A morphologically aware offline mobile electronic dictionary for St. Lawrence Island Yupik

Benjamin Hunt[1], Emily Chen[2], Lane Schwartz[2],
Wills Davis[1], Sylvia L.R. Schreiner[1], Scott Schreiner
[1]George Mason University [2]University of Illinois at Urbana-Champaign

Background

Yupik

- St. Lawrence Island Yupik (*Akuzipik/Yupigestun*)/Central Siberian Yupik (*Masingkestun*) (ISO 639-3:ess); Inuit-Yupik-Unangan
- Endangered (EGIDS 7), ~1000 speakers of an ethnic group of ~2500 individuals (St. Lawrence Island, AK; Chukotka, Russia; Alaskan Mainland)
- Rapid shift towards English in younger generations since 1990s
- Polysynthetic, ergative-absolutive, relatively free word order

The project

- *documenting – digitizing legacy materials – creating computer tools for speakers and researchers – aiding in community revitalization efforts*

The dictionary

- Badten, et al. (2008): two-volume Yupik-English dictionary, including roots ("bases") and derivational morphology ("postbases")
- Not widely available in the St. Lawrence Island community
- Was digitized; files are large—still not accessible for most people

Goals/motivations for the electronic dictionary

- Get the dictionary into the hands of community members
- Reduce the effort needed to use/access it
  - Currently only one print version exists in the school library in Gambell
  - No other community members have a personal copy

- Use readily available, free resources to create a useful, professional looking tool
- Many community members expressed unwillingness to ask older speakers for small clarifications or specific vocabulary items due to social pressures. The dictionary offers them another option.

Ultimately, it must be…
- Accessible with a smart phone and without internet access after the initial download
- Capable of taking full word forms as input and returning a parsed root
- Easily maintainable by the community, long term

  - Easy to add and edit entry data
  - Require little to no training
  - Comprehensive documentation to facilitate future maintenance

Development (v1 – functional)

- Implemented in vanilla HTML, JavaScript, and CSS
  - All documentation accessible online (w3schools, etc.)
- Simple string search function
  - Can be toggled between English and Yupik search
- Integrated finite state transducer for morphological parsing
  - Developed by Chen & Schwartz (2018) using FOMA FST toolkit
  - Translated to JavaScript and added to search routine
- Integrated Yupik Cyrillic orthography converter and IPA transcriber
  - These can also be toggled on and off to declutter the interface for users who don't read either system
- Alphabetical view
  - Display all words beginning with a given letter

Ivaghilghiit Naalkenaaghaat

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

Yupik-English   English-Yupik   Search

Show IPA   **Show Cyrillic**

qavagh- (қаваг-) [qɑ.vaʁ] *verb root* - to sleep; to fall asleep
qavaghtuq 'he slept' /

qavaghni- (қавагни-) [qɑ.vaʁ.ni] *verb root* - to be sleepy
qavaghniiq 'he is sleepy' /

qavaghsiq (қаваχсиқ) [qɑ.vɑχ.siq] *noun root* - yearling bearded seal (Erignathus barbatus)
cf. qavagh-

qavaghtagh- (қаваχтаг-) [qɑ.vɑχ.taʁ] *verb root* - "to stay several nights in the ocean, ice or wilderness; to camp out"
angyat mellaalluki taan'gani qavaghtaghtut 'a boat was trapped by ice in the ocean so they stayed several nights out there'; nunaaghmi yuuk ullghitiighulluni qavaghtaamalghii 'a man broke down on his wheeled/tracked vehicle in the tundra and stayed several nights out there'

qavaghvik (қавагвик) [qɑ.vaʁ.vik] *noun root* - bed; bedroom; place to sleep

qavaghyugtuq

<= qavagh- @~f+yug- -tuq

Yupik-English   English-Yupik   Search

qavaghtagh- (қавахтаг-) [qɑ.vɑχ.taʁ] *verb root* - "to stay several nights in the ocean, ice or wilderness; to camp out"
angyat mellaalluki taan'gani qavaghtaghtut 'a boat was trapped by ice in the ocean so they stayed several nights out there'; *nunaaghmi yuuk ullghitiighulluni qavaghtaamalghii* 'a man broke down on his wheeled/tracked vehicle in the tundra and stayed several nights out there'

qavaghvik (қаваг̇вик) [qɑ.vɑʁ.vik] *noun root* - bed; bedroom; place to sleep
No information given.

qavaghwaagh- (қаваӱг̇аг̇-) [qɑ.vɑ.ʁʷɑːʁ] *verb root* - to oversleep
No information given.

@~f+yug- (фьюг-) [fjuɣ] *verb-elaborating postbase* - to want to V; to want some N; to tend toward V-ing; to feel V (with emotional roots)
tagiyugtuq 'he wants to come over' (from *tagi-* 'to come over'); *neghyugaa* 'he wants to eat it' (from *negh-* 'to eat'); *taaqsugtuq* 'he wants to quit' (from *taaqe-* 'to quit'); *aaghhwsugtuq* 'he wants to crawl' (from *aaghhwe-* 'to crawl'); *tuqusugaa* 'he wants to kill it' (from *tuqute-* 'to kill'); *laalighsugaa* 'he wants to visit him' (from *laalighte-* 'to visit'); *atughyugaa* or *atuqsugaa* 'he wants to use it' (from *atugh-* or *atuqe-* 'to use'); *qupneghyugtuq* 'he wants some whitefish' (from *qupneq* 'whitefish'); *akuutaghyugtuq* 'he want some Eskimo ice cream' (from *akuutaq* 'Eskimo ice cream'); *kamegyugtuq* 'he wants skin-boots' (from *kamek* 'skin-boot'); *tuquyugtuq* 'he is deathly ill' (from *tuqu-* 'to die'); *qinuyugtuq* 'he is sick; he is overwhelmed by the tasks facing him' (from *qinu-* 'to be disinclined to act'); *yimiyugtuq* 'he feels shy, intimidated' (from emotional root *yimi-* 'shy, intimidated')

@~f+yuggaagh- (фьюхаг̇-) [fju.xɑːʁ] *verb-elaborating postbase* - to first V; to V in the beginning
Kii, Sivungaq neliyuggaaghaqngan iyataghniituq. 'Golly, while Savoonga was being established there was no cause for hunger.' (from *neli-* 'to establish (a) household(s)') (SIVUO, UNGIP, 19); = *-yuwhaagh-*

Development (v2 – in development)

- Implemented in vanilla HTML, JavaScript, and CSS
  - All documentation accessible online (w3schools, etc.)
- More robust search function
  - Still toggleable between English and Yupik search
  - Using JSON database (eventually moving to XML for longevity and compatibility)
  - Predictive search using string match/Levenshtein distance and word weights
- Integrated FST parser
  - Display parsed morphemes above search results
- Yupik Cyrillic orthography and IPA now stored in entry data
- Dynamically generated entry pages with full entry data (examples, paradigms, etc.)

Yupigestun Sangaawa
A dictionary for St. Lawrence Island Yupik

qavaghtuq

Search

parse   qavagh-   -~$_f$(g/t)ugh   -Ø

qavagh-, n. (қавағ-) [qɑ.vaʁ]   to sleep; to fall asleep

qavaghtuq 'he slept'...

qavaghni-, n. (қавағни-) [qɑ.vaʁ.ni]   to be sleepy

qavaghniiq 'he is sleepy'...

Implications and next steps

Hope to achieve...

- widespread dissemination within the community
- active feedback from community members regarding functionality and new additions/edits to existing entries

Improvements:

- Search functionality (results sorted by relevance; predictive search)
- "Dictionary feel"
  - Full entry-devoted page with etymological data, usage examples from corpus, expandable morphological paradigm tables, etc.
- Display parsed output above search results
- Integrate neural-net-based morphological analyzer currently in development (Schwartz, et al. 2019)
- "Word wheel"
  - Display nearby words in the dictionary and/or suggest related words in sidebar

Acknowledgments

References

Badten, L., V. Kaneshiro, M. Oovi, & C. Koonooka. 2008. St. Lawrence Island / Siberian Yupik Eskimo dictionary. U of Alaska Fairbanks: ANLC.

Chen, E. & L. Schwartz. 2018. A morphological analyzer for St. Lawrence Island/Central Siberian Y*upik. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018).

Hunt, Benjamin, Emily Chen, Sylvia L.R. Schreiner, and Lane Schwartz. 2019. Community lexical access for an endangered polysynthetic language: An electronic dictionary for St. Lawrence Island Yupik. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations), 122-126. https://doi.org/10.18653/v1/N19-4021.

Schwartz, Lane, Emily Chen, Benjamin Hunt, and Sylvia L.R. Schreiner. 2019. Bootstrapping a Neural Morphological Analyzer for St. Lawrence Island Yupik from a Finite-State Transducer. In *Proceedings of the 3rd Workshop on Computational Methods for Endangered Languages: Vol. 1, Article 12*, 85-96. https://scholar.colorado.edu/scil-cmel/vol1/iss1/12