Community Lexical Access for an Endangered Polysynthetic Language:
An Electronic Dictionary for St. Lawrence Island Yupik
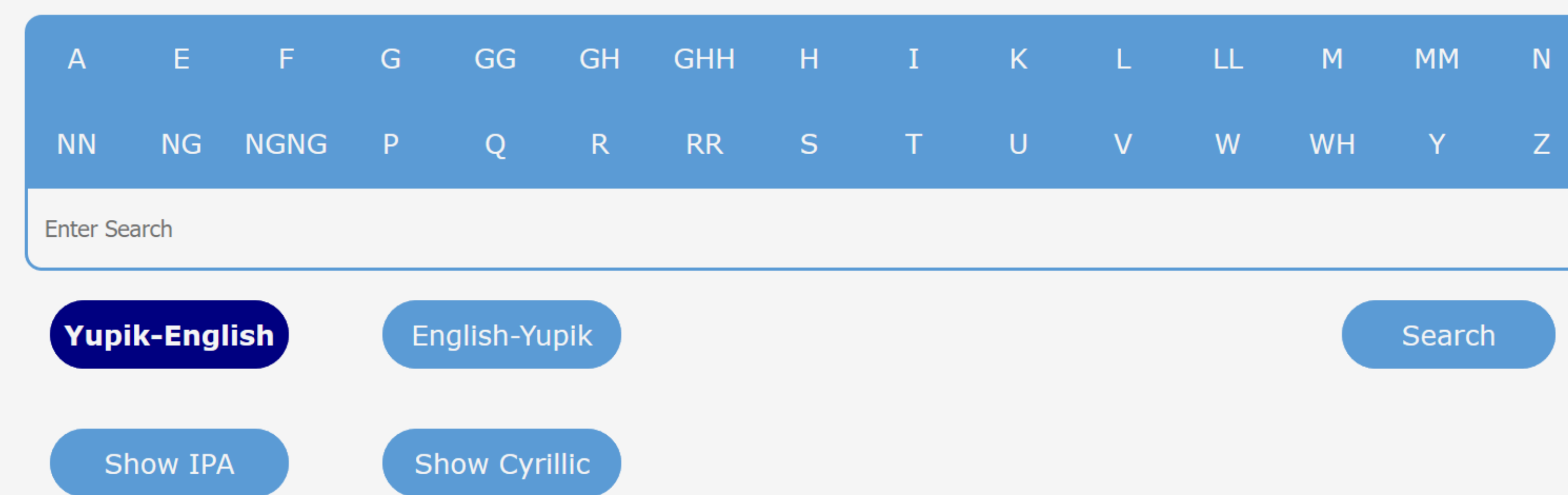
Benjamin Hunt [1]   Emily Chen [2]   Sylvia L. R. Schreiner [1]   Lane Schwartz [2]

[1]George Mason University   [2]University of Illinois Urbana-Champaign

Overview

St. Lawrence Island Yupik (hereafter *Yupik*) is an endangered, polysynthetic language, spoken on St. Lawrence Island of Alaska and the Chukotka Peninsula of Russia. It has undergone a radical language shift in the past few decades, with the youngest generation largely abandoning Yupik in favor of English and Russian. There is overt community interest on St. Lawrence Island in Yupik language revitalization, specifically in developing modern technologies to facilitate language-learning. We present one such resource: a morphologically-aware web-based version of the Badten et al. (2008) Yupik-English dictionary.

Ivalghiit Naalkigalnguut

St. Lawrence Island / Central Siberian Yupik Dictionary

A	E	F	G	GG	GH	GHH	H	I	K	L	LL	M	MM	N
NN	NG	NGNG	P	Q	R	RR	S	T	U	V	W	WH	Y	Z

Enter Search

Yupik-English   English-Yupik   Search

Show IPA   Show Cyrillic

Language Background

Yupik is a polysynthetic language with a relatively high average number of morphemes per word (Schwartz et al., 2019) that typically adheres to the following template:

Root + Derivation + NEG + TMMA + Infl + Encl

One such example word is:

angyasqughhalgunghitukung

angyagh-	-squghhagh-	-leg-	-ngu-	-nghite-	-tu-	-kung
boat-	-small-	-one.that.has-	-to.be-	-to.not-	-INTR.IND-	-1DU

'*We₂ don't, or didn't, have a small boat*'   (Jacobson, 2001, pg. 43)

Features of the Electronic Dictionary

Implemented using HTML, Javascript, and CSS, the dictionary has several search functionalities, exemplified in the seven examples given below. Because Yupik is a polysynthetic language, and handling of multi-morphemic word forms is critical, the most significant feature is perhaps (6) which demonstrates the electronic dictionary's capacity to perform morphological analysis prior to searching. For each matching result, the user is presented with the corresponding entry from the paper dictionary in the following format:

> **Yupik Morpheme** (Cyrillic Transcription) [IPA Pronunciation] *Part of Speech* - English Definition
> *Example sentences and/or notes about word origin and usage*

Search Functionalities

(1) Users may search by letter, by selecting it from the search interface. For example, after the user selects the letter *S*, all words beginning with *s* are displayed. The first three results are shown here.

sa (cя) [sɑ] *noun root* - what?; something; relative (kin)
*"sameng piyugsin? 'what do you want?'; sanguzin? 'what are you?'; sa tamaghhaan 'everything'; savut 'our relative', 'our thing(s)'; relative case is saam (rather than *sam); the expanded base sangaa- (underlyingly sangau-) rather than sangu- is used for asking what something (3rd person) is; thus: sangaawa? 'what is it?'; cf. sangwaa"*

saa (cя) [sɑ] *particle* - I don't know; it doesn't register in my mind; never mind
this exclamation is often pronounced with vowel as in English "hat"; = saami"

sa- (cя-) [sɑ] *verb root* - to do what?; to do something
"saa? 'what did he do?'; aatkaghten saat aghvingisafki 'what happened to your clothes when you didn't wash them?; ayveq saa guusavgu? 'what did the walrus do when you shot it?'; saaqat? 'what's going on?, what are they doing?'"

(2) The user can enter a search term either in Yupik or in English. Here, partial results are shown for the incomplete Yupik search term *aghna*. Yupik words containing that substring are displayed.

aghnagan (aгнаган) [ɑʁ.nɑ.ɣɑn] *particle* - hurry up
aghnagan uyuq emta 'hurry up you, knowing how you are'"

aghnagh- (aгнаг-) [ɑʁ.naʁ] *verb root* - to wear a dress
aghnaghluni 'wearing a dress'; direct verbalization of aghnaq 'woman'

(3) The user may alternatively search in English. Partial results are shown for the search term *family*. Note that derivational suffixes containing the English search term are returned in addition to roots.

-nkuk / -nkut (-нкук -нкут) [nkuk.n.kut] *noun-elaborating postbase* - N and partner; N and associate(s); N and family
see %e)nkuk / %(e)nkut

aalghaq (аӆгак) [ɑɬ.ʁɑq] *noun root* - another family in the same clan; the other of a pair of boats cooperating in a hunt; hunting partner; second wife

(4) Yupik searches may also return entries corresponding to Yupik derivational suffixes. Partial results are shown for the search term *nkut*.

-nkuk / -nkut (-нкук -нкут) [nkuk.n.kut] *noun-elaborating postbase* - N and partner; N and associate(s); N and family
see %e)nkuk / %(e)nkut

kinkut (кинкут) [kin.kut] *noun root* - who? (plural)
look under kina

(5) The user may search for fully inflected Yupik words. Results are shown for the search term *nagatunga* 'I listened.'

nagate- (нагаты-) [nɑ.ɣɑ.tə] *verb root* - to listen
nagatuq 'he is listened'; nagataa 'he listened to her' /

(6) Preliminary support is included for multi-morphemic Yupik word searches. Partial results for the Yupik word *mangteghaghrugllangllaghyunghitunga* 'I didn't want to make a huge house' are shown.

mangteghagh- *verb root* - to make a house for; to reassemble a house (in fall after its parts have been aired out)
direct verbalization of mangteghaq 'house'

–ghrugllak *noun-elaborating postbase* - big N; large N
mangteghaghrugllak 'big house', mangteghaghrugllaget 'very big houses' (from mangteghaq 'house')

(7) Preliminary support is included for Yupik searches using the Cyrillic orthography. Partial results for the Yupik word *qikmiq* are shown, where the original search was performed in Cyrillic.

qikmighaq (кикмигак) [qik.mi.ʁɑq] *noun root* - puppy

qikmiq (кикмик) [qik.miq] *noun root* - dog

Community and Research Impacts

Versions of the print dictionary have been available through the University of Alaska Press in various editions since 1983. However, the print edition is bulky and relatively expensive; while the school libraries in the St. Lawrence Island villages have copies, nearly all community members do not. As a result, the electronic dictionary should facilitate access to knowledge that was otherwise difficult to obtain, making it readily available to everyone in the community.

We expect the electronic dictionary to most positively shape the language experience of the younger generations, promoting language use among youth who have not yet mastered the polysynthetic aspects of the Yupik language and who may be unlikely (for social reasons) to ask elders about word-forms. The dictionary should be useful for linguists and researchers as well, and its capacity to perform morphological analysis may greatly expedite the construction of morphological interlinear glosses, which are critical for the processing and sharing of linguistic data.

References

Linda Womkon Badten, Vera Oovi Kaneshiro, Marie Oovi, and Christopher Koonooka. *St. Lawrence Island / Siberian Yupik Eskimo Dictionary*. ANLC, 2008.

Steven A. Jacobson. *A Practical Grammar of the St. Lawrence Island/Siberian Yupik Eskimo Language*. Alaska Native Language Center, Fairbanks, Alaska, 2nd edition, 2001.

L. Schwartz, E. Chen, B. Hunt, and S. Schreiner. Bootstrapping a neural morphological analyzer for SLI Yupik from a finite-state transducer. In *Proc. ComputEL*, Feb 2019.

Acknowledgements