

Disambiguating homonymous enclitics in Greenlandic

Liv Molich

Oqaasileriffik / The Language Secretariat of Greenland

livm@oqaasileriffik.gl

Abstract

In Greenlandic ambiguities are very common. Some of them concern enclitics which are widely used and can be interpreted as both conjunctive and adverbial particles. The disambiguation of such homographic enclitics was not attempted before 2018, even though a Greenlandic constraint grammar (CG) was initiated more than a decade ago.

In this paper I shall deal with the disambiguational challenges of enclitic particles and discuss where the disambiguation rules should be placed in the CG, and show how some disambiguation problems can be solved by looking at the combination of inflection and enclitic.

This is an important issue, because different renderings of enclitics – like many other morphemes – can change the syntax of the sentence completely.

1 Credits

The Greenlandic CG was initiated in 2008 by Per Langgård in collaboration with Eckhard Bick and Tino Didriksen. The grammar has continuously been improved, especially by Per Langgård, assisted by changing colleagues. A Greenlandic-Danish-Greenlandic machine translation project running from 2017 till the end of 2021 under the auspices of the Language Secretariat of Greenland (Oqaasileriffik 2016) has speeded up the improvement of the CG. Thanks to this work, the Greenlandic CG is now performing better than ever before, but some major issues still need attention.

¹This work is licensed under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>.

2 Introduction

Greenlandic is a polysynthetic, agglutinative, split-ergative pro-drop language with a rich morphology. The level of ambiguity is high, with an average number of 3-4 readings per cohort on the morphological level alone (Oqaasileriffik 2010; Molich 2019:4). Most Greenlandic words consist of a root and several derivational morphemes followed by inflection and sometimes one or more enclitics. When the morphemes are combined, a number of morphophonological phenomena (Langgård 1997) blur the morpheme boundaries, raising the level of ambiguity and resulting in a hard job of disambiguation.

In addition to ambiguities caused by morphophonology, there are also ambiguities at the word class level: Common nouns are not distinguishable from adjectives,¹ and nominal participles are often homonymous to verbal participles. Additionally, enclitics can be interpreted as adverbs or conjunctions, but graphically they do not differ. However, a correlation between participles and enclitics can be found, facilitating the disambiguation.

3 The enclitic particles can be conjunctions or adverbs

Three enclitic particles, {luunniit}, {lu} and {li}, share some features: They can be added to any word and may function as conjunctions or adverbs, without changing the syntactic potential or word class of the word they are added to. If the enclitic particle is conjunctive, it coordinates the phrase itself with the phrase to the left of it – disjunctively, additively, or adversatively. If the enclitic is adverbial, it only modifies the clause that it is part of.

The adverbial enclitics often lead to translations

¹For this reason common nouns and adjectives are treated as belonging to the same word class, nouns.

with subordinate conjunctions if they are added to verbs in the participle or contemporative mood. Such enclitics may therefore be equated to conjunctions, even though they are adverbial seen from a purely Greenlandic point of view.

3.1 An example: {luunniit}

The following quote² is an example of adverbial and conjunctive use of {luunniit} and shows that disambiguation between the two functions is needed:

EX *Maleruagassa-t*

rule-N.Abs.P³

@SUBJ>

suminngaanneer-aluar-aanni-luunniit

be.from.where-even-V.Par.Im-even.Encl.Adv

@CL-ADVL>

atuup-put,

be.in.use-V.Ind.3P,

@PRED

franski-u-gaanni

Frenchman-be-V.Par.Im

@CL-<ADVL

tyrkeri-u-gaanni

Turk-be-V.Par.Im

@CL-<ADVL

kalaali-u-gaanni-luunniit.

Greenlander-be-V.Par.Im-or.Encl.Conj

@CL-<ADVL

“The rules apply **no matter** where you are from, [no matter] if you are a Frenchman, a Turk **or** a Greenlander.”

Both {luunniit} enclitics are added to an impersonal participle -*gaanni*, “you” or “one”, but they clearly have different functions, the first being adverbial “**no matter** where you are from” and the second conjunctive, “**or** if you are a Greenlander”. Formerly, both {luunniit}s were translated by “even though” (*selvom* in Danish). This is clearly not what we intend, especially not where {luunniit} should be translated as a coordinating conjunction.

²KNR 2005. The example has been shortened.

³N=noun; Abs=absolute case; P=plural; V=verb; Ind=indicative mood; Par=verbal participle; Im=impersonal; 3=3rd person; Encl=enclitic particle; Adv=adverbial; Conj=conjunctive.

3.2 Writing disambiguational rules for {luunniit}

As the example shows, the syntactic function of the enclitic cannot be deduced from morphology alone, and syntactic rules must therefore be written in order to be able to do the disambiguation. The syntactic function of the enclitic is shown by adding a secondary tag to the word, *Gram/Adv-encl* or *Gram/Conj-encl*. These secondary tags facilitate correct translations of the enclitics into other languages.

If conjunctive, {luunniit} must coordinate two similar phrases. In the following rule, used in the example above, the secondary tag *Gram/Conj-encl* is added to the enclitic, if a verb in the same mood appears to the left of the enclitic, as well as to the right of or in the same word as the enclitic.

SUBSTITUTE:conjlluun

(LUUNNIIT) (LUUNNIIT Gram/Conj-encl)

TARGET LUUNNIIT + \$\$MOOD

IF (-*1 \$\$MOOD - Gram/Exclm

BARRIER (*) - KOMMA);

3.3 Placing the rules for {luunniit}

Rules such as the above one work well, but often word class disambiguation prior to enclitic disambiguation could be advantageous. The most logical thing to do is therefore to place the rules for {luunniit} after the word class disambiguation rules. However, this is problematic because the disambiguation of enclitics is useful for disambiguation of word class, and word class disambiguation is useful for the disambiguation of enclitics. Both wishes of course cannot be met at the same time.

Disambiguation of nominal and verbal participle is not done until late in the grammar because of the complexity of the task – as shown later. Because of these and other unsolved ambiguities, I have decided to place the enclitic disambiguation rules in the top of the grammar, rather than in the bottom.

Not being able to disambiguate word class before enclitic, I have decided to divide the rules for the conjunctive and adverbial enclitics into groups of safe, less safe and unsafe rules and arrange them by turns: The group of safe conjunctive, safe adverbial, less safe conjunctive and so on. The outcome of this strategy is a reliable disambiguation, as shown in the evaluation below.

3.4 Running the rules for {luunniit}

The Greenlandic CG is run twice, the first time only using the disambiguation rules and the second time using disambiguation rules as well as mapping rules.⁴ This feature is useful for proof-reading the enclitic disambiguation rules: If the secondary tag in the first run is different from the one in the second run, the rules might need an adjustment, or the difference may point to a true ambiguity.

At last, the secondary tag is used for the translation of the enclitic. In cases of true ambiguity or enclitics that for some reason have not been targeted by any of the disambiguation rules, the disambiguation must be performed in a later grammar or directly in the translation lexicon rules.

4 An inflectional suffix homonymous to a derivational morpheme

While the impersonal participle *-gaanni* in the example above can be ambiguous, a much more serious problem is represented by the personal participle written *-toq* or *-soq*, identical to a derivational morpheme, the nominal participle {tuq}.⁵

4.1 An example of *-toq*

The various possibilities of interpreting words ending in the ambiguous *-toq* call for radically different syntactic analyses. A word like *atuartog* can mean either “that he⁶ is⁷ reading” or “a pupil”:

EX <i>Atuar-toq</i>	<i>taku-ara</i>
read-Nzr.N.Abs.S	see-V.Ind.1S.3SO ⁸
The pupil	I saw him
@OBJ>	@PRED
“I saw the pupil.”	

EX <i>Atuar-toq</i>	<i>taku-ara</i>
read-V.Par.3Sg	see-V.Ind.1S.3SO ⁹
That he was reading	I saw it
@CL->CIT	@PRED

⁴According to Tino Didriksen, 97,5% of the disambiguation is done in the first run, and most of the remaining disambiguation is done on the basis of the syntactic tags. The grammar could therefore easily have been split into a disambiguating grammar and a mapping grammar (personal communication, August 2019).

⁵*u* and *o* are orthographic variations of the phonemic vowel /u/.

⁶In Greenlandic, gender is not a grammatical category. Here and later, “he” could just as well have been “she” or “it”.

⁷Present and past tense is normally not marked morphologically, so depending on the context, it could be translated as “was”.

“I saw that he was reading / I saw him read.”

Both interpretations are equally possible and depend on the context alone. Therefore, in cases such as this one, the choice between nominal and verbal participle should only be made with context taken into account.

4.2 Combination of participle *-toq* and enclitic {lu}

Fortunately, real ambiguities are rare. In some cases the ambiguities can be solved by looking at the combination of inflection and enclitic:

EX <i>Taama</i>	<i>oqar-tor-lu</i>
That.Part	say-V.Par.3S-when.Encl.Adv ¹⁰
That	when he said
@ADVL>	@ADVL>
<i>paatsiveerup-punga</i>	
become.confused-V.Ind.1S	
I became confused	
@PRED	

“When he said that, I became confused.”

Here, the adverbial, enclitic particle {lu} and the verbal participle *-toq* are used in combination to show that the first two words form a temporal subordinate clause, “**when** he said that”, which is not coordinated, and should therefore not be translated as ***and when** he said that”.

This combination of enclitic and participle can be written in a simple rule:

```
SUBSTITUTE:Adv02lux
(LU) (LU Gram/Adv-encl)
TARGET LU + Par
IF (NEGATE *-1 V) ;
```

Here, {lu} is marked adverbially if it is added to a verbal participle, and if no other verb is found to the left of it.

In this way, the word class can sometimes be determined by looking at the combination of enclitic and inflection.

5 Evaluation

The disambiguation of enclitic {luunniit}, {lu} and {li} is in most cases done in the first run of

⁸Nzr=nominalizer; N=noun; Abs=absolutive case; S=singular; V=verb; Ind=indicative mood; 1=1st person; 3=3rd person; O=object.

⁹V=verb; Ind=indicative mood; Par=participle mood; 1=1st person; 3=3rd person; S=singular; O=object.

¹⁰Part=particle; V=verb; Ind=indicative mood; Par=participle mood; 1=1st person; 3=3rd person; S=singular; Encl=enclitic particle; Adv=adverbial.

the grammar. In a minor test corpus (Lynge 1976) of 16,400 tokens in 1,951 sentences, 1,395 of the tokens contained one of the three enclitics – almost one per sentence. Among these were 118 {luunniit}s.

Disambiguation of enclitic {luunniit}			
Total	Correct	Incorrect	Not disamb.
118	98	3	17

Of the 17 {luunniit}s that could not be disambiguated after the implementation of my rules, 7 had no morphological analysis at all, and 2 lacked a morphological analysis of the next word in the sentence concerned. 2 were differently marked in the first and the second run of the grammar, and 1 was not disambiguated until the second run. I expect that the rate of correctly disambiguated enclitics will increase when completely analyzed text becomes available.

The effectiveness of the {luunniit} rules in the top of the grammar was tested by the use of a minimal grammar only consisting of {luunniit} rules. The tagging was here very similar to that of the full grammar, but of course without the word disambiguation. This shows that the internal rule order is effective, and that the potential advantage of running the grammar twice is minimal.

As regards the correlation between enclitic {luunniit} and participle, the table below clearly shows that the combination of nominal participle and adverbial enclitic is the least probable one.¹¹

Correlation between participle and enclitic		
	Nominal	Verbal
Conjunctive	611	456
Adverbial	318	429

This knowledge can be used for disambiguating the last enclitics and/or participles, on a statistical basis.

6 Conclusion

Disambiguation problems abound in Greenlandic, and they sometimes call for creative thinking. Some of the problems can be solved by distinguishing between adverbial and conjunctive enclitics. Disambiguation of word class and disambiguation of enclitic are interdependent, and when the disambiguation rules for enclitics are

¹¹In Oqaasileriffik's big corpus with more than 13,000,000 words, 2,614 words contain the combination of participle and {luunniit}. Not all enclitics were analyzed or disambiguated so the numbers in the table do not sum up to the total.

put in the right order, reliable disambiguations can be achieved, resulting in better disambiguation of participles, and more correct translations.

Acknowledgements

My dear husband Flemming A. J. Nielsen deserves big thanks for correcting linguistic errors, and for always cheering me up.

References

- Kleist, Mira, Juana Petrusen and Carla Rosing Olsen. 2017. *Ataqinartuaraq*. Nuuk: Milik Publishing. Translated from Antoine de Saint-Exupéry. 1943. *Le Petit Prince*. New York: Reynal & Hitchcock.
- KNR (Kalaallit Nunaata Radioa). 2005. *Birthe Rønn Hornbech: Siunnersuut akuersaarneqarsinnaanngilluinnarpoq*, <https://knr.gl/kl/nutaarsiassat/birthe-r%C3%B8nn-hornbech-siunnersuut-akuersaarneqarsinnaanngilluinnarpoq>.
- Langgård, Per. 1997. *Forsøg til en Forbedret Grønlandsk Pædagogisk Grammatica*. Nuuk: Atuagkat.
- Lynge, Hans Anton. 1976. *Seqajuk*. Atuakkiorkfik.
- Molich, Liv. 2019. *Solving translational problems through constraint grammar - a practical case study of possession homographs*. Bachelor thesis. Nuuk: University of Greenland. <https://uni.gl/media/4822434/solving-translational-problem-s-through-constraint-grammar-a-practical-case-study-of-possession-homographs-official.pdf>.
- Oqaasileriffik. 2010. *A bit of History*. <https://oqaasileriffik.gl/langetech/a-bit-of-history/>.
- Oqaasileriffik. 2016. *Ukiumoortumik nalunaarut 2016-imoortoq*. <https://oqaasileriffik.gl/wp-content/uploads/2018/08/2016-Oqaasileriffik-final.pdf>.