

Tools for the maintenance and revitalization of St. Lawrence Island Yupik

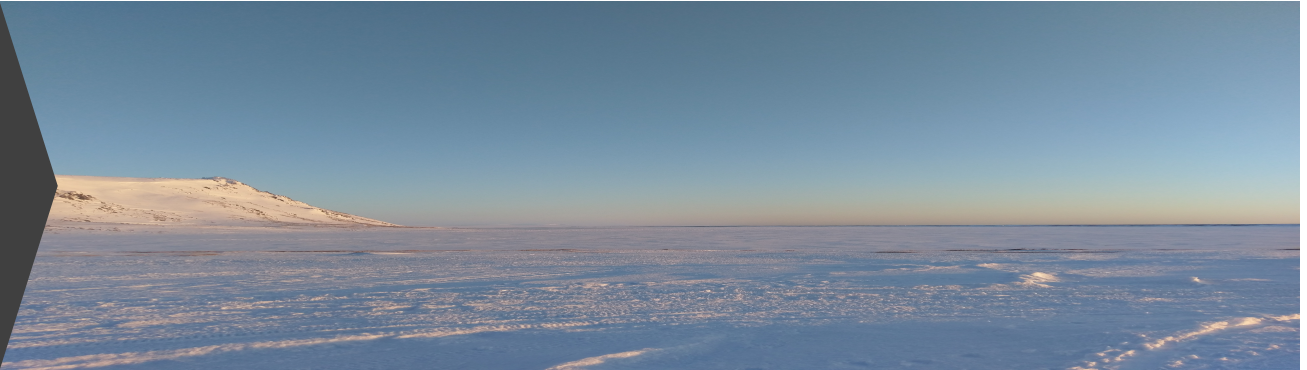
Lane Schwartz¹, Sylvia L.R.
Schreiner², Emily Chen¹, Hyunji
Hayley Park¹, Peter Zuckerman¹,
Benjamin Hunt², Giulia Masella
Soldati²

¹University of Illinois at Urbana-Champaign

²George Mason University

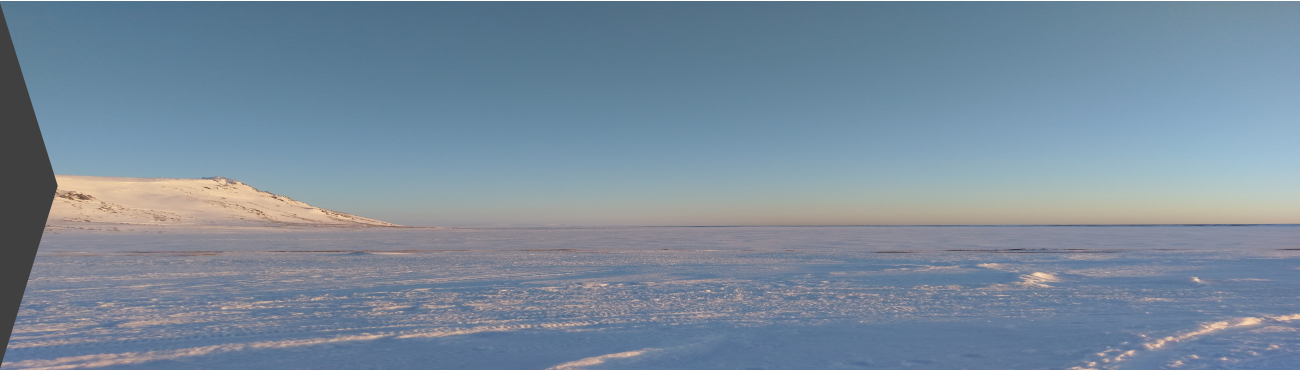


Background



- St. Lawrence Island Yupik / Central Siberian Yupik / *Akuzipik* (ISO 639-3:ess)
- Yupik branch of Inuit-Yupik-Unangam family
- Endangered, under-resourced language spoken on St. Lawrence Island, AK; Chukotka peninsula of Russia; mainland AK (~1000 speakers) (Schwartz et al. 2020)

Background



- Rapid generational shift since 1990s; youth now largely L1 English speakers (Schwartz et al. 2019)
- Polysynthetic, ergative-absolutive (case), relatively free word order



Revitalization Efforts

- Revitalization group organized by the Yupik-speaking community in Gambell
- Community hopes to eventually establish an immersion curriculum



Revitalization Efforts

- Our work:
 - (1) Digitizing existing Yupik-language materials and developing a community-accessible repository for these materials; and
 - (2) Developing a suite of tools for use by speakers of the language.
- We are attempting to develop these tools in such a way as to lower the barrier to entry by community language activists.

Digitization



- Materials at the Alaska Native Language Archive
 - 44 of the 55 sets of items filed under 'St. Lawrence Island Yupik' or 'Chaplinksi'
- Materials stored in the Materials Development Center at the school in Gambell
 - 67 comb-bound texts, 3 elementary short story collections, and 7 collections of oral narratives



Digitization

- We are building a digital corpus of these materials (Schwartz, et al. 2021)
 - 41,310 Yupik sentences
 - 269,878 total words
 - 87,137 unique words

Nagaten .

Nagaqughsigu-u ?

Nakaa .

Sangaawa ?

Esghaqaghhuqun tazigna .

Maaten nagaqughaqa .

Enta aqfaatelta tazingavek .

Hilikaptera .

Digitization



- **Goals:** Make materials accessible to community members; support the development of language technologies

1. Listen.

Do you hear it?

2. No.

What is it?

3. Look over there.

4. Now I hear it.

Let's run over there.

5. Helicopter.

Orthography & Pronunciation Tool

- We developed a web-based utility (Schwartz & Chen 2017) that serves as
 - orthotactic spell-checker
 - pronunciation tool
- Spelling errors are in red

St. Lawrence Island / Central Siberian Yupik

Latin	Doubled	Phonetic Transcription	Cyrillic	Stress Patterning
<p>Alcuzilleq: Taawa umiilek akuziiquq. "Avaqutaghpesi-i sanqutangit aqfayulghiiim aqfaliki. Yugighsaghtegnuq tuqutuumaaglce!"</p> <p>Enter</p>				

Latin

Alcuzilleq:

Taawa umiilek akuziiquq,

Avaqutaghpesiii sanqutangit aqfayulghiiim aqfaliki.

Yugighsaghtegnuq **tuqutuumaaglce!**

Orthography & Pronunciation Tool

- There are several pronunciation (phonetic transcription) options to choose from:
 - standard IPA
 - systems used by other Yupik scholars

Phonetic (IPA)

alcuziłəq
ta:ɣ^wa umi:lək akuzi:quq
avaqutaχpəsi:i sanqutaŋit aqfajulɬi:m aqfaliki
juɣiχsaχtəɣŋuq tuqutu:ma:ɣlcət

Phonetic (Krauss 1975)

alcuziłəq
taay^wa umiilək akuziiquq
avaqutaχpəsiii sanqutaŋit aqfayulɣiim aqfaliki
yuyiχsaχtəɣŋuq tuqutuumaaylcət

Phonetic (Nagai 2001)

alcuziłəq
taay^wa umiilək akuziiquq
avaqutaχpəsiii sanqutaŋit aqfayulɣiim aqfaliki
yuyiχsaχtəɣŋuq tuqutuumaaylcət

Orthography & Pronunciation Tool

- This utility can also toggle between the standard Latin and Cyrillic orthographies

Latin

Alcuzilleq:

Taawa umiilek akuziiquq,
Avaqutaghpesiii sanqutangit aqfayulghiim aqfaliki.
Yugighsaghtegnguq **tuqutuumaaglcet!**

Cyrillic

Алсузильық:

Тәуә умйлык акузййкуқ,
Авакутахпысйи сянкутанит ақфаюлғйим ақфалики.
Югйхсяхтыгнук тукутүмәглсыт!

Orthography & Pronunciation Tool

- It also has an option for a Latin orthography that is more orthographically transparent
(as requested by teachers)

Latin

Alcuzilleq:

Taawa umiilek akuziiquq,
Avaqutaghpesiii sanqutangit aqfayulghiim aqfaliki.
Yugighsaghtegnguuq **tuqutuumaaglcet!**

Doubled

Alcuzilleq:

Taawa umiilek akuziiquq,
Avaqutag**hh**pesiii sanqutangit aqfayulghiim aqfaliki.
Yugi**gh**sag**hh**tegnguuq tuqutuumaaglcet!

Finite-State Parser & GUI

- We developed a finite state morphological parser, which takes a Yupik word as input and returns its possible morphological breakdowns (Chen & Schwartz 2018)

siqineghllugughsimakanga

siqineghllug - ~:(ng)ughte - @~:(i/u)ma - ke - nga

excess.sun - for.there.to.start.to.be - PAST - PARTICIPIAL - 3sg

'the sun came out' (Apassingok et al. 1993)



Finite-State Parser & GUI

- This parser is an implementation of Steven Jacobson's (2001) reference grammar
- It was first developed to help non-speaker researchers parse or build Yupik words
- This parser is the precursor to several other potential language technologies, such as a lexical spellchecker and predictive text


Finite-State Parser & GUI

- The finite-state parser can be used to break down or build up complex words
- It was initially only usable at the command line

```
[foma[1]: apply up akuziquq  
akuzi(V)^~(g)aqe(V→V)^[Ind.Intr]^[3Sg]  
akuzi(V)^~i(V→V)^+qe(V→V)^[Ind.Intr]^[3Sg]  
[foma[1]: apply down akuzi(V)^~(g)aqe(V→V)^[Ind.Intr]^[3Sg]  
akuziquq  
foma[1]: █
```

Finite-State Parser & GUI

- But we have since designed a GUI that is much more user-friendly

Ivaghilghiit Naalkenaaghaat 

Morphological Analyzer for St. Lawrence Island Yupik

Enter Yupik word

Enter base + Part of Speech Case/Mood Person/Number

Parse **Generate** **Engage**



Neural Parser

- We are developing a *neural* morphological parser using machine learning methods (Schwartz, et al. 2019)
- The neural parser is expected to improve upon the finite-state parser:
 - Alternate spellings
 - Undocumented morphemes
 - Exceptions to documented morphotactics and morphophonology

Web-Based Dictionary

- We also implemented a web-based digital version (Hunt, et al. 2019) of the Yupik-English dictionary (Badten, et al. 2008)
- The user can search in both Yupik and English

The screenshot shows the interface for the 'Ivaghilghiit Naalkenaaghaat' dictionary, which is the 'St. Lawrence Island / Central Siberian Yupik Dictionary'. The interface features a header with the title and a row of social media icons. Below the header is a search bar with a grid of letters for navigation. The search bar contains the text 'Enter Search'. Below the search bar are four buttons: 'Yupik-English', 'English-Yupik', 'Search', and 'Show IPA'. The 'Show IPA' and 'Show Cyrillic' buttons are also visible.

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



Web-Based Dictionary

- Moving this resource to the web means community members (most of whom have smartphones) can access the dictionary
- This dictionary also employs the parser, so users can search for an inflected word
- We hope to later allow users to choose from a variety of outputs

Interactive E-Books & GUI

- We are creating interactive e-books from digitized elementary primers that integrate text, images, and audio
- Users can read the words and listen to a Yupik speaker read the story
- Readers can also click on each Yupik word to hear its pronunciation

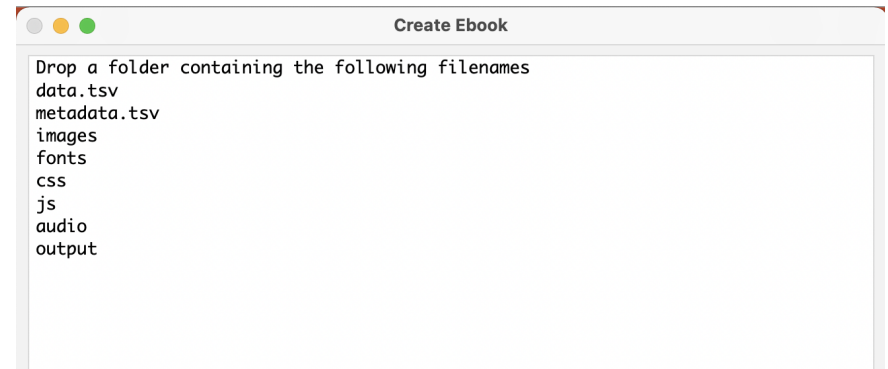
Una Tinana.




Una Tinana.

Interactive E-Books & GUI

- We are also developing a GUI for this process.
- With limited training, community members will be able to learn how to edit images, text, and audio files, and "drag and drop" the files to create interactive e-books.






Read-along
Pledge of
Allegiance

NEKEVGHAVIGANUN, **ATAASIQ**
NUNAGHLLAK ASINGANI



Future Work

- Completing digitization of all relevant materials at ANLA
- Improvements to the neural morphological parser
- Improvements to the dictionary: offline app, additional functionality
- Completing e-books
- Additional apps and pedagogical tools as requested by the community
- Language investigation and archiving tool (corpus, lexicon, parser functionality, etc. in user-friendly interface)



For more
information...

- <https://github.com/SaintLawrenceIslandYupik>
- <http://www.sylvialrschreiner.com/research/yupik-project/>
- <https://ldl.linguistics.gmu.edu/>

Acknowledgments

The authors' institutions lie on unceded land traditionally held by the Powhatan Algonquian, Manahoac, and Monacan peoples (Fairfax, Virginia), and by the Kiiikaapoi (Kickapoo), Myaamia, Očhéthi Šakówinj, and Peoria peoples (Champaign-Urbana, Illinois).

Special thanks to the Yupik speakers who have shared their language and culture with us. Igamsiqayugvikemsi. We take sole responsibility for any errors. Part of a collaborative research project also including RAs Preston Haas, Emma Evans, David Ngo, and Kaelynn Mae Kenny (GMU). Portions of this work were funded by NSF Documenting Endangered Languages Grants #BCS 1761680 and 1760977, a Mathy Junior Faculty Award in the Arts and Humanities, and a George Mason University Presidential Scholarship, and a GMU CHSS Faculty Research and Development Award. Many thanks to the members of the GMU Language Documentation Lab Group for their help with data processing.

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 Yupik. 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, L., & E. Chen, E. 2017. Liinnaqumalghiit: A Web-based Tool for Addressing Orthographic Transparency in St. Lawrence Island/Central Siberian Yupik. *Language Documentation and Conservation* 11, 275-288.
- 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>
- Schwartz, L., E. Chen, H.H. Park, E. Jahn, & S.L.R. Schreiner, 2021. A Digital Corpus of St. Lawrence Island Yupik. 2021. *Proceedings of Comput-EL 4*. arXiv preprint arXiv:2101.10496.
- Schwartz, L., S.L.R. Schreiner, & E. Chen. 2020. Community-focused language documentation in support of language education and revitalization for St. Lawrence Island Yupik. *Études Inuit Studies* 43(1/2). <https://doi.org/10.7202/1071949ar>.