Community-Focused Language Documentation in Support of Language Education and Revitalization for St. Lawrence Island Yupik

Lane Schwartz,[i] Sylvia L.R. Schreiner,[ii] and Emily Chen[iii]

ABSTRACT

St. Lawrence Island Yupik, an endangered language of the Bering Strait region spoken by fewer than one thousand people in western Alaska and far eastern Russia, is currently in a state of generational transition. We survey the existing body of Yupik literature and pedagogical resources developed during the twentieth century, examine the context and use of Yupik in the current educational setting, and describe current challenges for teaching the language in the schools. We then outline our integrated approach to language documentation currently being applied to Yupik, and address how existing resources can be integrated into research and development processes in a way that both supports research efforts and results in tangible modern educational tools for the Yupik community on St. Lawrence Island, and eventually in Russia. This approach is intentionally designed to closely integrate research processes from language documentation and computational linguistics such that the results of each research endeavour positively support the other, and such that both disciplines concretely support community-based efforts to revitalize and teach the language.

KEYWORDS
St. Lawrence Island Yupik, endangered languages, language education, computational tools, language revitalization

RÉSUMÉ
La documentation linguistique axée sur la communauté en soutien à de l'éducation et à de la revitalisation pour le yupik de l'île Saint-Laurent

Le yupik de l'île Saint-Laurent, une langue du détroit de Béring en péril, (moins de 1 000 locuteurs habitent dans l'ouest de l'Alaska et l'extrême est de la Russie)

i. Department of Linguistics, University of Illinois, Urbana-Champaign, Illinois, United-States. lanes@illinois.edu
ii. Department of English, George Mason University, Fairfax, Virginia, United-States. sschrei2@gmu.edu
iii. Department of Linguistics, University of Illinois, Urbana-Champaign, Illinois, United-States. echen41@illinois.edu

est actuellement en période de transition générationnelle. Cet article examine la littérature disponible sur le yupik et les ressources pédagogiques qui se sont développées au cours du XXe siècle. Il étudie également le contexte et l'utilisation du yupik dans le cadre éducatif actuel et décrit les défis liés à l'enseignement de la langue dans les écoles. On décrit ensuite l'approche intégrée de la documentation linguistique appliquée actuellement au yupik et aborde aussi les façons dont les ressources existantes peuvent être intégrées aux processus de recherche et de développement, de manière à soutenir les efforts de recherche et produire des outils pédagogiques modernes et tangibles pour la communauté yupik de l'île Saint-Laurent (et à plus largement en Russie). Nous avons conçu cette approche pour intégrer étroitement les processus de recherche de la linguistique documentaire et de la linguistique computationnelle, de sorte que les résultats de chaque projet de recherche puissent se soutenir mutuellement, et appuient concrètement les efforts communautaires de revitalisation et d'enseigner de la langue.

******

S t. Lawrence Island Yupik, while thriving in many ways, is currently in a period of rapid generational transition. While St. Lawrence Islanders born before 1980 are nearly all L1 Yupik speakers, the situation has changed drastically in the generations born since that point. The current youth on St. Lawrence Island are predominantly L1 English speakers with limited Yupik proficiency.

In this paper, we examine the context and use of St. Lawrence Island Yupik in the schools and in the broader community. We detail some of the current challenges surrounding the teaching of the language in the schools, as well as our current efforts to document the language and produce digital materials and computational tools to aid the language education, maintenance, and revitalization efforts of the Yupik-speaking community.

In the rest of this section, we provide background information on the language and its current status. In the second section, Educational Resources, we discuss the present educational situation, including challenges and desired changes. Then, in the third and fourth sections—Computational Tools for Yupik and An Integrated Approach in Support of Yupik Language Documentation and Education, respectively—we describe the tools we have developed and the processes we have undertaken to gather language data and existing materials, and to use these resources to produce useful and accessible materials and tools for the language and its speakers.

Language background and status

The Inuit-Yupik language family spans the northern coast of North America, ranging from Greenland to the Chukotka Peninsula of Russia; it spans the entirety of northern Canada, as well as northern, western, and south-central Alaska (Krauss et al. 2011). The Yupik branch of this family includes Sugpiaq in south-central Alaska, Central Yup'ik in western Alaska, Naukan in Chukotka, and a fourth variety spoken both in Chukotka and on St. Lawrence Island. In this work we examine this fourth member of the Yupik language family, ISO 639-3 ess, an endangered language of the Bering Strait region. During the twentieth century, this language was called Central Siberian Yupik or St. Lawrence Island Yupik in the English-language linguistics literature, and Chaplinski Yupik in the Russian-language literature (so named for the Russian name of the Chukotkan Yupik old village of Ungaziq); the language is called Yupik, Yupigestun, or Akuzipik on St. Lawrence Island (de Reuse 1994). Because our work directly involves only the St. Lawrence Island variety and because we consider the label "Central Siberian" to be inapt, we will use the term *St. Lawrence Island Yupik*, or simply *Yupik*, using the term *Chaplinski Yupik* when necessary to refer to the Chukotkan variety. Readers with an interest in the grammar of Yupik may wish to consult, for example, Krauss (1975), Jacobson (1977, 1985, 2001), Krauss and colleagues (1985), and de Reuse (1994). A substantial amount of grammatical documentation and texts also exist for related languages, especially Central Alaskan Yup'ik (see Jacobson 1995 and Miyaoka 2012).
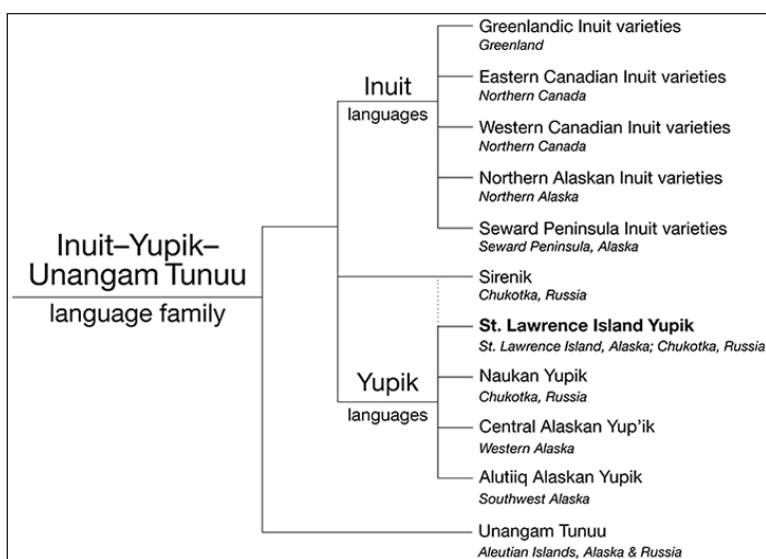
Figure 1. Inuit-Yupik language family.

The term *Yupik* will also be used to refer to a Yupik person, with the plural *Yupiget* referring to multiple persons. There are currently two settlements on St. Lawrence Island, the Yupik villages of Sivuqaq (English: Gambell) and Sivunga (English: Savoonga). Nearly all of St. Lawrence Island's 1,300 permanent residents are Yupik; another 300 to 400 Yupiget reside on the Alaskan mainland (Schwalbe 2017). In Russia, most Yupik settlements were closed during the early to mid-twentieth century as a result of Soviet relocation programs (Krupnik and Chlenov 2013); today, most of the 800 Russian Yupiget live in the villages of Novoe Chaplino and Sirinek (Schwalbe 2017). Despite the geographic and political separation, the differences between the varieties of Yupik on St. Lawrence Island and Chukotka are generally assumed to be relatively minor (Krauss 1975).

Yupik is currently in a state of rapid language shift that began in Russia in the late 1950s (Krupnik and Chlenov 2013) and in Alaska in the late 1990s (Koonooka 2005). Substantial Yupik language materials were developed in Russia between the 1930s and the 1950s (Krauss 1971; Sevil'gaev 1985), and in Alaska between the 1970s and the 1990s using federal funding through the Title VII Bilingual Education Act (Public Law 90-247 1967). In 1980 University of Alaska linguist Michael Krauss reported that St. Lawrence Island Yupik was, at the time, "the only Alaskan language which is still being learned by all the children" of that language's ethnic group (Krauss 1980, 105). Using 2010 US Census demographic reports and using a birth year of 1980 as a delineator, we estimate that (as of 2010) there were approximately 540 Yupiget on St. Lawrence Island (representing 41 per cent of the island's Yupik population at that time) who were definitely fluent L1 Yupik speakers. The use of Yupik among younger generations has substantially declined in recent decades (Koonooka 2005; Morgounova 2007), resulting in correspondingly decreasing percentages of L1 Yupik speakers over time in the generations born after 1980. The decrease in Yupik usage among Russian Yupiget began much earlier; Vakhtin (2001) estimated no more than 200 fully fluent Yupik speakers in Chukotka, the youngest of whom would now be in their seventies. In total, we estimate that there are currently 800 to 900 fully fluent L1 Yupik speakers out of an ethnic population of between 2,400 and 2,500. There exist no up-to-date estimates of the number of children on St. Lawrence Island currently learning Yupik as their L1 at home. Based on our recent fieldwork in Gambell and personal communication with the Gambell school administration and bilingual staff, it seems clear that most children on St. Lawrence Island are now English-dominant, with relatively few, if any, learning Yupik as their L1. During our recent fieldwork (November 2016, July 2017, November 2017, July 2018, November 2018), we did not encounter any Yupik-dominant children, and at this point it seems very possible that no children on St. Lawrence Island are Yupik-dominant.

Educational Resources

During field visits to St. Lawrence Island conducted since November 2016, we met with various community stakeholders to discuss the language situation. Across generations, we have observed an explicit stated desire for preservation, documentation, and revitalization of the St. Lawrence Island Yupik language. Community members expressed two concrete goals: increased use of Yupik-language educational materials in the school, and the eventual establishment of a Yupik immersion program.

In this section, we examine the legacy educational materials developed for Yupik during the twentieth century, as well as the existing grammars of Yupik, and consider current educational challenges with regard to materials' accessibility, pedagogy, and goals.

Existing Yupik language materials

Numerous Cyrillic-orthography Yupik materials were developed in Russia, primarily from the 1930s through the 1950s (Krauss 1973). The bulk of this material is currently accessible only in the form of original paper books and booklets of Yupik publications, or, in some cases, rebound or loose-leaf photocopies of Soviet publications, stored in the archives of the Alaska Native Language Archive (ANLA) in Fairbanks (Michael Krauss, pers. comm.). While the ANLA index includes many Yupik publications, we have already identified a number of Yupik materials published in Alaska that have not been indexed. Almost none of the numerous Cyrillic-orthography Yupik publications are in the ANLA index (Siri Tuttle, pers. comm.).

During the 1970s, a number of Yupik pre-primers and primers were developed by the Nome Agency Bilingual Education Resource Center of the Bureau of Indian Affairs and by the Alaska Native Language Program at the University of Alaska. A small number of Yupik-language pedagogical resources targeting adults were also developed by Dave Shinen of the Summer Institute of Linguistics (e.g., Shinen 1976).

In the 1980s and early 1990s, a substantial amount of additional Yupik material was developed by the St. Lawrence Island bilingual program staff under the aegis of the Title VII federal Bilingual Education Act. This material includes the bilingual Yupik-English *Lore of St. Lawrence Island* trilogy (Apassingok et al. 1985, 1987, 1989), which collects transcribed and translated oral stories of St. Lawrence Island Elders; the Sulpik stories (Apassingok and Tennant 1987), a collection of short stories whose design and creation was meant to promote literacy among elementary school children; and a series of three bilingual Yupik-English elementary readers (Apassingok et al. 1993, 1994, 1995). This material also includes a full set of Yupik bilingual/bicultural curricula, including lesson plans, for kindergarten through high school grade levels.

While the first descriptions of Yupik were Russian-language publications developed by Soviet linguists (Menovshchikov 1960, 1962, 1967; Rubtsova 1971; Menovshchikov and Vakhtin 1990), the most current and most thorough description of Yupik is the English-language grammar by Jacobson (2001). Other Yupik-language materials include texts that address the basics of the Yupik alphabet (Kaneshiro and Smart 1980; Nome City Schools 1983) and Yupik spelling in the Latin (Tennant 1985a, 1985b) and Cyrillic (Jacobson 1990) orthographies, other elementary readers (Kaneshiro and Kaneshiro 1974, 1975; Kaneshiro and Apatiki 1975; Orr and Wetmore 1975; Poage 1975; Apassingok and Waghiyi 1985a, 1985b, 1985c, 1985d, 1985e), elementary school workbooks (Badten 1974; Poage and Apatiki 1975; Crispin et al. 1976; Hargraves, Oozevaseuk, and Apatiki 1976; Apatiki, Apatiki, and Apangalook 1982; Poage et al. 1983), a series of pre-primers (Blanchett et al. 1972a, 1972b; Teeluk, Chikoyak, and Badten 1972) with accompanying workbooks (Chrispin et al. 1975a, 1975b) and instructional guides, anthologies of folk stories (Slwooko 1977; Oosevaseuk and Waghiyi 1985), a collection of legends (Koonooka 2003), a collection of personal narratives of visits to Chukotka (Kaneshiro and Badten 1975), and other stories (Apassingok et al. 1972; Slwooko and Kulukhon 1975; Pungowiyi and Gologergen 1975; Slwooko, Imergan et al. 1975; Slwooko, Rookook et al. 1975).

Materials' accessibility and use

Almost none of the material described in the previous section is currently being used for St. Lawrence Island Yupik education. Koonooka (2005) describes a confluence of factors that led to the current situation. These factors include the retirement of the most experienced St. Lawrence Island Yupik bilingual staff, coupled with various mandates from the school district, state, and federal levels (most notably the federal No Child Left Behind Act) that de-prioritized Yupik language instruction in favour of other educational aims.

Another major challenge is the physical accessibility of materials. These materials exist primarily in two places: the Alaska Native Language Archive at the University of Alaska Fairbanks and the Materials Development Center storage room in the Gambell School. In Gambell, some of the primers exist in sufficient printed quantities that they could in principle, be rolled out for use at the early elementary level. For some of the other materials, scans are available for download through the Alaska Native Language Archive. However, a substantial amount of the available material is not easily physically or electronically accessible to the Yupik instructors or community members who might make use of them. For example, to the best of our knowledge, high school students in Gambell do not have access to individual copies of the Jacobson (2001) grammar for home study (see Pedagogical Issues below). More broadly, there are members of the St. Lawrence Island Yupik community

on the island and in the mainland Alaskan diaspora who have expressed a desire for convenient access to existing Yupik materials. Numerous community members have also expressed a desire for an accessible Yupik electronic dictionary.

Pedagogical issues

At the early elementary level, a number of pre-primers and primers exist, along with a preliminary edition of an elementary Yupik reading textbook (Tennant 1985a). Yet even when materials are physically or electronically available, many are not well suited to immediate use within modern pedagogical frameworks. The St. Lawrence Island Yupik K–12 bilingual/ bicultural curricula mentioned above, for example, will need to be adapted to fit within the context of the current educational practices adopted by the Bering Strait School District and the Gambell and Savoonga schools. To support this adaptation, the curricula will first need to be digitized, verified, and exported into an easily editable format. We describe our initial work on this process below in Digitization of Printed Materials.

At the high school level, the Yupik grammar of Jacobson (2001) is currently being employed for Yupik language instruction in Gambell, under the direction of an L1 Yupik instructor. This grammar is written at the level of a college textbook, and uses descriptive linguistic terminology to explain the structure of Yupik to a presumed audience of L1 Yupik-speaking college students. Prior to his retirement, Jacobson used this grammar when he taught St. Lawrence Island Yupik at the University of Alaska Fairbanks. The grammar includes coverage of the Yupik Latin and Cyrillic orthographies, noun cases, verb moods, a selection of Yupik roots and derivational suffixes, Yupik morphophonological and phonological processes, and tables of inflectional morphology, demonstratives, and pronouns. The end of each chapter contains basic exercises in translating between the two languages; neither an instructor's guide nor answer keys exist.

The Jacobson grammar is an excellent resource. Despite the advanced target audience of the book and the lack of supplementary materials, it is the only existing English-language textbook available for Yupik language instruction. However, the assumption made by the author of an audience of college-aged L1 Yupik students does not match what is found in practice today—an audience in Gambell of English-dominant high school students.

Goals

One immediate educational objective is the digitization and verification of pedagogical and cultural materials that currently exist in print form only (see below, Digitization of Printed Materials). Our discussions with community members, however, have revealed a longer-term goal of developing and

implementing an immersion curriculum at the lower grade levels. A Yupik-language immersion program would require pedagogical materials for the entire curriculum for the targeted levels (e.g., K–3 or K–6). Middle and high school curricula could be modified towards a dual language immersion program, with some topics taught in Yupik and some in English.

A total immersion experience for younger students would necessarily include core subjects such as math, science, language arts and reading, and social studies, as well as music, arts, etc. in Yupik. Each of these areas presents its own challenges. Modern textbooks on such topics do not currently exist in Yupik. While an early approach might combine instruction in Yupik with English-language textbooks, the eventual goal would be to have the ability to teach from Yupik-language books. In the next section, we examine the potential for fundamental computational enabling technologies to facilitate the more rapid development of such Yupik-language resources. For example, the translation of textbooks into Yupik could be greatly facilitated by the ability to undertake even a passable machine translation, which could then be verified and edited by native speakers. The first step towards machine translation into (or out of) Yupik is the development of a working morphological analyzer; we discuss our work on such an analyzer in Data Sparsity and Morphological Analysis.

The existing early readers and other Yupik-language texts could be modified to be used as supplementary materials. We have begun work on this through digitization and e-book creation, which we discuss in the fourth section below. The bilingual/bicultural curricula (once updated to reflect modern pedagogical aims) could also be implemented in this environment. Many of the scope and sequence materials are written in English, but these could also be translated into Yupik, or even used at first in their existing format (since all Yupik speakers from St. Lawrence Island are bilingual in English) to help teachers phase in the curriculum. Translating all the materials into Yupik is a monumental task for teachers or other speakers to undertake in addition to their regular work. Thus, an early goal for machine translation would be to provide educators with a first-pass Yupik version of the materials. Digitization of these curricula is the first step to increasing their usability in the school setting (we detail our work on this in Community Access to Legacy and New Materials).

Computational Tools for Yupik

In this section, we present the computational resources for Yupik that we have developed thus far. The primary tools available to date are a web-based utility (Schwartz and Chen 2017) and a finite-state morphological analyzer (Chen and Schwartz 2018), respectively described in the first and second sections

below. We also have preliminary implementations of a neural-network morphological analyzer (Schwartz et al. 2019) and an electronic dictionary (Hunt et al. forthcoming).

Orthographic conversions and basic spellchecking

The first known publicly available computational tool for St. Lawrence Island Yupik is our recently developed web-based utility (Schwartz and Chen 2017). It is similar in function and intent to the Kleinschmidt and International Phonetic Alphabet (IPA) converters for Greenlandic.[1] Our utility is implemented as a standalone web page using HTML and Javascript, and can be used online or offline. The functions included in this utility were selected for development based on their capacity to assist with ongoing language learning and pedagogical efforts, as well as their value to linguists conducting fieldwork. For example, the orthographic conversion function allows a Yupik instructor to quickly show a text to students in both the standard Latin orthography and in a fully orthographically transparent variant; this functionality was created in direct response to feedback from Yupik language instructors that some L1 English students consistently mispronounce words that are not fully orthographically transparent. Similarly, the IPA conversion functionality allows field linguists to rapidly access the IPA transcription of an entire sentence or paragraph.

The utility accepts Yupik text written in the St. Lawrence Island standard Latin orthography (developed by linguists working with Yupiget in the 1970s) or in the Chaplinski Yupik Cyrillic orthography (developed several decades earlier). Based on this input, the utility performs basic spellchecking using a non-lexical orthotactic spellchecker. For example, long vowels in Yupik are represented orthographically as double letters. However, no diphthongs surface phonemically in Yupik; therefore, any word containing two unlike adjacent vowels is necessarily misspelled. Yupik words provided by the user that fail to conform to Yupik orthotactic standards are highlighted in red.

The user can also select from a set of optional transforms to be applied to the Yupik input. The first addresses a particular absence of transparency in the Latin orthography. In this orthography, when two unvoiced consonants appear next to each other in a word, in most cases one of the two consonants is written using the grapheme for its voiced counterpart, a process called *orthographic undoubling* (Krauss 1975; Jacobson 2001). While orthographic undoubling was intuitive to speakers when the Latin orthography was first developed and introduced (Krauss 1975), this is no longer necessarily true for students learning Yupik as an L2. Yupik instructors have reported that some

---

1. See the converters at https://oqaasileriffik.gl/langtech/kleinschmidt/ and https://oqaasileriffik.gl/langtech/ipa/.

students struggle with undoubling, and when reading a Yupik word with an undoubled consonant will incorrectly pronounce the consonant as voiced (as the grapheme appears to indicate) rather than unvoiced (as the phonological rule requires). If this transform is selected by the user, the utility returns an orthographically transparent rendering of any words containing orthographic undoubling, thus allowing the user (such as a student learning Yupik) to contrast the orthographic standard form of a word with its fully transparent counterpart as an asset in language-learning.

The second available set of transforms is primarily targeted at linguists working on St. Lawrence Island Yupik. Because of the unambiguous design of the St. Lawrence Island Yupik orthography, it is possible to deterministically convert each Yupik word into an equivalent phonological transcription. The default transform converts a word from the standard Latin or Cyrillic orthography into the equivalent phonemic representation in IPA. We are currently working to fully document allophony in the language, which will allow future iterations of this transform to yield more precise phonetic representations. Prior linguistic work on St. Lawrence Island Yupik by Krauss (1975) and Nagai (2001) used variants of Americanist phonetic notation; the tool also allows a user to transform Yupik words from the standard orthography into these phonetic notations.

The third set of transforms transliterates Yupik input from the St. Lawrence Island Yupik Latin orthography into the Chaplinski Yupik Cyrillic orthography, as well as from the Cyrillic orthography into the Latin orthography. As legacy Yupik texts are digitized and made available to the Yupik community, it will be desirable for Yupik texts written in the Latin orthography to be viewable by the Chaplinski Yupik community in Cyrillic and for Yupik texts written in the Cyrillic orthography to be viewable by the St. Lawrence Island Yupik community in the Latin orthography. As we consider, in consultation with the Yupik community, how to make legacy Yupik texts most easily available to Yupik community members, we anticipate that the Javascript implementation that underlies these transliteration utilities will be incorporated into any web-based solutions.

The fourth set of transforms has a target audience of both Yupik language learners and linguists documenting Yupik. The Yupik grammar of Jacobson (2001) presents stress-assignment rules documenting which syllables in a Yupik word receive stress. Jacobson also presents an annotation standard for the Latin orthography that marks syllable boundaries and stress assignment. If this transform is selected, the tool converts Yupik words input by the user into the syllable/stress marking notation of Jacobson. Inclusion of this notation is expected to be primarily useful for Yupik high school students on St. Lawrence Island who use the Jacobson grammar as a text. The web utility also extends Jacobson's notation to present syllable/stress marking

in Cyrillic. Finally, this set of transforms allows for syllable and stress to be marked in standard IPA notation for use by linguists.

Data sparsity and morphological analysis

In recent decades, substantial progress has been made in the field of computational linguistics, including such areas as machine translation, speech recognition, and parsing, with major achievements in statistical natural language processing techniques beginning in the 1990s (e.g., Brown et al. 1993) and explosive growth in neural network techniques in the past several years (Goodfellow, Bengio, and Courville 2016). Many, if not most, of the most successful techniques were developed for English, and are therefore based around the assumption that the word is the primary meaning-bearing unit within the language, often completely ignoring both derivational and inflectional morphology. For a resource-rich, largely analytic language such as English, this assumption typically works reasonably well. When a modern machine translation system is trained to translate from English to German, for example, there is enough training data available that the system can treat *cat* and *cats* (for example) as completely unrelated types with little degradation in performance.

Yupik, in contrast, is a polysynthetic language that relies very heavily on suffixation of both derivational and inflectional morphemes. In addition, Yupik is a low-resource language, meaning that there exist relatively few texts in the language and even fewer that are easily accessible in digitized textual form. The result of these two facts is that any attempt to build computational tools for Yupik using standard techniques from computational linguistics is bound to encounter major issues in data sparsity. To quantify the effect of polysynthesis on data sparsity, we can directly measure the word type sparsity of a corpus using a metric proposed by Hasegawa-Johnson and colleagues (2017). Starting at the beginning of corpus $\mathcal{C}$, every time a previously unseen word type u_n is encountered, record the number of previously seen word tokens encountered since the previous novel word type u_{n-1}. Let $(\mathcal{C})$ be a measure of data sparsity of corpus $\mathcal{C}$, defined as the average over these records over the corpus. In a hypothetical degenerate corpus $\mathcal{C}_\infty$ in which the entire corpus consists of exactly one word type over and over again, $(\mathcal{C}_\infty)$ approaches infinity. In a hypothetical degenerate corpus $\mathcal{C}_1$ where every word type is found exactly once in the corpus, $(\mathcal{C}_1) = 0$. When we measure the word type sparsity using d on the Yupik text $\mathcal{C}_{ess}$ and English translation $\mathcal{C}_{eng}$ of volume 1 of *Lore of St. Lawrence Island* (Apassingok et al., 1985), we observe that $d(\mathcal{C}_{eng}) = 5.27$ and $d(\mathcal{C}_{eng}) = 0.53$. In other words, on average in this text, a new English word type is encountered approximately every 5 to 6 words. In contrast, a new Yupik word type is encountered approximately every 0 to 1 word.

One technique for mitigating data sparsity, especially in synthetic languages, is the use of a morphological analyzer. A morphological analyzer is a computational tool that takes the surface form of a word and produces the corresponding sequence of underlying morphemes in a format very similar to an interlinear gloss. This approach has been successfully used, for example, as a foundation for various natural language processing technologies for numerous languages, including Finnish (for examples, see Beesley and Karttunen 2003).

We have built a finite-state morphological analyzer for Yupik (Chen and Schwartz 2018) using *foma* (Hulden 2009), an open-source toolkit for implementing finite-state morphological analysis tools. This analyzer includes the lexical roots, derivational suffixes, and inflectional suffixes, as well as the morphological, morphophonological, phonological, and orthotactic rules of Yupik as described by Jacobson (2001).[2] The analyzer also includes a preliminary integration of the remaining lexical roots and derivational suffixes of Yupik from the Badten and colleagues (2008) Yupik-English dictionary.

When we run the morphological analyzer over the Yupik texts that we have digitized to date, it fails to produce a morphological analysis for approximately 25 per cent of the Yupik tokens (corresponding to 51 per cent of the Yupik word types) in the texts. The analyzer failing to produce an analysis for a Yupik word indicates one of several possible issues. One possibility is that the word's root or one of its derivational suffixes is unattested in the Badten and colleagues (2008) dictionary and the Jacobson (2001) grammar. In that case, lexicographic fieldwork must be performed to validate the missing morpheme prior to adding it to the dictionary. Another possibility is that the word was misspelled in the original text or was mistranscribed during digitization. In that case, fieldwork is also indicated. A final possibility is that all of the word's morphemes are included in the analyzer, but a bug in the analyzer is causing analysis failure. A secondary issue involves words for which the analyzer produces more than one morphological analysis; these we address through a combination of additional elicitation and grammar engineering.

In the next section we discuss our ongoing development of further resources in support of Yupik language education and our own language documentation work.

---

2. As future work, we also plan to consult the Central Alaskan Yup'ik descriptions of Jacobson (1995) and Miyaoka (2012). These are more thorough in their descriptions of Yup'ik than Jacobson (2001) is for Yupik, and we anticipate that insights from these descriptions may be helpful in improving our Yupik analyzer.

An Integrated Approach in Support of Yupik Language Documentation and Education

The resources described in the previous section have been intentionally designed to closely integrate research processes from language documentation and computational linguistics. This has been done so that the results stemming from each approach might positively support the efforts of the other, and so that both disciplines might concretely support community-based efforts to revitalize and teach the language. Our work on this integrated approach involves several interlocking components, which we discuss below.

Digitization of print materials

Our goal in digitization is to enable access to existing Yupik-language materials that are not currently broadly accessible to linguists or Yupik community members. As a first step in this process, we have begun to systematically index existing Yupik materials that are not already in the Alaska Native Language Archive (ANLA) index. In doing so, we are creating a prioritized list of existing texts for digitization.

We have successfully piloted an end-to-end approach to rapidly digitizing and validating existing Yupik texts. In this process, we scan books and loose pages as 600 dpi TIFF files using a dedicated book scanner with sloped edge. Files undergo post-processing in ScanTailor, where page orientation correction, content selection, deskewing, dewarping, and despeckling operations are performed. Optical character recognition (OCR) is performed using Abbyy FineReader14, with the resulting digitized text exported in UTF-8 format. The majority of OCR errors are able to be identified via the non-lexicalized spellchecker. The resulting text is validated, and any remaining errors corrected, by hand. Using this process we have digitized and validated the entire 700 page *Lore of St. Lawrence Island* trilogy (Apassingok et al. 1985, 1987, 1989), as well as several of the primers.

While our goal is to scan, post-process, and archive in digital form as much material as possible, we are prioritizing materials that are considered most valuable and relevant to Yupik community members and researchers. Highest priority is being given to dual-use materials that linguists can search for examples of linguistic phenomena and that Yupik instructors can quickly incorporate into Yupik language instruction. In addition to the *Lore of St. Lawrence Island* trilogy, this will include Title VII materials such as the series of three elementary readers (Apassingok et al. 1993, 1994, 1995) and the Sulpik stories (Apassingok and Tennant 1987), a collection of short stories designed to promote literacy among elementary school children. We will also determine the priority for digitization of the dozens of Cyrillic-orthography Yupik publications developed in Russia (see Krauss 1971 for a partial list).

Language documentation and supporting technologies

These digitized materials form a growing searchable digital corpus of Yupik texts, which will allow broad access by community members, Yupik language teachers and learners, and other linguists engaged in documentation efforts. As educators work to build the Yupik curriculum and work towards an immersion program, they will be able to access these texts and pedagogical materials easily, which will facilitate the integration of folklore and traditional cultural knowledge into the course of study.

Ongoing fieldwork and ongoing computational research proceed in tandem, with the results of each process supporting the other. The top priority with regard to computational research is augmentation of the morphological analyzer through elicitation with the goal of complete coverage over existing Yupik texts. Access to the morphological analyzer enables rapid, often immediate, creation of interlinear-gloss style morphological analysis while on-site with our Yupik consultants. The completed morphological analyzer will be coupled with operating system spellcheck APIs to create a full-fledged lexically aware spellchecker. This spellchecker will be provided to the Yupik community, and will also speed up validation of scanned legacy materials. Using the morphological analyzer with our corpus of elicited field recordings, we plan to develop novel techniques for very low-resource speech recognition and spoken language identification. Such research will serve as a novel contribution to computational linguistics, and will serve the highly practical function of speeding up the process of segmenting and transcribing elicited field recordings. The corpus of digitized Yupik texts will also be crucial to such efforts, providing a corpus on which to train a probabilistic Yupik language model.

Using the online language learning infrastructure of Little (2017), we have begun building Yupik language-learning lessons, using an existing set of prototype Inuktitut lessons as a starting point. As part of our recent fieldwork, we worked with our primary consultant to adapt existing Inuktitut lessons to Yupik, and recorded two widely respected Yupik speakers (one male, one female) speaking the Yupik phrases in the lessons. We have observed generational, inter-family, and inter-speaker variation with regard to things such as willingness to be creative with word formation, use of certain words and morphemes, and pronunciation. As such, we are working to ensure that the Yupik in these lessons will reflect a consensus of existing fluent speakers as far as possible, and include usage notes when consensus is not possible.

Community access to legacy and new materials

Community members have iterated a desire for increased use of Yupik materials in the local school and for easier community access to existing

materials. To that end, it is our intention to integrate all digitized Yupik texts and resources into a searchable corpus that is accessible both online and offline, such that students, community members, and linguists can not only access the documents, but also perform any number of functions on them such as transliteration or undoubling the text into a more orthographically transparent form. In addition to their use by researchers as standalone utilities, we anticipate integration of the analyzer and the dictionary into the aforementioned corpus portal. In this way, our computational utilities and the digitized texts will together form a single repository of language learning and language documentation resources. We have developed an HTML and Javascript version of the Yupik-English (Badten et al. 2008) dictionary, which we plan to field-test with Yupik speakers in summer 2019.

In addition to making available legacy Yupik books and pedagogical materials, we have developed a new interactive mode for accessing these materials: interactive narrated e-books. After scanning and post-processing one of the Yupik pre-primers (Teeluk, Chikoyak, and Badten 1972), we recorded our primary consultant reading the text. We split the audio into segments corresponding to each page of the book, and created an interactive e-book in ePub format. The result is a platform-neutral electronic book that a reader can listen to while they read; the reader can replay each page as many times as desired to hear the words again. Platform neutrality is especially important in this case, as learners may want to access the books on their smartphones or on the Chromebook laptops that are available for student use during the school day. We are in the process of integrating our Javascript transliteration utility within the e-books to enable a user to transparently select between the standard Latin orthography, a fully transparent Latin variant orthography, the Cyrillic orthography, or IPA (with or without syllable boundaries and stress markings).

The issue of language learners' access to materials is closely tied to the issue of internet access. Most community members have smartphones (on various platforms), but not all have personal computers. However, while cellular service is relatively decent on the island, data access is generally poor. For electronic resources such as the dictionary and e-books to be of use to the most people, offline functionality will be important.

Conclusion

As a result of the geographical isolation of St. Lawrence Island coupled with the cultural strength of the Yupik community, St. Lawrence Island Yupik retained very high levels of natural parent-to-child language transmission into the 1980s, far later than was the case for most other Indigenous languages of North America. The language today is at a critical inflection point. A very high

percentage of St. Lawrence Island Yupiget at or above the age of forty are Yupik-dominant L1 Yupik speakers. A very high percentage of St. Lawrence Island Yupiget at or below the age of twenty are English-dominant L1 English speakers.

The Yupik community on St. Lawrence Island has iterated an explicit desire for the role of Yupik in the local educational curriculum to be strengthened. A sizable collection of Yupik-language cultural and pedagogical materials exists. It is our hope that as the digitization efforts of this project progress, these extant materials can be adapted by St. Lawrence Island Yupik educators for incorporation within modern pedagogical best practices.

The computational linguistic utilities that we are producing will also enable more widespread access to these materials. By utilizing our Javascript transliteration utility, we can trivially convert any Latin-orthography e-books that we create into the Yupik Cyrillic orthography for use by Yupiget in Russia. Once complete, it should be straightforward to extend our Yupik morphological analyzer into a full-fledged lexical spellchecker for distribution to and use by the Yupik-speaking community. With some degree of additional work, it should be possible to integrate the analyzer into a Yupik-language text completion module for mobile phones.

We believe that modern efforts in language documentation, computational linguistics, and language revitalization can and ideally should be integrated. The model we have described is one in which computational linguistic research and development proceeds in concert with documentary fieldwork, and where both are strongly guided by the goals and input of the language community. We hope that this model may prove useful, both for other languages in the Inuit-Yupik language family and for language communities more broadly.

Acknowledgements

References

APASSINGOK, Anders, and Dorothy WAGHIYI. 1985a. *Qula*. Gambell, AK: St. Lawrence Island Bilingual Education Center.
———. 1985b. *Talli*. Gambell, AK: St. Lawrence Island Bilingual Education Center.

———. 1985c. *Yupigem Antonym-ngi: Akuzitet Sameng Pesiitangi Kipunqulghiit*. Gambell, AK: St. Lawrence Island Bilingual Education Center.

———. 1985d. *Yupigem Synonym-ngi: Akuzitet Atillghi Tawatelgutkullghiit Sameng Pesiitangi Allakullghiit*. Gambell, AK: St. Lawrence Island Bilingual Education Center.

———. 1985e. *Yupigem Synonym-ngi: Akuzitet Sameng Pesiitangi Tawatekutevzilghiit*. Gambell, AK: St. Lawrence Island Bilingual Education Center.

APASSINGOK, Anders (Iyaaka), and Edward TENNANT (Tengutkalek), eds. 1987. *Sulpik*: *Sivuqaghmiitlu, Sivungaghmiitlu apeghtughistengita ulimaaghat* (The adventures of Sulpik). Unalakleet, AK: Bering Strait School District.

APASSINGOK, Anders (Iyaaka), Willis WALUNGA (Kepelgu), and Edward TENNANT (Tengutkalek), eds. 1985. *Sivuqam Nangaghnegha — Siivanllemta Ungipaqellghat / Lore of St. Lawrence Island — Echoes of Our Eskimo Elders*. Vol. 1, *Gambell*. Unalakleet, AK: Bering Strait School District.

———.1987. *Sivuqam Nangaghnegha — Siivanllemta Ungipaqellghat / Lore of St. Lawrence Island — Echoes of Our Eskimo Elders*. Vol. 2, *Savoonga*. Unalakleet, AK: Bering Strait School District.

———. 1989. *Sivuqam Nangaghnegha — Siivanllemta Ungipaqellghat / Lore of St. Lawrence Island — Echoes of Our Eskimo Elders*. Vol. 3, *Southwest Cape*. Unalakleet, AK: Bering Strait School District.

APASSINGOK, Anders (Iyaaka), Jessie UGLOWOOK (Ayuqliq), Lorena KOONOOKA (Inyiyngaawen), and Edward TENNANT (Tengutkalek), eds. 1993. *Kallagneghet / Drumbeats*. Unalakleet, AK: Bering Strait School District.

———. 1994. *Akiingqwaghneghet / Echoes*. Unalakleet, AK: Bering Strait School District.

———. 1995. *Suluwet / Whisperings*. Unalakleet, AK: Bering Strait School District.

APASSINGOK, Thomas, Imingan SMITH, Hazel OMWARI, and Jimmie TOOLIE. 1972. *Ayumiim Ungipaghaatangi I (Stories of Long Ago I)*. Fairbanks: Alaska Native Language Center. https://www.uaf.edu/anla/collections/search/resultDetail.xml?id=SY972BKK1972.

APATIKI, Edna, Lydia APATIKI, and Charlene APANGALOOK. 1982. *Whangaperegaaghmeng (All About Me)*. Fairbanks: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY982AAA1982.

BADTEN, Adelinda Womkon, ed. 1974. *Atightuusim Aallghi (Second Reading Book)*. Fairbanks: Alaska Native Language Center. https://www.uaf.edu/anla/collections/search/resultDetail.xml?id=SY970B1974b.

BADTEN, Linda Womkon (Aghnaghaghpik), Vera Oovi KANESHIRO (Uqiitlek), Marie OOVI (Uvegtu), and Christopher KOONOOKA (Petuwaq). 2008. *St. Lawrence Island / Siberian Yupik Eskimo Dictionary*. Fairbanks: University of Alaska Fairbanks, Alaska Native Language Center.

BEESLEY, Kenneth R., and Lauri KARTTUNEN. 2003. *Finite-State Morphology*. Chicago: University of Chicago Press for the Center for the Study of Language and Information.

BLANCHETT, Marie N., Martha TEELUK, Paschal AFCAN, Andrew CHIKOYAK, Adelinda BADTEN, and Vera Oovi KANESHIRO. 1972a. *Kulusiq*. Fairbanks: University of Alaska Fairbanks, Alaska Native Language Center. https://www.uaf.edu/anla/collections/search/resultDetail.xml?id=SY970B1972c.

BLANCHETT, Marie N., Martha TEELUK, Paschal AFCAN, and Ora GOLOGERGEN. 1972b. *Qepghaghaqukut Naghaaghaqukut (We Work and Play)*. Fairbanks: University of Alaska Fairbanks, Alaska Native Language Center.

BROWN, Peter E., Stephen A. Della PIETRA, Vincent J. Della PIETRA, and Robert L. MERCER. 1993. "The Mathematics of Statistical Machine Translation: Parameter Estimation." *Computational Linguistics* 19 (2): 263–311.

CHEN, Emily, and Lane 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 18)*, edited by Nicoletta Calzolari et al., 2623–30. Miyazaki, Japan: European Language Resources Association.

CHRISPIN, Barbara, Sharon ORR, Christine ALOWA, and Jeffrey APATIKI. 1975a. *Workbook for Kulusiinkut*. Fairbanks: University of Alaska Fairbanks, Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY971SAC1975b.

———. 1975b. *Workbook for Kulusiq*. Fairbanks: University of Alaska Fairbanks, Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY971SAC1975d.

CRISPIN, Douglas M., Sharon ORR, Dorothy WAGHIYI, and Jeffrey APATIKI. 1976. *Sivuliit Yupigestun Spelling-ngesit Iganka (My First Yupik Spelling Book)*. Nome, AK: Nome Agency Bilingual Education Resource Center.

DE REUSE, Willem J. 1994. *Siberian Yupik Eskimo: The Language and Its Contacts with Chukchi*. Salt Lake City: University of Utah Press.

GOODFELLOW, Ian, Yoshua BENGIO, and Aaron COURVILLE. 2016. *Deep Learning*. MIT Press. http://www.deeplearningbook.org.

HARGRAVES, Savannah, Raymond OOZEVASEUK, and Michael APATIKI. 1976. *Kaleret (Colors)*. Nome, AK: Nome Agency Bilingual Education Resource Center.

HASEGAWA-JOHNSON, Mark, Mohamed ELMAHDY, and Eiman MUSTAFAWI. 2017. "Arabic Speech and Language Technology." In *Routledge Handbook of Arabic Linguistics*, edited by Elabbas Benmamoun and Reem Bassiouney, 299–311. Oxford: Taylor and Francis Group Ltd.

HULDEN, Mans. 2009. "Foma: A Finite-State Compiler and Library." In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session*, 29–32. Stroudsburg, PA: Association for Computational Linguistics.

HUNT, Benjamin, Emily CHEN, Sylvia L.R. SCHREINER, and Lane SCHWARTZ. Forthcoming. "Community Lexical Access for an Endangered Polysynthetic Language: An Electronic Dictionary for St. Lawrence Island Yupik." *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*.

JACOBSON, Steven A. 1977. *A Grammatical Sketch of Siberian Yupik Eskimo as Spoken on St. Lawrence Island, Alaska*. Fairbanks: Alaska Native Language Center.

———.1985. "Siberian and Central Yupik Prosody." In *Yupik Eskimo Prosodic Systems: Descriptive and Comparative Studies*, edited by Michael E. Krauss, 25–46. Fairbanks: Alaska Native Language Center.

———. 1990. *Reading and Writing the Cyrillic System for Siberian Yupik*. Fairbanks: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY975J1990a.

———. 1995. *A Practical Grammar of the Central Alaskan Yup'ik Eskimo Language*. Fairbanks: Alaska Native Language Center.

———. 2001. *A Practical Grammar of the St. Lawrence Island/Siberian Yupik Eskimo Language*, 2nd ed. Fairbanks: Alaska Native Language Center.

KANESHIRO, Vera Oovi, and Jeffrey APATIKI. 1975. *Teketaatenkuk Kinunkuk (Teketaat and Kinu)*. Fairbanks, AK: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY972K1975c.

KANESHIRO, Vera Oovi, and Clyde T. KANESHIRO. 1974. *Unkusequlghiik (Going to See the Fox Traps)*. Fairbanks: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY972K1974e.

———. 1975. *Piyaataalghiit (Going for a Walk)*. Fairbanks: Alaska Native Language Center.

KANESHIRO, Vera Oovi, and George SMART. 1980. *Atightumun Liitusit III: Kumyugnalghiit Latat (Primer Book Three: Undoubling Letters)*. Fairbanks: Alaska Native Language Center.

KANESHIRO, Vera Oovi (Uqiitlek), and Adelinda Womkon BADTEN, eds. 1975. *Pangeghtellghet / Visits to Siberia*. Fairbanks: Alaska Native Language Center.

KOONOOKA, Christopher (Petuwaq). 2003. *Ungipaghaghlanga — Quutmiit Yupigita Ungipaghaatangit / Let Me Tell a Story – Legends of the Siberian Eskimos*. Fairbanks: Alaska Native Language Center.

———. 2005. "Yupik Language Instruction in Gambell (St. Lawrence Island, Alaska)." *Études Inuit Studies* 29 (1–2): 251–66.

KRAUSS, Michael E. 1971. "Developing a Literature in the Language of the Eskimos of St. Lawrence Island." Unpublished manuscript. Alaska Native Language Center Identifier SY970K1971c. https://uafanlc.alaska.edu/Online/SY970K1971c/SY970K1971c.pdf.

———. 1973. *St. Lawrence Island and Siberian Eskimo Literature*. http://www.uaf.edu/anla/item.xml?id=SY970K1973a.

———. 1975. "St. Lawrence Island Eskimo Phonology and Orthography." *Linguistics: An International Review* 13 (152): 39–72.

———. 1980. "Alaska Native Languages: Past, Present and Future." Alaska Native Language Center Research Papers no. 4.

KRAUSS, Michael E., Gary HOLTON, Jim KERR, and Colin T. WEST. 2011. *Indigenous Peoples and Languages of Alaska* [Map]. Fairbanks and Anchorage: Alaska Native Language Center and UAA Institute of Social and Economic Research. Alaska Native Language Archive Identifier G961K2010. http://www.uaf.edu/anla/collections/search/resultDetail.xml?id=G961K2010.

KRAUSS, Michael E., Jeff LEER, Steven A. JACOBSON, and Lawrence KAPLAN, eds. 1985. *Yupik Eskimo Prosodic Systems: Descriptive and Comparative Studies*. Fairbanks: University of Alaska Press.

KRUPNIK, Igor, and Michael CHLENOV. 2013. *Yupik Transitions: Change and Survival at Bering Strait, 1900–1960*. Fairbanks: University of Alaska Press.

LITTLE, Alexa N. 2017. "Connecting Documentation and Revitalization: A New Approach to Language Apps." In *Proceedings of the 2nd Workshop on the Use of Computational Methods in the Study of Endangered Languages*, edited by Antti Arppe, Jeff Good, Mans Hulden, Jordan Lachler, Alexis Palmer, and Lane Schwartz,

151–55. Honolulu, HI: Association for Computational Linguistics. https://www. aclweb.org/anthology/W17-01.

MENOVSHCHIKOV, G.A. 1960. *Eskimosskii iazyk*. Leningrad: Gosudarstvennoe uchebno-pedagogicheskoe izdatel'stvo.

———. 1962. *Grammatika iazyka aziatskikh eskimosov*, vol. 1. Moscow and Leningrad: Izdatel'stvo akademii Nauk.

———. 1967. *Grammatika iazyka aziatskikh eskimosov*, vol. 2. Moscow and Leningrad: Izdatel'stvo akademii Nauk.

MENOVSHCHIKOV, G.A., and Nicolai B. VAKHTIN. 1990. *Eskimosskii iazyk.* Leningrad: Prosveshchenie. Preliminary edition 1983.

MIYAOKA, Osahito. 2012. *A Grammar of Central Alaskan Yupik (CAY).* Boston: De Gruyter Mouton.

MORGOUNOVA, Daria. 2007. "Language, Identities and Ideologies of the Past and Present Chukotka." *Études Inuit Studies* 31 (1–2): 183–200.

NAGAI, Kayo. 2001. *Mrs. Della Waghiyi's St. Lawrence Island Yupik Texts with Grammatical Analysis*. Kyoto, Japan: Nakanishi Printing.

NOME CITY SCHOOLS. 1983. *Yupigestun Igallghet*. Fairbanks: Alaska Native Language Center.

OOSEVASEUK, Edith, and Dorothy Waghiyi. 1985. *Ateghyiighaghhaankuk Meteghllugenkuk (The Little Bird and the Raven)*. Gambell, AK: St. Lawrence Island Bilingual Education Center.

ORR, Sharon (Kesliq), and Luann Wetmore. 1975. *Kiluuq*. Fairbanks: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY971S1975.

POAGE, Myra. 1975. *Naten Teghikusat Qavaghaqat? (How Do Animals Sleep?)*. Nome, AK: Nome Agency Bilingual Education Resource Center.

POAGE, Myra, and Jeffrey Apatiki. 1975. *Atightughtek Kalereqitek Naghaaghtek (Read Color Play)*. Nome, AK: Nome Agency Bilingual Education Resource Center.

POAGE, Myra, Raymond Oozevaseuk, Linda S. Gologergen, and Jeffrey Apatiki. 1983. *Latam Liitellghi (Letter Recognition)*. Rev. ed. Nome, AK: Nome Agency Bilingual Education Resource Center.

PUBLIC LAW 90-247. 1967. *Title VII of the Elementary and Secondary Education Amendments of 1967: Bilingual Education Act*. https://www.govinfo.gov/content/ pkg/STATUTE-81/pdf/STATUTE-81-Pg783.pdf.

PUNGOWIYI, Laura, and Edward GOLOGERGEN. 1975. *Ayumiim Ungipaghaatangi III (Stories of Long Ago III)*. Fairbanks, Alaska, Alaska Native Language Center.

RUBTSOVA, E.C. 1971. *Eskimossko-russkii slovar'*. Moscow: Izdatel'stvo "Sovetskaia entsiklopediia."

SCHWALBE, Daria Morgounova. 2017. "Sustaining Linguistic Continuity in the Beringia: Examining Language Shift and Comparing Ideas of Sustainability in Two Arctic Communities." *Anthropologica* 59 (1): 28–43. https://muse.jhu.edu/article/658679.

SCHWARTZ, Lane, and Emily CHEN. 2017. "Liinnaqumalghiit: A Web-Based Tool for Addressing Orthographic Transparency in St. Lawrence Island/Central Siberian Yupik." *Language Documentation and Conservation* 11 (September): 275–88. https://doi.org/10125/24736.

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." *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.

SEVIL'GAEV, G.F. 1985. "Razvitie narodnogo obrazovaniia: Sozdanie pis'mennostei na iazykakh narodnostei Severa" [Progress in education: Creating literacies for northern indigenous minorities]. In *Narody Dal'nego Vostoka SSSR v 17-20 vv.*, edited by I.S. Gurvich, 177–187. Moscow: Nauka.

SHINEN, David C. 1976. *Siberian Yupik Literacy Manual*. Nome: Nome Agency Bilingual Education Resource Center.

SLWOOKO, Grace. 1977. *Qateperewaaghmeng Aatkaqelghii Yuuk (The Man Dressed in White)*. Fairbanks: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY975Sw1977a.

SLWOOKO, Grace, and Rose KULUKHON. 1975. *Ayumiim Ungipaghaatangi II (Stories of Long Ago II)*. Fairbanks: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY972K1975d.

SLWOOKO, Grace, Flora IMERGAN, Homer APATIKI, Dorothy WAGHIYI, and Ruby ROOKOK. 1975. *Ayumiim Ungipaghaatangi IV (Stories of Long Ago IV)*. Fairbanks: Alaska Native Language Center. https://www.uaf.edu/anla/collections/search/resultDetail.xml?id=SY972K1975.

SLWOOKO, Grace, Ruby ROOKOOK, Beda SLWOOKO, Abraham KANINGOK, and Fred OKOMIALNGUK. 1975. *Ayumiim Ungipaghaatangi V (Stories of Long Ago V)*. Fairbanks: Alaska Native Language Center. https://www.uaf.edu/anla/collections/search/resultDetail.xml?id=SY972K1975g.

TEELUK, Martha, Andrew CHIKOYAK, and Adelinda BADTEN. 1972. *Kulusiinkut (Kulusiq and His Family)*. Fairbanks: Alaska Native Language Center. http://www.uaf.edu/anla/item.xml?id=SY970B1972d.

TENNANT, Edward, ed. 1985a. *Yupik Formula Three Reading-Spelling-Learning Program: Instructor's Manual*. Unalakleet, AK: Bering Strait School District.

———. 1985b. *Yupik Formula Three Reading-Spelling-Learning Program: Study Guide*. Unalakleet, AK: Bering Strait School District.

VAKHTIN, Nikolai B. 2001. *Yazyki narodov Severa v XX veke: Ocherki yazykovogo sdviga* [Languages of the peoples of the north in the 20th century: Essays on language shift]. St. Petersburg: European University at St. Petersburg.