANLC Archive had not been extensively catalogued, and lack of attention to preservation rendered paper and audio recordings at risk of long-term degradation.

To address these issues, the DATA project created an Online Archive of the Dena'ina language materials at the ANLC, using the technological expertise of the LINGUIST List and the recommendations for archiving standards of groups like E-MELD and OLAC.<sup>2</sup> At the moment, the Archive contains metadata for more than 500 entries, all of which are searchable by a range of fields, including title, date, dialect, content type, and participants. The Archive is built on an Oracle database with a ColdFusion interface (for a thorough description, see Holton, Berez & Williams 2006). Most of the resources are available for download in digital format. The Online Archive is designed to protect intellectual property rights by permitting restrictions to be placed on access to materials deemed sensitive by the Dena'ina community. The access system is flexible and can easily be adapted to changing community consensus. The Online Archive represents the first *A* of the acronym DATA: that is, archiving.

The second *A* stands for access, which is covered by the Qenaga.org website (*qenaga* means 'language' in Dena'ina). The website acts as an online portal to Dena'ina language and community information. Here, users can learn of local news and goings-on in the Dena'ina community, as well as access resources for learning Dena'ina language. The site contains information on grammar, phonology, verb structure, music, literature and history, as well as instructional modules and suggestions for self-study. The web portal provides a 'value-added' product, synthesising existing archival documentation in order to make it more accessible to users. The web portal also supplements existing documentation by providing access to 'known' but undocumented information, like short biographical descriptions of people who have worked toward language maintenance, including speakers, educators, and linguists. Until now this kind of information had, for the most part, gone unrecorded.

The *T* in DATA represents its significant training component. In the summer of 2005 we developed and taught a course in language technology at the Dena'ina Language Institute, a three-week intensive summer language program in Soldotna, Alaska.<sup>3</sup> The course covered basic HTML programming and audio recording and digitisation. Students were Dena'ina heritage speakers and friends of the community, and their ages and levels of experience varied widely. By the end of the three-week class, each student produced two small websites about Dena'ina language or culture, with incorporated audio clips of Dena'ina language collected from elders during the Institute.

The training goals of DATA reached beyond the limits of the summer class, and we worked to provide more in-depth instruction to interested students. One of the language products resulting from DATA is a CD-ROM containing nineteen Dena'ina narratives with time-aligned audio and text (Kari & Berez, forthcoming; see Figure 2). The audio comes mostly from legacy audio recordings, some nearly forty years old. These recordings had been digitised as part of the archiving component of the DATA project, and those recordings are available in the Online Archive. However, recordings with linked translations had never before been available to the Dena'ina community at large.

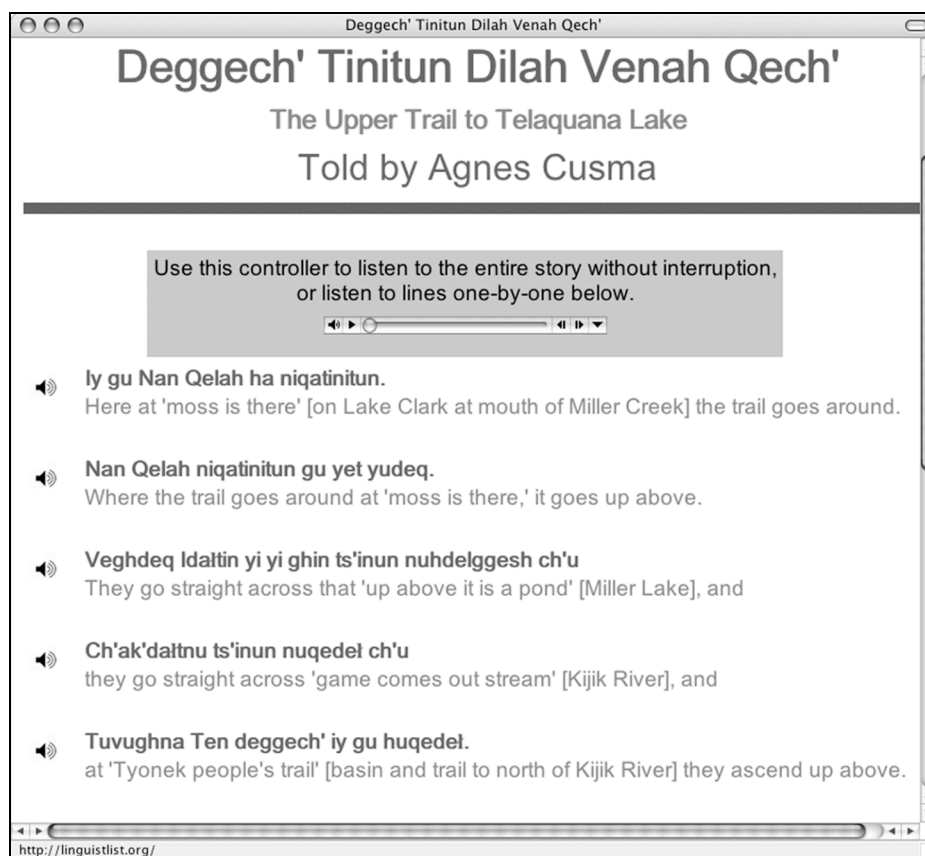


Figure 5: Screenshot from the Dena'ina story CD (Kari & Berez, forthcoming).

The nineteen stories included on the CD-ROM do not begin to scratch the surface of the supply of legacy narrative recordings available in Dena'ina. Given the positive community response to the collection, their nearly constant quest for new language materials, the eagerness of some

students to broaden their technology skills, and the project directors' commitment to good practice technology infrastructure, we decided to train community members to create a similar product. In the remainder of this paper we describe this training process and reflect on our experience with community technology training.

### Why Community Training?

Why bother training community members, when we were perfectly capable of producing the story CD ourselves and then handing the finished product over to the community? In truth, training in the production of CDs was not our initial objective, and in the early months of the project we set out to simply deliver the finished CD to the community. Before we started, we had very little knowledge of the software and programming languages we eventually used, and selecting among the many available options required careful consideration of each in terms of learnability, ease of use, and long-term viability. But as we developed the workflow—a process that took several months, during which we relied heavily on our professional contacts in the field of language technology for advice—we realised that it would be a good idea to put as much attention into the process itself as we were putting into the finished product. We knew that we were going to be available to do the work in the short term, as the grant had a time frame of three years, but it seemed a shame to waste the months of investigation and self-training. By focusing on streamlining the process of making the CD, according to principles we outline here, we were able to develop a community training procedure.

Certainly, developing such a procedure is a larger undertaking than simply making the product and handing it over to the community, so the potential benefits of doing so need to be made clear. We have identified several potential benefits:

- Often, the person who makes language materials gains as much benefit, in terms of language learning, as the consumers who use the materials later. We learned a great deal of Dena'ina language simply from manipulating texts and recordings.
- Teaching and learning materials of all kinds are very desired.
- For severely endangered languages, many aspects of language work are in a race against the clock. Many hands make light work, and an army of technology 'experts' can do more than just a few.

- Community members have access to documenting authentic, spontaneous language-use contexts and events that outsiders do not have (see Villa 2002).

Most importantly, perhaps, community technology training keeps financial and intellectual resources inside the community. Often, technology work is contracted to non-community members at great expense. As Hinton (2001) cautions:

(a)n easy pitfall is for a language program to get a grant to buy computers and hire a consultant to make some CD-ROMs or programs for the language—and then be left without the money or expertise to upgrade or make repairs or transfer the program or CD-ROM to the next stage of technology. (268)

We were nearly witness to such an experience. During the course of the DATA project, a local tribal organisation received a grant that included USD85,000 budgeted as a subcontract for digitising audio recordings. The DATA project was able to provide training in digitisation to community members who were then able to do the work themselves, thereby allowing the budgeted funds to be spent inside the community, while at the same time developing technological expertise within the tribal staff.

To invoke an adage, a good workflow is like teaching a person to fish, to become self-sufficient in the use of technology as a tool for language revitalisation. Villa (2002) also advocates technology training as a response to the ills of cultural mining:

Certain researchers have entered indigenous groups for study, taken the data they sought, and then left to publish their research without returning any information to the people with whom they worked. As (a) result ... many indigenous groups in the US closely control who is allowed to enter the community to carry out work on language.... One solution to this problem is to train members of the language group in the use of advanced technology in order that materials for language preservation and teaching can be collected, archived, and prepared by in-group members for other in-group members. (93-94)

Thus, the benefits of technology training are not limited to improvements in the quality of products produced during or shortly after the training period. Technology training can bring ownership, in a larger sense, to language documentation and maintenance efforts within the community.

## Goals

In designing the training procedure we established a number of goals, both general and project-specific. The general goals, listed in Figure 3, would probably apply to any language technology project, because they are the methodological goals that remain constant from application to application (that is, these goals would still apply if the finished product were a dictionary or a language-related game, rather than an interactive CD of narratives). The project-specific goals, on the other hand, define the function of the final product.

The first goal in Figure 3, 'adhere to best practice procedures for data preservation', reflects our desire for materials to be of lasting value. Even if the materials are not intended to be of documentary value, we cannot ignore best practice recommendations (see Boynton et al, this volume). We see two reasons for this. First, even when documentation is not the intent, language maintenance materials may end up being the best record of a language. An example is the case of Han Athabascan, one of Alaska's most severely endangered languages, for which some of the best documentation is arguably a series of pedagogical recordings that were never intended to be of documentary value. Second, without attention to best practice, the resulting products may not even serve their immediate pedagogical purpose. This can occur when products have such long development times that the underlying technology is nearly obsolete when development is complete. Examples from Alaska include numerous products produced with Hypercard and Macromedia Director.

- adhere to best practice procedures for data preservation
- be easy enough to be learned in a short time by a motivated student
- include a significant training component
- include instructions for getting assistance if a problem occurs
- require only free or inexpensive software
- employ bug-free fully functioning software
- allow easy installation of the final product

Figure 3: General goals for training.

The remaining general goals in Figure 3 are motivated primarily by our experiences with teaching language technology. A good workflow must be intuitive and easy to learn, and users must have access to help resources. Successful trainees need to acquire sufficient expertise to continue development and maintenance of the product. The next two goals in Figure 3, as we shall see below, are apparently in conflict and must be balanced against each other: while it is important that inexpensive tools be made available for community work, such tools frequently contain more bugs and are prone to crashing.

As a final general goal, we wanted the resulting product to be portable and to require very little configuration of home computers to use the CD. Previous experience with HTML-based projects intended for community use has shown the tolerance for installation difficulties, such as configuring browser plug-ins, to be extremely low. Indeed, after a preliminary beta distribution of the Dena'ina CD in 2005, we received numerous requests to hand-configure users' machines to run the required browser/plug-in combination. Similarly, a Dena'ina multimedia phrasebook produced recently at the ANLC (Balluta & Evanoff 2004) has been inaccessible to many potential users because it requires the QuickTime audio plug-in with a specific web browser. Even though detailed and accurate instructions for installing and configuring the plug-in with the requisite browser are included on the CD, many users have proven reluctant or unable to complete the installation and configuration progress. Additionally, in trial versions of a Tanacross multimedia CD that includes QuickTime video clips (Holton & Thoman 2006), the CD automatically attempts to install and configure the QuickTime plug-in if it is not already present. Many users, however, are reluctant to click 'OK' in the confirmation dialog box.

We also developed a number of project-specific goals, which may or may not apply more generally to other language projects. In order to implement our best practice goal, we chose to employ an XML structure and to include metadata for each narrative. However, in order to meet the goal of easy installation, we chose to include support for non-Unicode enabled machines—a choice apparently in conflict with best practice recommendations supporting Unicode. This decision reflects our experience with browsers and machines that still fail to display some Unicode characters correctly.



## Developing the Workflow

In this section, we describe the process of developing the workflow. This was largely a process of self-training, in which we investigated possible options for converting legacy materials into the finished presentation product. While we have worked hard to ensure the technologies we used will be relevant in the years following publication of this paper, it is possible they will not be (indeed, even at the time of writing some of the procedures described here are outdated). We urge readers to look beyond the details of the workflow, and to consider the larger issues that remain relevant.

To select the stories for inclusion in the collection, we searched the existing archived materials to find matching pairs of transcriptions and audio recordings. Some transcriptions had been previously published without accompanying audio, while others were catalogued in the ANLC Archive as manuscripts. Unfortunately, existing metadata were generally not rich enough to aid significantly in the matching of text and audio, because recordings and transcriptions had been archived at the ANLC as separate parts of larger resources. Transcribed audio recordings for the most part did not exist as individual resources, but rather as part of larger field recordings containing material from entire field sessions. Other audio-transcription pairs were in the personal collection of linguist James Kari. In only one case did we select matching text and audio that had been previously published together (Johnson 2005).

Kari provided transcriptions as either handwritten documents or as WordPerfect files. We typed the previously published stories and the handwritten transcriptions into a text editor and passed them back to Kari for proofreading. As for the WordPerfect files, we simply opened them in the text editor and checked for character conversion problems. Dena'ina has only two non-Latin characters in its orthography, *Ł* and *ł*, which in pre-Unicode days were typed with the backslash symbol (\). This was fortunate, because we discovered no real character conversion problems that can plague older files in other languages. We quickly cleaned up small formatting issues, and the new text file became our master copy of the transcription.

Most of the audio recordings, with the exception of a few recent digital recordings, had been previously digitised from original cassette or reel-to-reel. The process of selecting audio recordings thus involved editing large audio files. Generally, we tried to select audio that was free of background

noise, but there is a significant hiss in at least one very old recording. We did not attempt to remove this noise.

After selection of the corpus, we were ready to begin creating a time-aligned, interlinearised version of the story. We already knew that the ELAN software could be used to create time-aligned XML files in accordance with best practice archiving recommendations.<sup>4</sup> After some experimentation, we found the easiest approach to alignment to be from the 'back-end': we first created the time divisions in the ELAN GUI, then we opened the ELAN file in a text editor and cut-and-pasted the Dena'ina and English transcriptions directly into the ELAN XML. Our alignments contained only two tiers, one for Dena'ina and one for an English free translation, but this back-end method would work for any level of interlinearisation, provided the user takes care in navigating through the XML structure. This method was generally easier than pasting text directly into the ELAN interface.

With our annotation files in hand, it was time to turn our attention to presentation. We created HTML pages in order to be consistent with our goal of making the product readily available on home computers. All modern computers have a web browser installed, and most Dena'ina community members, even those living in the more remote villages with little or no internet access, know how to use them. We used XSLT to transform the ELAN XML into HTML. As neither of us was familiar with XSLT, we relied on books and our colleagues at the LINGUIST List, and managed to teach ourselves enough XSLT to create a passable, though not elegant, stylesheet.<sup>5</sup> The stylesheet iterates through the ELAN tiers and transforms the result into 'triplets' of Dena'ina and English lines of text, and the accompanying Flash file. Regrettably, this stylesheet does not manipulate the ELAN timecodes. We then used the LINGUIST List's ColdFusion server, where the Qenaga.org Online Archive is housed, to perform the transformation.

The stylesheet, as written, does not produce immediately useable HTML, and the resulting output must be edited further before it can be used. In addition, the stylesheet handles only the visual portion; it does not deal with presentation of audio.

We investigated several different audio presentation technologies, including SMIL, QuickTime, and Flash. Our stated goals for the project included allowing a choice between uninterrupted audio and line-by-line clips. Offering clips proved to be the most challenging aspect of the entire

project, one that took us several months, and two versions of the CD-ROM, to solve. Early in the project, we had no easy way of breaking the full audio file into smaller segments, and cutting audio by hand would have proved too costly in terms of time and the potential for mistakes. Our first solution came in the form of QuickTime embedded media players. QuickTime recognises <STARTTIME> and <ENDTIME> tags, and so can be programmed to play only a section of the entire file. This method has two advantages: there is no need to cut the audio into smaller files, and memory space on the CD is saved by including only one audio file for each story.

However, our QuickTime solution also had a number of significant disadvantages. First, we had to copy the timecodes from the ELAN file header into the HTML by hand for each line of text. This function could probably have been incorporated into the stylesheet, but that was beyond our limited expertise. The second disadvantage is that while Apple uses QuickTime as the default media player plug-in for its machines, PCs do not. Because most of the PC default players do not recognise the <STARTTIME> and <ENDTIME> tags, PC users would be forced to download QuickTime and then configure their computers to use it. As described above, community testing showed this to simply be too troublesome for most users, even when we provided hands-on instruction. The final disadvantage of QuickTime is that numerous players in one page will often crash the computer, or may simply fail to load. In short, QuickTime, as used here, was neither very programmer-friendly nor very user-friendly.

A better solution came in the fall of 2005, when Aric Bills, a graduate student at the University of Alaska Fairbanks, made a small piece of software with big potential for ELAN users. Using the scripting language Tcl,<sup>6</sup> he created an 'audio chopper' that reads the timecodes from the ELAN file header and creates audio clips with unique file names with the push of a button. Now we were free to investigate Flash as a method for playing audio. We created Flash buttons for each line of audio and embedded those into the HTML.

This solved several problems at once. First, it was fairly easy to modify the XSLT to include the unique audio filenames, which meant much less massaging of the HTML output. Second, Flash has none of the user-friendliness problems that QuickTime has. Recent figures from Macromedia (2006) claim 97% penetration of Flash Player as of December

2005. Pages with multiple Flash buttons are not nearly as buggy or susceptible to browser crashes as those containing multiple QuickTime players.

One of our general goals was to include support for non-Unicode-enabled machines, which we met by creating a parallel set of pages wherein all the *Ě* and *Ľ* characters are replaced with a backslash, as was done in the early days of Dena'ina transcription. The navigation within the parallel pages is designed to be as unobtrusive as possible. Once the user has navigated to a page containing backslashes, all the linked pages also contain backslashes, eliminating the need to switch back and forth for each story. Of course, the user can choose to return to the pages containing the Unicode characters at any time. Finally, we added metadata for each story to a credits page.

In the end, we were able to meet most of our general goals for the workflow and many of our project-specific goals. Our level of success in reaching our stated technology goals for the workflow is summarised in Figure 4.

### Limited Success of Community Training

As Figure 4 reveals, training was a significant area of under-achievement. Undeniably, community training has been one of the most challenging aspects of DATA.<sup>7</sup> The three-week technology class at the 2005 Dena'ina Language Institute successfully met the goals stated in the syllabus, and each student who participated learned something new. However, as technology training was only one of three components of the DATA project, there was simply not enough scheduled contact time to address our advanced training concerns. Most of our time during the three years of the project was spent developing the Online Archive and the Qenaga.org website from our offices in Fairbanks and Michigan.

Providing advanced training was further complicated by difficulty in locating and recruiting interested community members. Only a few students in the 2005 summer class undertook projects with the intention of building on them after the end of the instruction period. One student created an interactive map of the Lake Clark area, with clickable hotspots linked to Dena'ina audio of the placenames. The student has continued to develop this project in her spare time during the last year. Her experience with this project has helped her to contribute in a significant way to

development of the United States National Park Service's GIS placenames database for Lake Clark National Park.

<i>Goal</i>	<i>Achieved?</i>	<i>Comments</i>
Easy to learn	maybe	With sufficient student motivation. Some parts easier/more appealing than others.
Training	maybe	Attempted but perhaps not successful.
Bug-free	yes	As of now, yes. Future bugs difficult to predict.
Free or inexpensive software	no>yes	Flash is expensive, but may be available through a university license; QuickTime is free.
Able to contact linguist later	yes	Linguists are in close contact with the community and readily available by email.
Audio available whole or clips	yes	
Built upon Elan XML	yes	
Metadata included	yes	However, metadata is not integrated into archival version.
Non-Unicode support	yes	
Little configuration of computer to use	yes	Except computers without Flash Player (a small number).

Figure 4: Achievement of workflow goals.

The two most successful students from training class, two non-Indigenous men with close ties to the Dena'ina community, have secured more than USD20,000 from the University of Alaska and Kenaitze Indian Tribe to upgrade their equipment and produce digital products on the Kenai dialect of Dena'ina. One of these products is an interactive CD-ROM suitable for internet publication (Boraas & Christian 2006). Linked pages cover Kenai Dena'ina phonology, vocabulary, grammar, verb structure, narratives, and ethnogeography of Dena'ina territory. Our continued assistance to these students has taken the form of emails and a few extra hours in front of a computer, but no additional formal training has taken place. These

students also have plans to produce a collection of narratives similar to the one produced during the DATA project, but to our knowledge they have not yet started work on it. Whether they will adopt the workflow discussed in this paper, in whole or in part, is questionable. It will be interesting to see which aspects of the workflow are attractive to them.

### Balancing Goals: the BP Cline and the Locus of Best Practice

For the Dena'ina stories project, we were largely able to meet the technology workflow goals we established. However, in the end we realised that some of these goals were actually in conflict, causing a 'cline of best practice' to emerge within our workflow. We now believe that such a 'BP cline' may be inherent in community-oriented technology workflows more generally. Best practice goals are easily violated in favour of learnability and user-friendliness goals, and vice versa. Figure 5 illustrates this concept. For example, we segmented the audio and used Flash because we wanted an end product that could be easily installed and was bug-free. However, the pathway from the archival time-aligned XML version to the Flash-enabled end product is rather non-intuitive, violating our workflow goal of easy learnability. The high cost of Flash production software violates our goal to use only inexpensive software.

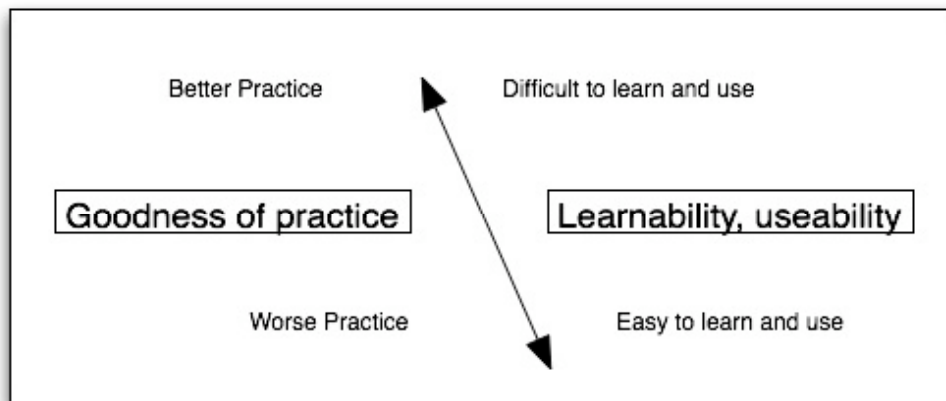


Figure 5: The cline of best practice.

Another example of the BP cline can be found in the transformation of XML into HTML. We used the LINGUIST List ColdFusion server to perform the transformation, which had the advantage of forcing the storage of the alignment file in a secure location. It is also possible to perform the transformation in XML editing software on a local machine.

While doing the latter is certainly quicker and easier than uploading the file to the server, the archival-quality file is not automatically stored in a secure location.

While all community trainees believe intrinsically in the value of best practice as it applies to language technology, in reality best practice goals are often traded for other more immediate or more practical goals. The hardest thing for our trainees to accept is the value of ensuring the longevity of their work; they prefer instead to focus on only those stages of the workflow that directly result in an attractive final product. Our students are very interested in learning HTML, but they would prefer to skip the alignment altogether in favour of transcribing directly in the source code. While many of our students have now become proficient in editing HTML pages, none has yet mastered the art of massaging the XML representation into HTML. Thus it seems that at several points in the workflow, two paths are available: an easy way, and a best practice way. Given the choice, people will very often choose the easy way.

This failure—if it is one—is at least in part owing to a shortcoming of our workflow design. We failed to develop (or locate) an appropriate tool for managing the conversion to presentation format, and we failed to develop (or locate) an audio playback tool that did not rely on expensive proprietary software. While better solutions could surely have been found for this project, similar problems are bound to arise in future projects. In any case, the stated goal of the DATA project was not to develop new tools but instead to investigate a subset of the existing tools to see how they can be applied to Dena'ina. As a case study, our experiences with the Dena'ina narrative CD indicates that the field needs better tools for making presentation products.

In fact, in order to better understand the shortcomings of our training program, we need to revisit not only our stated technology workflow goals, but also our unstated ideologies regarding those goals. From the outset, we assumed best practice to be an integral part of the community technology workflow, and this assumption became our first workflow goal. As linguists who have been involved with digital language documentation and archiving, the need for best practice seemed not only obvious, but universal. 'Best practice' had become our mantra and our creed. In incorporating best practice recommendations into the community workflow we felt we were participating in a form of technology repatriation, responding to Villa's (2002) admonition to

provide training in advanced technology. And yet in doing so we may have inadvertently engaged in a form of cultural hegemony, imposing the culture of best practice upon a community that may not want it. Indeed, the Dena'ina community may prefer to develop its own Indigenous sense of best practice.

Whether or not this is the case, it is worth re-examining the locus of best practice in community technology training. This is not an issue of whether the community cares about best practice; of course they care. Linguists and community members alike want language materials to be preserved and to be of lasting value. But to care about best practice is not necessarily to assume responsibility for its implementation. A more effective approach to community-oriented language technology might strive to assign tasks in a manner that best takes advantage of relevant skills. In small communities with limited resources it may be most reasonable to place concern for best practices in data preservation in the charge of linguists, while community members focus their efforts on designing and implementing presentation and pedagogical formats.

## Conclusion

In the best collaborative efforts of speaker communities and linguists, each party works to help the other meet its goals. Whether it is necessary—or even possible—for each party to adopt the goals of the other as its own is unclear. As field linguists, we cannot expect communities to commit to implementing best practice recommendations for long-term data preservation as eagerly as we do (indeed, it is hard to convince all linguists to commit to such a task).

It is a good idea for linguists to help communities become self-sufficient in language technology when requested. But speaker communities may not approach language technology with the same goals as us, the linguists; indeed, they may not use their new knowledge in ways we expect. If a community requests training, we need to be sure to serve their needs when doing so. Asking them to serve our best practice needs can lead to frustration for everyone.

Any best practice component of a community-tech workflow should be as unobtrusive as possible. There is a tenuous line between ensuring long-term data preservation and creating a workflow that is too cumbersome to be attractive. The last thing we want to do is create a roadblock between the community and their language goals (Dauenhauer & Dauenhauer



1998). In the case of the DATA project, the responsibility for best practice fell primarily on the linguists rather than on the community. This may or may not prove true for other projects. As linguists, we cannot be disappointed or frustrated if community members are not interested in best practice. They simply may not be.

Then again, they may be. In the end the goals of linguists and the community are not really all that different. Both are interested in creating materials to document the language and aid in language revitalisation. And both linguists and community members are interested in ensuring the long-term preservation and accessibility of those materials. To the extent that technology training undertaken with good digital practice procedures in mind will help ensure the long-term viability of language materials, both linguists and community members will surely benefit from such training. Thus, while the results of the community training program developed for the DATA project are not yet encouraging, we remain optimistic that future initiatives to develop community language technology courses will be successful. And we hope that others will benefit from knowledge of our experiences.

## Endnotes

<sup>1</sup> We gratefully acknowledge the assistance of our colleagues Sadie Williams, for her technology expertise and her help developing the workflow described herein, and James Kari for selecting and editing the corpus. A version of this paper was presented at the 2006 E-MELD Workshop on Tools and Standards: The State of the Art, held in Lansing, Michigan, USA in June 2006. Support was provided by a US National Science Foundation grant (NSF-OPP 0326805, PIs Helen Aristar-Dry, Verónica Grondona, and Gary Holton). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

<sup>2</sup> Dena'ina Online Archive: <http://qenaga.org/archive>; E-MELD (Electronic Metastructures for Endangered Languages Data): <http://emeld.org>; OLAC (Open Language Archives Community): <http://www.language-archives.org>.

<sup>3</sup> ANL 295: Technology for Alaska Native Languages was offered as a University of Alaska Fairbanks course at Kenai Peninsula College by instructors Andrea Berez and Sadie Williams.

<sup>4</sup> ELAN Multimedia Annotator: <http://www.mpi.nl/tools>.

<sup>5</sup> The stylesheet is available from the authors upon request.

<sup>6</sup> Tool Command Language: <http://www.tcl.tk>.

<sup>7</sup> In contrast, graduate assistant training was highly successful, providing two years of training to two students, and limited training to two other students. Despite extensive recruitment efforts, none of these students were from the Dena'ina community.

## References

- Berez, A., & Holton, G. (2005). *Dena'ina qenaga: Previous linguistic work*. Retrieved February 25, 2006 from <http://qenaga.org/lingwork.html>
- Balluta, A., & Evanoff, G. (2004). *Dena'ina Qenaga Du'idnaghelnik/Dena'ina words sound pretty*. Fairbanks: Alaska Native Language Center.
- Boraas, A., & Christian, M. (2006). *Kabtnuht'ana Qenaga/The Kenai people's language*. Kenai, Alaska: Kenai Peninsula College and Kenaitze Indian Tribe, I.R.A.
- Dauenhauer, N.M., & Dauenhauer, R. (1998). Technical, emotional, and ideological issues in reversing language shift: Examples from southeast Alaska. In L.A. Grenoble, & L. J. Whaley (Eds.), *Endangered languages: Language loss and community response* (pp.57-98). Cambridge: Cambridge University Press.
- E-MELD. (2005). *E-MELD School of Best Practice: Dena'ina*. Retrieved April 18, 2006 from <http://emeld.org/school/case/denaina/index.html>
- Hinton, L. (2001). Audio-video documentation. In L. Hinton, & K. Hale (Eds.), *The green book of language revitalization in practice* (pp. 265-271). San Diego, London: Academic Press.
- Holton, G., & Thoman, R. (2006). *The sounds of Tanacross*. Fairbanks: Alaska Native Language Center.
- Holton, G., Berez, A., & Williams, S. (2006). Building the Dena'ina language archive. In Dyson, L., Hendricks, M., & Grant, S. (Eds.), *Information technology and indigenous people*. Hershey, Pennsylvania: Idea Group.
- Johnson, W. (2005). *Sukdu Neł Nuhtghelnek/I'll tell you a story: Stories I recall from growing up on Iliamna Lake*. Fairbanks: Alaska Native Language Center.
- Kari, J. & Berez, A. (Eds.). (forthcoming). *Dach' Dena'inaq' Qeyegh Nuqelnixch'/They tell about this in Dena'ina*. Fairbanks: Alaska Native Language Center.
- Macromedia. (2006). *Flash Player penetration survey*. Retrieved April 29, 2006 from <http://www.macromedia.com/software/flash/survey>
- Villa, D.J. (2002). Integrating technology into minority language preservation and teaching efforts: An inside job. *Language learning and technology*, 6, 92-101.