

## Language archives: They're not just for linguists any more

Gary Holton

*Alaska Native Language Center, Fairbanks*

While many language archives were originally conceived for the purpose of preserving linguistic data, these data have the potential to inform knowledge beyond the narrow field of linguistics. Today language archives are being used by people without formal linguistic training for purposes not necessarily envisioned by the original creators of the language documentation. The DoBeS Archive is particularly well-placed to become an important resource for cultural documentation, since many of the DoBeS projects have been interdisciplinary in nature, documenting language within its broader social and cultural context. In this paper I present a perspective from a legacy archive created well before the modern era of digital language documentation exemplified by the DoBeS program. In particular, I describe two types of non-linguistic uses which are becoming increasingly important at the Alaska Native Language Archive.

**1. LESSONS FROM THE ANALOG ERA.** Within the language archiving community we now face the problem of how to preserve and access a continually growing body of born-digital documentation of endangered languages. But at the Alaska Native Language Archive (ANLA, <http://www.uaf.edu/anla>) we have always been playing catch-up in the digital realm. For the past ten years I have been involved with ANLA in various capacities, and throughout that time the Archive has been not a dry, quiet repository but rather a dynamic part of ongoing language work in Alaska. Indeed, the Archive remains an integral part of language research and language revitalization efforts in Alaska. In many ways the Archive is less a repository and more an active research tool.

The ANLA collection includes nearly everything written in or about each of Alaska's twenty indigenous languages. This amounts to some 15,000 items, including everything from primary field notes to published grammars and dictionaries. In addition, ANLA holds significant collections in related languages outside Alaska; in particular, the coverage of Eskimo languages spoken in the Russian Far East is the most extensive in the world. Not all materials are original or unique; in many cases the Archive holds copies of documents housed elsewhere. (These copies may turn out to be the only extant copies, as happened recently with Bittle's Kiowa Apache field notes.) The aim is for comprehensive coverage. In addition to the print materials, the Archive also contains more than 5,000 audio recordings, though in contrast to the print materials no attempt has been made at comprehensive coverage.



As a legacy archive still struggling to enter the 21st century, ANLA cannot tell us much about best practices in digital preservation or about new technologies for accessing and sharing language resources. What ANLA and other legacy archives can tell us is how language archives have been and are being used. More than half a century has passed since collection began at ANLA, and in that time two important points have emerged regarding archive users and the uses they make of the archive. First, the primary users of the Archive are Native speakers and their descendants, that is, members of Alaska Native language communities. This situation is typical of established legacy archives (Austin 2011). Second, many users seek information which is not primarily linguistic in nature. Neither of these points should be taken as absolutes. The Archive continues to be used by academic researchers, and users continue to seek linguistic documentation. What is notable is that academics are not the *only* users, and linguistic information is not the *only* type of documentation sought. These observations may be relevant as we consider how to make use of language documentation being generated by current projects.

**2. LANGUAGE ARCHIVES AS SOURCES OF CULTURAL DOCUMENTATION.** As linguists – creators or collectors of language documentation – we tend to think first of the *linguistic* uses of archival materials. How can these materials inform our understanding of language? What does a text tell us about how serial verbs are employed? What does a recording tell us about prenasalization? Linguists tend to ask linguistic questions.

But language documentation encodes much more than just linguistic information. Language documenters are first and foremost field workers, interacting with speaker consultants whose interests lie in the documentation of many types of knowledge, be they marriage traditions or navigational techniques. A field worker documenting names for kin terms or stars is very likely to also document knowledge of marriage customs or stellar navigation, respectively, even if inadvertently. Furthermore, much language documentation has been collected without regard to the nature of non-linguistic content. A text may be recorded because it represents a particular genre or style, such as narrative or conversation. The content of that recording – i.e., what that narrative or conversation is about – is generally not constrained by the documenter. As a result language archives now present a veritable treasure trove of non-linguistic information encoded in the signal of the subject language. And users of language archives are very often interested in this type of information.

This interest can be gauged by considering some examples of recent inquiries at ANLA. These include requests for: information on ethnobotany in the Yukon Flats region; a copy of a eulogy for a 19<sup>th</sup> century missionary; a copy of Yup'ik music recordings, for use in writing a libretto; information on genealogy in the Upper Koyukuk region; and information on Russian influence in the middle Kuskokwim River region. All of these requests have a linguistic component; for example, ethnobotanical documentation includes indigenous names for plants, and the eulogy for Father Rysev was delivered in Aleut, translated into English. What distinguishes these patron requests is that they reflect an interest specifically in the *non-linguistic* content. A user sought information on ethnobotany in the Yukon Flats region not because she was interested in indigenous plant names but rather because she was interested in how plants were used. A user sought out the eulogy for Father Rysev not to study the use of Aleut language but rather because he was annotating Rysev's diaries. In this type of research it is the content rather than the linguistic code which is relevant.

Given the diversity of non-linguistic research at ANLA and constraints on space in this article I will elaborate here on just one area: ethnoastronomy. In late 2011 I began working with a user who was attempting to identify names for stars, constellations, and other atmospheric phenomena across Alaska. Before we met, this patron, Chris Cannon, had spent more than two years researching star names without much success; only two relevant published sources of information had been identified. MacDonald's (1998)s study of Inuit ethnoastronomy is one of the best works in this genre for any language; however, it is focused on the Eastern Canadian Arctic and includes only a few comparative terms in the Alaskan dialect of the Inuit language. The remaining 19 Alaska Native languages are completely outside the scope of this work. Bradley (2002) is a study of celestial navigation in Yup'ik and is easily the best modern documentation available. However, it is far from a comprehensive study and is focused on only one Alaska Native language.

It turns out that the best record of Alaska Native ethnoastronomy is to be found within the collections of ANLA. Cannon ((Unpublished ms.)) has worked tirelessly to compile this information using ANLA sources. Some language documentation contains extremely detailed information about stars. For example, Knut Bergsland's (1952) Aleut field notes include a star chart and a hand-drawn, labeled map of Aleut star names. While most linguists have not included such detail, it is still possible to compile star names in many languages. A name is recorded for the constellation *Ursa major* or some portion thereof in each of the twenty Alaska Native languages. In many cases these names are buried in obscure sources. They are not always accompanied by literal translations, but comparing the literal translations reveals some interesting patterns, as shown in Table 1.

LANGUAGE	FAMILY	LITERAL TRANSLATION
Aleut	Eskimo-Aleut	'caribou'
Yup'ik	Eskimo-Aleut	'caribou'
Siberian Yupik	Eskimo-Aleut	'caribou'
Inupiaq	Eskimo-Aleut	'caribou'
Tsimshian	Tsimshianic	'spoon'
Haida	Haida	'sea-otter-stretching-board'
Tlingit	Na-Dene	'all stone'
Ahtna	Na-Dene	'the one that moves above us'
Den'ina	Na-Dene	'one that turns over us', 'stars stretched'
Koyukon	Na-Dene	'it rotates its body', 'according to it the year is measured'
Tanacross	Na-Dene	'dipper'
Upper Tanana	Na-Dene	'I'm sitting'
Gwich'in	Na-Dene	'the seat'

TABLE 1: Literal names for constellation *Ursa major* in Alaska Native languages

Even without examining the actual linguistic form of the names, it is immediately clear from the literal translations that the Eskimo-Aleut languages share a common conceptualization of *Ursa major* as a caribou – something not found in languages of the other families. Examination of other star names reveals further insights. For example, the Inupiaq

constellation *iglupqaqtalik*, literally, ‘the turf house’, is formed from a combination of stars in constellations known to English speakers as Orion, Auriga, and Gemini. These facts are not purely linguistic observations; rather, they are facts about how indigenous Alaskan communities conceive of the sky. They reflect indigenous knowledge embedded in the linguistic code.

**3. CREATION OF DERIVED LANGUAGE MATERIALS.** All Alaska Native languages are extremely endangered, and for most, the youngest speakers are already age 70 or older. As the number of speakers continues to decline, there has been a marked increase in language revitalization efforts (cf. Gaul & Holton 2005). And these efforts increasingly turn to ANLA as a resource for developing derived or secondary language materials. For no language is this more evident than it is for Eyak. In 2008 Eyak became the first Alaska Native language to disappear in recent times. (The last speaker of Tsetsaut, an Athabaskan language formerly spoken on the Portland Canal, passed away in the first half of the twentieth century.) While all of Alaska’s 19 remaining languages are severely endangered, the Eyak situation is in many ways exceptional. In most situations of language shift, it is difficult to identify a “last speaker”. Rather, as knowledge of language erodes, the criteria which define a fluent speaker adapt, creating new last speakers (cf. Evans 2001). But in the case of Eyak the break-up and scattering of the community of speakers led to a large generational gap in transmission, with the last few speakers actually outliving their immediate descendants. In a very real sense, Marie Smith Jones was indeed the last Native speaker of Eyak; no partial or semi-speakers survive. Thus, for Eyak the only surviving sources of information about the language are found in the archival documentation at ANLA.

Since the passing of the last speaker there has been a surge of community interest in Eyak language. Truly an Eyak renaissance. The Eyak Language Project has created a multimedia website ([sites.google.com/site/eyaklanguageproject](http://sites.google.com/site/eyaklanguageproject)) using information harvested from ANLA. Language workshops have been organized in Anchorage and Cordova. While some might characterize these efforts as too little, too late, that characterization is not appropriate, owing to the large body of documentation materials. Thanks to this documentary record, Eyak language revitalization is not “too late”. The first known documentation of Eyak is found in a 1308-item vocabulary by Rezanov (1805). Eyak was then “re-discovered” in the 1930s through the work of de Laguna (Krauss 2006). Then, in 1961, serious linguistic documentation of the language began under the direction of Michael Krauss. At that time there were only four remaining speakers with a good command of Eyak; in other words, Eyak was already severely endangered nearly half a century ago. Krauss began documentation in earnest, eventually compiling about 3,000 pages of field notes; a massive dictionary of some 3,600 pages and 7,000 lexemes (Krauss 1970); and about ten hours of transcribed narrative texts (Krauss et al. 1970). Krauss continued his documentation work sporadically through the 1970’s and 1980’s, but by 1992 only one speaker remained and the corpus became essentially closed. Nonetheless, in spite of the small number of speakers and the lack of a thriving language community, in the span of less than a decade Eyak went from being almost unknown to being among the best documented of the Alaska Native languages. This invaluable documentary record forms the basis for new language initiatives today.

For Alaska Native languages other than Eyak, the situation is not so dire. Fluent speakers remain and can continue to contribute to ongoing language documentation and revitalization

programs. Nevertheless, archival documentation continues to be of great value because it documents an earlier stage of the language – a stage no longer accessible or known to speakers today. This documentation can form the basis for new language materials. For example, the Jesuit missionary Jules Jetté made elaborate and detailed records of place names in the middle Yukon River region in the early 20<sup>th</sup> century. Information from these maps can serve as the basis for new projects such as MapTeach ([www.mapteach.org](http://www.mapteach.org)), which encourages geographic understanding and sense of place among descendants of Koyukon speakers, as well as new types of cultural reference materials (cf. YRDLA 2008). The Alutiiq Museum has converted Jeff Leer's grammatical notes into a multimedia website (<http://www.alutiiqmuseum.org/language/learn-alutiiq.html>).

**4. CONCLUSION.** The massive amount of documentation compiled over the past decade by various DoBeS projects and others represents an unprecedented contribution to the documentary linguistic record. As linguists begin to analyze this body of data, they will no doubt uncover new insights into the nature of human language. But the experience of legacy archives such as ANLA shows us that the DoBeS archive has a great potential to impact other areas of knowledge as well. Legacy archives can thus inform the way we think about the utilization of emerging language archives. In particular, archive users may be interested not only in linguistic applications but also in non-linguistic information encoded in the linguistic record; this highlights the importance of including information of relevance to non-linguists in the metadata. Furthermore, users may wish to repurpose linguistic data to create derived language materials for use in language revitalization efforts. Archives will better serve users to the extent that they facilitate these two types of uses.

#### REFERENCES

- Austin, Peter. 2011. Who uses Digital Language Archives? Post to Endangered Languages and Cultures Blog. <http://www.paradisec.org.au/blog/2011/04/who-uses-digital-language-archives/> (21 March, 2012).
- Bergsland, Knut. 1952. Notes on Aleut wordlists regarding the natural world. Alaska Native Language Archive, Ms., ANLA Item AL950B(B170)1952.
- Bradley, Claudette. 2002. Traveling with Fred George: The changing ways of Yup'ik star navigation in Akiachak, Western Alaska. In Igor Krupnik & Dyanna Jolly (eds.), *The Earth is Faster Now: Indigenous observations of arctic environmental change*, 240–265. Fairbanks: Arctic Research Consortium and Smithsonian Arctic Studies Center.
- Cannon, Chris. (Unpublished ms.). Working list of indigenous star and constellation names of Alaska and their European counterparts.
- Evans, Nicholas. 2001. The last speaker is dead – Long live the last speaker! In Paul Newman & Martha Ratliff (eds.), *Linguistic Fieldwork*, 250–281. Cambridge: Cambridge University Press.
- Gaul, Karen K. & Gary Holton. 2005. Speaking across generations: Dena'ina language revitalization in Southcentral Alaska. *Alaska Park Science* 4(1). 26–31.
- Krauss, Michael E. 1970. Eyak Dictionary. Massachusetts Institute of Technology, Ms., ANLA Item EY961K1970b.

- Krauss, Michael E. 2006. A history of Eyak language documentation and study: Fredericæde Laguna in memoriam. *Arctic Anthropology* 172–217.
- Krauss, Michael E., Anna Nelson Harry, Lena Saska Nacktan, Marie Smith, George Johnson & Galushia Nelson. 1970. Eyak Texts. Massachusetts Institute of Technology, Ms., ANLA Item EY961K1970a.
- MacDonald, John. 1998. *The Arctic sky: Inuit astronomy, star lore, and legend*. Ottawa: Royal Ontario Museum and Nunavut Research Institute.
- Rezanov, Nikolai Petrovich. 1805. Slovar' Unalashkinskago, Kad'yakskago, Kinayskago, Kolyuzhskago; Ugalykmuskago i Chugatskago Yazykov. Fond Adelunga, Academy of Sciences Leningrad, Leningrad. Unpublished ms.
- Yukon River Drainage Fisheries Association (YRDFA). 2008. *Middle Koyukuk River of Alaska: At Atlas of Fishing Places and Traditional Place Names*. Anchorage, Alaska: Yukon River Drainage Fisheries Association.

Gary Holton  
gary.holton@alaska.edu