



## The distribution of nominal quantifiers in a digitized corpus of St. Lawrence Island Yupik



Benjamin Hunt  
bhunt6@gmu.edu

Sylvia L.R. Schreiner  
sschre12@gmu.edu



Linguistics Program - Department of English - George Mason University

### Overview

- St. Lawrence Island Yupik. St. Lawrence Island / Central Siberian Yupik / *Akuzipik* (ISO 639-3:ess)
- Endangered, under-resourced language spoken on St. Lawrence Island, AK; Chukotka peninsula of Russia; mainland AK (~1000 speakers) (Schwartz et al. 2021)
- Rapid generational shift since 1980s; youth now largely L1 English speakers (Schwartz et al. 2021)
- Polysynthetic, ergative-absolutive (case), relatively free word order
- Part of a larger project to document Yupik, digitize legacy materials, create computer tools for researchers and speakers, aid in revitalization efforts
- Existing work on agreement, noun incorporation (de Reuse 1994, Jacobson 2001)
- Goals:
  - Arrive at an exhaustive list of nominal quantifiers
  - Is there a preferred order between head N and Q?
  - Are case and number agreement obligatory in N-Q?

### Methods

#### Initial search

- Badten, et al. (2008) Yupik-English dictionary, Jacobson (2001) grammar searched for Q-like elements
- Found possible examples of Generalized Existential, Generalized Universal, Proportional, Comparative, and Partitive Qs (Keenan 2017)
- Examples:
  - GE: *ataasiq* 'one', *maaghuq* 'two'
  - GU: *ingunagh* 'all; whole', *amaagghagh* 'all, every, both [for dual]'
  - Proportional: *ugligh* 'large number or amount; many; much; large quantity', *aveg* 'half' (nominal root)
  - Comparative: *elimaagghagh* 'a little bit more'
  - Partitive: *naali/naighil(a)* 'any of, which'

#### Elicitation

- To gather naturally produced examples of known Qs, potentially elicit others
- Procedure adapted from the Quantifier Questionnaire (Keenan 2017)
- Series of visualizations created with each of the categories in mind (following Kinloch 1971, Gregg 1992, and Woods 1999, Mathewson 2004)
- Elicitation conducted asynchronously via Facebook Messenger due to COVID-19: speaker asked to comment on each visualization
- Known Qs confirmed; several new senses of known roots added to list of Qs



Figure 1: Quantifier visualization 9 – reservations  
*pis* "most", "almost all", "many"  
*ng* "two", "few/a few", "not many"

Kyang *laagaghthengut*.  
"Most are singers."  
(Elicitation 09/2020; Speaker 21)

*Maaghuq ukughvagneg* *laagan* *ghituk*.  
"Two out of all these didn't sing."  
(Elicitation 09/2020; Speaker 21)

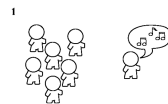


Figure 2: Quantifier visualization 1 –  
*pis* "one", "only one"  
*ng* "most", "almost all", "many", "several"

Team *laani* *laagastagil*.  
"That one is singing to the others he's with."  
(Elicitation 09/2020; Speaker 21)

*Ataasighthinaq* *laagaghaaquq*.  
"Only one is singing."  
(Elicitation 09/2020; Speaker 21)

### Corpus Preparation and Search

- Yupik digital corpus (Schwartz, et al. 2021)
- Digitized text files
  - primers, storybooks, and fables (all primarily fiction)
  - Yupik NT Translation and other religious texts were excluded
- ~100,000 tokens, ~40,000 types
- Corpus annotated using finite-state morphological analyzer (Chen & Schwartz 2018)
- Series of regular expressions used to extract targeted quantifier contexts, yielding 17,937 tokens
- Included: Target word forms with unambiguous analyses (4,953 tokens) or two analyses that posited the same root and agreed in part of speech
  - Also any quantifiers that would've been totally excluded by these criteria
- Subcorpus: 1076 sentences, 581 coded, 172 identified as nominal quantification
- Coded manually for
  - Quantifier
  - Pre- or postnominal
  - Case and number of Q & N

Quantifier	Meaning	Identified Root	Root definition (Badten et al. 2008)
<i>laani</i>	other	<i>la</i>	relative; associate; part; one or some (of)
<i>laanghini</i>	only some		
<i>aveghil/aveghale</i>	half	<i>aveg</i>	half
<i>ugligh</i>	more	<i>ugligh</i>	large number or amount; many; much; large quantity
<i>kyang</i>	most	<i>kyang</i>	mostly; usually
<i>qerngita</i>	all	<i>qerngagh</i>	together
<i>ingunagh</i>	all	<i>ingunagh</i>	all; whole
<i>natangighaq</i>	little bit	<i>natangighagh</i>	little bit; slightly; temporarily
<i>naighghitini</i>	only part	<i>naighgh</i>	some part or the other
<i>naighghok</i>	another		
<i>naighghit</i>	anyone of these		
<i>naighghini</i> / <i>naighghinay</i>	some	<i>naigh</i>	which one(s)
<i>naighghinaq</i> / <i>naighghinaam</i>	only some little bit of these/it		

Figure 3: Quantifiers identified through elicitation task (novel and new senses of known quantifiers)

### Word Order

- Previous work posited relatively free word order in Yupik; determined primarily by discourse considerations (Jacobson 2001; de Reuse 1994)
- De Reuse (1994): possessors tend to precede the noun; demonstratives and nominal modifiers tend to follow the noun

- The corpus sample demonstrated the opposite tendency: quantifiers precede the noun more often, even when non-numeral quantifiers and numerals are considered separately
- Quantifier-noun (head-final) sequences make up ~68% of all sampled word orders, including those with another intervening word.



Figure 4: Word order frequency in corpus sample

- Influence from English seems unlikely, given the nature of the texts in the corpus sample (primers, storybooks, fables; published in the 1970s-1990s)
- Preference for one order over the other may be linked to discourse considerations (which may or may not be extractable from the corpus data)
- Next steps: Analysis based on the type of text; analysis of naturally-occurring speech in elders vs. younger speakers

### Case Agreement

- Head nouns and quantifiers were predicted to show agreement in case and number
  - Other nominal modifiers in the language (e.g., demonstratives) agree with the modified noun in both dimensions
- 83% of the quantifiers in the sample matched the modified noun in grammatical case
- Of those that did not match:
  - instances of a particular set of roots that can only take ergative case
  - caseless loanwords
  - undocumented morphology or misspellings (requires further investigation)
  - noun incorporation (no overt case marking on incorporated nouns)



Figure 5: Case agreement between nouns and quantifiers in the sample

### Case Agreement

- The sample showed a high incidence (95%) of number agreement between quantifier and head noun.
- The few instances of non-agreement:
  - morphologically singular nouns in a case that can be used to express plural semantics or
  - involve a particular quantifier (*lang* 'some') being used substantively along with another quantifier



Figure 6: Number agreement between nouns and quantifiers in the sample

### Implications and Future Directions

- Purposes:
  - Provide a descriptive picture of nominal quantification in Yupik
  - Contribute to the documentation of the language and further scholarship
- Several gaps identified in the expected quantifiers filled as a result of elicitation sessions
- The literature usually describes Yupik as having "free word order"
  - This analysis found that quantified nominal phrases appear to tend toward a head-final ordering
  - This could lend credibility to arguments for contextually preferred word order in languages which have traditionally been analyzed as having free word order
  - Possible implications for similar tendencies in related languages, particularly others on the Inuit-Yupik branch
- Additionally, our corpus annotation efforts lay the groundwork for the development of pedagogical materials for the Yupik community (e.g. a fully tagged, parsed corpus) to be used in...
  - advanced Yupik language study
  - development of natural language exercises for all learning levels
  - larger data-driven research projects on various aspects of Yupik grammar

#### Acknowledgments

The authors are presenting this poster from unceded land traditionally held by the Powhatan Algonquian, Manahoac, and Monacan peoples, the descendants of whom live in Northern Virginia and the surrounding areas today, including those on the Pamunkey and Mattaponi reservations.

Special thanks to the Yupik speakers who have shared their language and culture with us. Special thanks to Ukaal Crystal Aningayou for her help with the data in this presentation. Igamsiyagvikemsi. We take sole responsibility for any errors. Part of a collaborative research project with Lore Schwartz (UIUC) and RAs Giulia Masella Scatini, Preston Haas, Emma Evans, David Ngo, Kaelyn Map Kenny (GMU), and Emily Chen, Hayley Park, Peter Zuckerman (UIUC). Portions of this work were funded by an NSF Documenting Endangered Languages Grant #BCS 1705977, a Mattay 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. 2016. A morphological analyzer for St. Lawrence Island / Central Siberian Yupik. In Proceedings of (LREC'18), pp. 2623-2630. Miyazaki, Japan. de Reuse, W. 1994. Siberian Yupik Eskimo. *Studies in Indigenous languages of the Americas*. Salt Lake City: U of Utah P. Jacobson, S. 2001. A practical grammar of the St. Lawrence Island/Siberian Yupik Eskimo language. 2nd edn. Fairbanks, Alaska: ANLC. Keenan, E. 2017. The quantifier questionnaire. In *Handbook of quantifiers in natural language: Volume II*, 1-20. Dordrecht: Springer. Schwartz, L., E. Chen, H.H. Park, E. Jahn, & S.L.R. Schreiner. 2021. A digital corpus of St. Lawrence Island Yupik. *Proceedings of Comput-EL 4*.