

The Distribution of Nominal Quantifiers in a Digitized Corpus of St. Lawrence Island Yupik

Benjamin Hunt and Sylvia L.R. Schreiner

1. Introduction

In this paper we present a study of the distributional characteristics of lexical nominal quantifiers in a digitized corpus of St. Lawrence Island Yupik. St. Lawrence Island / Siberian Yupik / Akuzipik (ISO 639-3: *ess*, ‘Yupik’) is an endangered, under-resourced language spoken in the Bering Strait region. Approximately 800-900 L1 speakers of Yupik reside on St. Lawrence Island, Alaska, on the Chukotka peninsula of Russia, and in mainland Alaska (Schwartz, et al. 2019).

While use of Yupik is fairly robust in older generations, the language on St. Lawrence Island is undergoing a rapid generational shift that began in the 1990s, and youth today are predominantly English L1 speakers (Schwartz, et al. 2019). The work described here is part of a project to further document Yupik, digitize legacy materials, create computer tools for researchers and speakers, and aid in revitalization efforts.

Yupik is ergative-absolutive (in its case system) and displays relatively free word order. It is generally considered to be polysynthetic for a variety of reasons including its highly productive, concatenative derivational morphology and frequent noun incorporation (de Reuse 2009). Yupik verbs inflect for twelve moods as well as person and number for core arguments. Nouns inflect for seven cases (relative (~ergative), absolutive, ablative-modalis, localis, terminalis, vialis, and equalis), four persons (first, second, third, and third reflexive), and three numbers (singular, dual, and plural) (Jacobson 2001).

1.1. Motivation and research questions

Several authors have explored nominal behaviors in Yupik such as agreement and noun incorporation (de Reuse 1994; Jacobson 2001), but nominal quantification in Yupik has received little attention. Jacobson (2001) provides some examples of what appear to be verbal quantifier roots that he claims are only used in conjunction with the “quantifier-qualifier” construction. It is unclear from Jacobson’s treatment of these roots whether they are strictly verbal quantifiers or if they can be used in conjunction with nominal roots as well.

The ordering of elements in nominals is addressed briefly in de Reuse (1994). De Reuse highlights some of the common word orders that occur in Yupik clauses and notes that nominal phrases tend toward head-final order when a possessor is present and head-initial otherwise (pp. 25-26). This would suggest that we should expect quantifiers to follow the noun they modify. As will be demonstrated in section 3, the sample shows the opposite tendency for quantifier-noun pairs.

* Benjamin Hunt, George Mason University, bhunt6@gmu.edu. Sylvia L.R. Schreiner, George Mason University, sschrei2@gmu.edu. The authors’ institution lies on 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 Ukaall Crystal Aningayou for her help with the data in this paper. Igamsiqayugvikamsi. We take sole responsibility for any errors. Part of a collaborative research project with Lane Schwartz (UIUC) and RAs Giulia Masella Soldati, Preston Haas, Emma Evans, David Ngo, Kaelynn Mae Kenny (GMU); and Emily Chen, Hayley Park, Peter Zuckerman (UIUC). Portions of this work were funded by an NSF Documenting Endangered Languages Grant #BCS 1760977, a Mathy Junior Faculty Award in the Arts and Humanities, 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.

Apart from these works, the only treatment of any aspect of Yupik nominals is a preliminary analysis of DP structure based on evidence from the use of demonstratives as agreeing nominal modifiers (Hunt & Schreiner 2019). While this work is not conclusive, it does open the door to positing some sort of nominal constituency beyond the noun itself, which is relevant to any treatment of nominal quantification. This work also confirms the obligatory agreement that manifests between nominals and demonstratives, which may indicate that this is a feature of the nominal phrase as a whole. In light of this, we might expect that this same agreement (case and number) should be present between nouns and quantifiers as well. This obligatory agreement is shown to be present in the sample, despite a number of mismatches that can be explained by other syntactic or idiosyncratic means.

Compared to work on nominals and quantification in other languages, there is very little descriptive groundwork laid for such an investigation in Yupik. For example, Bittner's (1995) work on West Greenlandic quantification and the challenge it poses for compositional semantics begins with a description of the classes of quantifiers that exist in the language and provides a (presumably) exhaustive list of each type. Likewise, Nishiguchi's (2007) critique of Generalized Quantifier theory as it applies to Japanese quantification is prefaced by a description of Japanese predicative adjectival quantifiers before the discussion of how GQ theory may not be as useful for Japanese as it is for English. This preliminary focus on empirical description is no less true of each entry in Paperno & Keenan's (2017) *Handbook of Quantifiers in Natural Language: Volume II*, a collection of studies each characterizing quantificational elements and their distributions in 18 natural languages. The descriptive groundwork that serves as a necessary preface for theoretical investigations is lacking for the Yupik nominal domain; moreover, this descriptive groundwork cannot be undertaken if the targets of description (namely, Yupik nominal quantifiers) are entirely unknown.

This paper seeks first to identify and provide a descriptive picture of nominal quantifiers in Yupik both through elicitation and analysis of a written corpus. In order to achieve this goal, steps were taken to construct an exhaustive list of nominal quantifiers in Yupik, identify potentially interesting phenomena in the behavior and distribution of quantifiers—especially those patterns of behavior otherwise unidentifiable through individual data points—and ultimately lay the groundwork for theoretical study on nominal quantification. The secondary outcome of this study is the continued documentation of an endangered, understudied language to further scholarship on Yupik and typologically similar languages. This documentation work will aid in the production of computational and pedagogical resources to the benefit of the larger documentary effort on the language. To these ends, this paper examines the following research questions: 1. Is there a preferential order between head noun and quantifier? 2. Are case and number agreement obligatory between head noun and quantifier?

The paper proceeds as follows. In Section 2 we detail the methods of our search for quantificational elements and our corpus search. Section 3 presents the results of our corpus search. Section 4 discusses the results. Section 5 concludes with a discussion of the contributions of this project and the implications this work presents for future work on Yupik nominals and quantification in general.

2. Methods

2.1. Initial search

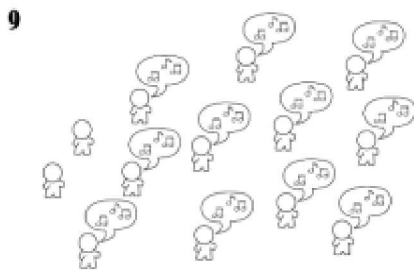
In order to establish an initial list of quantifiers to search for in the corpus, we undertook a preliminary search in the electronic version of Badten et al. (2008), a two-volume Yupik-English dictionary, as well as Jacobson's (2001) grammar, for lexical items whose English translations indicated that they might be quantificational. In this study we focus solely on independent lexical quantifiers. Paperno & Keenan (2017) provide a useful characterization of the types of nominal quantification found cross-linguistically. In their model, determiner-like quantifiers are broken down into three main categories: Generalized Existential Quantifiers (GEQs), Generalized Universal Quantifiers (GUQs), and Proportional Quantifiers (PQs). GEQs can be further subdivided into a) cardinals such as *a*, *some*, and numerals, b) interrogatives like *how many?* and *which?*, and c) value judgements like *many*, *few*, and *enough*. GUQs include quantifiers like *all*, *every*, and *each*, while PQs include *most*, *half*, and some uses of *many/few*. The cross-linguistic nature of Paperno & Keenan's characterization as well as the diversity of languages contained within their volume indicate that we should expect a similar array of quantifiers in Yupik. At the very least, we would anticipate finding GEQs, GUQs, and PQs.

In our search, possible examples of Keenan’s (2017) Generalized Existential, Generalized Universal, Proportional, Comparative, and Partitive types were found. We also identified quantifiers that are potentially “missing” when compared to English and Paperno & Keenan’s list. After completing the elicitation procedure, it appears that Yupik does express some of these missing nominal quantifiers distinctly and others through verbal quantification or the use of non-structural case marking. The results of our search are as follows: **Generalized Existential:** Cardinal: *ataasiq* ‘one’, *maalghuk* ‘two’, etc...; *ilangi* ‘some’. Interrogative: *qafsiniite* ‘which (in a series)’; *naligh** ‘which one(s)’; *qafsina* ‘how many?’. Value judgement: *qafsin’get* ‘several’. Potentially missing: ‘no/none’, ‘few’, group numerals (‘couple’, ‘dozen’, etc.). **Generalized Universal:** *iingunagh** ‘all; whole’, *tamaghhagh** ‘all, every, both (for dual)’, *qamagglu* ‘all, all of, each’. Potentially missing: ‘each’, ‘any’, ‘ever’. **Proportional:** *uglagh* ‘large number or amount; many; much; large quantity’, *aveg* ‘half’ (nominal root), *qafsinagnegh/qafsina* ‘several; a few’. Potentially missing: *most*. **Comparative:** *alla* ‘more; some more; again’, *ellmaaghaghghagh** ‘a little bit more’. Potentially missing: *fewer/less*. **Partitive:** *naliit/naliighiit(a)* ‘any of’. Potentially missing: ‘some of, few/several of’.

2.2. Elicitation

A series of elicitation sessions were conducted with one native speaker (female, age ~40, bilingual, English a.o.a. ~7 y.o.) to gather semi-naturalistically produced examples of known quantifiers, and to potentially elicit others not identified via the dictionary search. Following the example of successful pictorial elicitation studies like Kinloch (1971), Gregg (1992), and Woods (1999), a series of visualizations was created with the semantic categories given in Table 2 in mind. These visualizations were presented to the native speaker to prompt natural productions of appropriate quantifiers. The decision to conduct these elicitations with the use of images was based on work such as Matthewson (2004), which demonstrates the pitfalls of translation-based elicitation when targeting semantic phenomena. Each visualization was developed with a specific English quantifier or group of quantifiers in mind, but with an effort to leave the image abstract enough to allow multiple interpretations. For example, visualization 9, pictured below in Figure 1, could be interpreted in any of the following ways, and likely others: “Most of the/many/almost all of the/several people are singing”, “Two/few/a few of the/not many/a couple of the people are not singing.”

Figure 1: Example of elicitation prompt



Due to widespread travel restrictions in response to the COVID-19 pandemic, visualizations were sent individually to the speaker via Facebook Messenger over a two-week period. The participant was sent one image at a time and was asked to comment on what was pictured, without any context given. Once the initial response was received, a number of questions were asked of the participant to clarify the usage of any part of the response, and, if nothing identifiable as the targeted quantifier or quantifiers was provided, additional context was given to direct the speaker towards the desired semantic interpretation. These sessions resulted in the confirmation of the previously known quantifiers and the identification of a number of new quantificational senses of known roots. The roots deduced from these quantifier surface forms were added, in truncated form, to the list of regular expressions used in the extraction procedure outlined in the following section. The following quantifiers were identified; root definitions taken from Badten, et al. (2008): *ilaani* ‘other’, *ilaanghinii* ‘only some’ (root: *ila-* ‘relative; associate; part; one or some (of)’); *avelghhi/avegluku* ‘half’ (*aveg-* ‘half’); *uglan’ghi* ‘more’ (*uglagh-* ‘large number or amount; many; much; large quantity’); *kiyang* ‘most’ (*kiyang* ‘mostly; usually’);

qerngita ‘all’ (*qergagh-* ‘together’); *ingunaghmeng* ‘all’ (*iingunagh-* ‘all; whole’); *natengiighhaq* ‘little bit’ (*natengiighhagh-* ‘little bit; slightly; temporarily’); *nalighneghiini* ‘only part’ (*nalighnegh-* ‘some part or the other’); *nalighhqe* ‘neither’, *naliighhiit* ‘any/none of these’, *nalighhiiniit/naliighhiinaq* ‘none’, *naliighhiinaat/ naliighhiinaan* ‘only some/little bit of them/it’ (*naligh-* ‘which one(s)’).

2.3. Preparing the corpus

The Yupik digital corpus (Schwartz, et al. 2021) is composed of scanned text files from various written sources including primers, storybooks, and fables. It currently consists of ~100,000 tokens and ~40,000 types. There are also a number of Yupik translations of religious texts including a full translation of the New Testament; these were excluded from the current analysis in an attempt to eliminate effects of register or surface word ordering from the English source texts.

We annotated the corpus using a finite state morphological analyzer developed by Chen & Schwartz (2018). The analyzer is still in development, and yields multiple possible parses for ~66% of tokens. A series of regular expressions was used to extract target quantifier contexts, yielding 17,937 tokens. From those, all examples of target word forms with unambiguous analyses (4,953 tokens) or with two analyses that posited the same root and agreed in part of speech were marked for inclusion. The latter criterion was a necessary addition as there are a number of homophonous (and in some cases zero-derived) roots that can appear in possible analyses alongside their counterparts of a different part of speech. All other target quantifier tokens were rejected for this study, except for quantifiers whose data points would have been completely excluded by these criteria.

Given the prevalence of obscurative morphophonemic processes that affect Yupik morpheme boundaries, it is rarely the case that the full character sequence of any underlying quantifier form would appear in any token in the corpus. To ensure that all forms were captured by the regular expression, a truncated version of each quantifier root was used. This meant dropping the final consonant sequence from a root if one was present, or dropping the final vowel (in the case the quantifier *sangwaa* ‘some’, which sometimes appears as *sangwa*). Unrelated forms that happened to share the searched orthographic sequence were eliminated manually. Sentences were also eliminated if they were taken from a corpus file for which there was no available English translation.

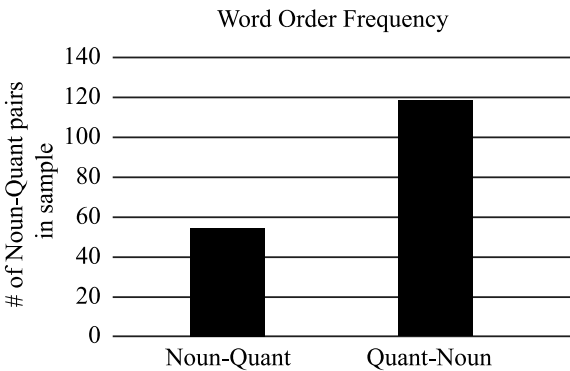
These criteria reduced the total subcorpus to 1076 sentences. These remaining sentences were coded manually with the assistance of their parallel English translations and the corresponding output from the morphological analyzer. Each quantifier-noun sequence was coded for several factors: which quantifier was present, whether it was pre- or postnominal, case and number of the quantifier, and case and number of the noun. Of the 1076 sentences, 581 have been coded, and 172 have been positively identified as cases of nominal quantification. It is with these 172 examples that this analysis was conducted.

3. Results and discussion

3.1. Word order

Figure 2 shows the distribution of word orders in our sample.

Figure 2: Word order frequency in the corpus sample



Previous work has posited that clausal word order in Yupik is relatively free and determined primarily by discourse considerations (Jacobson 2001; de Reuse 1994). De Reuse (1994, p. 25-26) states that possessors tend to precede the noun and that demonstratives and nominal modifiers tend to follow the noun. The corpus sample demonstrated the opposite tendency. Overall, the data show that quantifiers tend to precede the noun more often, even when only non-numeral quantifiers and numerals are considered separately. Quantifier-noun sequences make up ~68% of all sampled word orders, including those with another intervening word. A two-tailed z-test for proportions ($H_0 : p_1 = p_2$) showed a statistically significant difference between the proportions of noun-quantifier and quantifier-noun orders of all word orders ($p < .01$).

Additionally, a number of examples were identified in which the quantifier and noun were separated by another word. In most of these cases, the intervening word-form was another nominal-modifying element, such as a nominal possessor in (1). Other examples of an intervening nominal modifier are demonstratives, as in (2), and participial phrases, as in (3). The quantifier phrase is in italics and the intervening element is bolded.

- (1) Siqlugam paaynga ilangqullaq*katam amiiqam* putungi aliilluteng.
katam amiiqagh -m putu -ngi
only skin -RL.S¹ slit -AB.P.POSS.3S
'The opening of the cellar was icy, *only the slits of the skin* were showing.'
(Koonooka 2003, ungi.part4-alghalek.text27.ess_content, LN:28)

- (2) Nalighni *taakwani* ayumiqulleghni kaviiq akitupigtuq.
naligh -ni taakwa -ni ayumiqullegh -ni
which -LC.P D:R(NEAR) -LC.P year -LC.P
'[In] *some of those* years the price for fox pelts ran very high.'
(Apassingok 1987, volume2.part3.text3.ess_content, LN:335)

- (3) Enta *kalngaggaak* *kumlugpetun* *angkalghiik*
kalngag -ghhagh -k kumlu -gpetun angka -lghii -k
sack -small.N -AB.DU thumb -EQ.DU.POSS.2S be.a.certain.size-PTCPL.INTR-3DU
maalghuk ulimakek.
maalghu -k
two -DU
'Ok, make *two small sacks the size of your thumb*.'
(Koonooka 2003, ungi.part1-ayveghlaq.text2.ess_content, LN:165)

3.2. Agreement

3.2.1. Case

Head nouns and quantifiers were predicted to show agreement in case and number, based on the behavior of other nominal modifiers in the language (e.g., demonstratives), which agree with the modified noun in both dimensions. This appears to hold true for quantifiers as well, based on the corpus sample. 83% of the quantifiers in the sample matched the modified noun in grammatical case. Of those that did not match, approximately one third were instances of a particular set of roots that can only take ergative case. Several other instances were loanwords. A few more seemed to involve undocumented morphology or misspellings, and require further investigation.

¹ Abbreviations: AB - absolutive case, D:E - demonstrative: extended, D:O - demonstrative: obscured, D:R - demonstrative: restricted, DU - dual, IND - indicative, INTR - intransitive, MD - ablative-modal case, N - noun, OPT - optative, P - plural, PART - participial mood, POSS - possessed, PRES - present, PST - past, PTCPOBL - participial oblique case, QQ - quantifier/qualifier, RL - relative (~ergative) case, S - singular, SUB - subordinative mood, TM - terminalis case, TRN - transitive, VL - vialis case

occurring speech (particularly in elders vs. younger speakers), and analysis considering distinctions of mass vs. count nouns, etc., would be beneficial. Additionally, our corpus annotation work lays the groundwork for the development of pedagogical materials for the Yupik community, such as a fully tagged and parsed corpus that can be used in advanced Yupik language study and in the development of natural language exercises for all learning levels, as well as for larger data-driven research projects on various aspects of Yupik grammar.

References

- Apassingok, Anders (Iyaaka), Walunga, Willis (Kepelgu), and Tennant, Edward (Tengutkalek), eds. (1987). *Sivuqam Nangaghnegha — Siivanllemta Ungipaqellghat / Lore of St. Lawrence Island — Echoes of our Eskimo Elders, Volume 2: Savoonga*. Bering Strait School District, Unalakleet, Alaska. Accessible at: https://github.com/SaintLawrenceIslandYupik/digital_corpus.
- Apassingok, Anders (Iyaaka), Walunga, Willis (Kepelgu), and Tennant, Edward (Tengutkalek), eds. (1989). *Sivuqam Nangaghnegha — Siivanllemta Ungipaqellghat / Lore of St. Lawrence Island — Echoes of our Eskimo Elders, Volume 3: Southwest Cape*. Bering Strait School District, Unalakleet, Alaska. Accessible at: https://github.com/SaintLawrenceIslandYupik/digital_corpus.
- Badten, Linda Womkon (Aghnaghaghpiq), Kaneshira, Vera Oovi (Uqiitlek), Oovi, Marie (Uvegtu) & Koonooka, Christopher (Petuwaq). 2008. *St. Lawrence Island / Siberian Yupik Eskimo Dictionary*. Fairbanks: University of Alaska Fairbanks, Alaska Native Language Center.
- Bittner, Maria. (1995). Quantification in Eskimo: A challenge for compositional semantics. In *Quantification in Natural Languages* (pp. 59-80). Springer, Dordrecht.
- de Reuse, Willem J. (1994). *Siberian Yupik Eskimo: The Language and Its Contacts with Chukchi*. University of Utah Press, 101 University Services Building, Salt Lake City, UT 84112.
- de Reuse, Willem J. (2009). Polysynthesis as a Typological Feature: An attempt at a characterization from Eskimo and Athabaskan perspectives. University of North Texas, Denton.
- Gregg, Robert J. (1992). The Survey of Vancouver English. *American Speech* 67(3), 250-267.
- Hunt, Benjamin & Schreiner, Sylvia L.R. (2019). On the status of the determiner phrase in St. Lawrence Island Yupik. Paper presented at the Meeting of the Society for the Study of the Indigenous Languages of the Americas, New Orleans.
- Jacobson, Steven A. (2001). *A Practical Grammar of the St. Lawrence Island/Siberian Yupik Eskimo Language*. Fairbanks: Alaska Native Language Center.
- Kinloch, A. M. (1971). The Use of Pictures in Elicitation. *American Speech*, 46(1/2), 38-46.
- Koonooka, Christopher (Petuwaq). (2003). *Ungipaghaghlanga: Let Me Tell You A Story*. Alaska Native Language Center, Fairbanks, Alaska. Accessible at: https://github.com/SaintLawrenceIslandYupik/digital_corpus.
- Matthewson, Lisa. (2004). On the methodology of semantic fieldwork. *International journal of American linguistics*, 70(4), 369-415.
- Nishiguchi, Sumiyo. (2007). Quantifiers in Japanese. In *International Tbilisi Symposium on Logic, Language, and Computation* (pp. 153-164). Springer, Berlin, Heidelberg.
- Paperno, Denis, & Keenan, Edward L. (eds). (2017). *Handbook of Quantifiers in Natural Language: Volume II*. Germany: Springer International Publishing.
- Schwartz, Lane, Schreiner, Sylvia L.R. & Chen, Emily. (2019). Community-focused language documentation in support of language education and revitalization for St. Lawrence Island Yupik. *Études Inuit Studies*, 43(1-2). 291-311.
- Schwartz, Lane, Chen, Emily, Park, Hyunji Hayley, Jahn, Edward & Schreiner, Sylvia L.R. (2021). A Digital Corpus of St. Lawrence Island Yupik. In *Proceedings of the 4th Workshop on Computational Methods for Endangered Languages*.
- Woods, Howard B. (1999). *The Ottawa survey of Canadian English*. Kingston, ON: Strathy Language Unit, Queen's University.

Proceedings of the 39th West Coast Conference on Formal Linguistics

edited by Robert Autry,
Gabriela de la Cruz Sanchez,
Luis A. Irizarry Figueroa,
Kristina Mihajlovic, Tianyi Ni,
Ryan Smith, and Heidi Harley

Cascadilla Proceedings Project Somerville, MA 2024

Copyright information

Proceedings of the 39th West Coast Conference on Formal Linguistics
© 2024 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-481-2 hardback

A copyright notice for each paper is located at the bottom of the first page of the paper.
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

Ordering information

Orders for the printed edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at www.lingref.com. Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Hunt, Benjamin and Sylvia L.R. Schreiner. 2024. The Distribution of Nominal Quantifiers in a Digitized Corpus of St. Lawrence Island Yupik. In *Proceedings of the 39th West Coast Conference on Formal Linguistics*, ed. Robert Autry et al., 322-329. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #3644.