Per Langgård

How to fight for a digital future – the case of Greenlandic

Abstract

The language technology project that was launched in Greenland in 2005 has attracted quite a lot of attention internationally as one of the few examples of a successful technology project for a lesser resourced language and disproving a hitherto widespread belief that language technology was unrealizable for a language with extreme morphological richness and only a few resources. In this presentation the historical and political background for the project will be outlined and the project's actual progress set out as seen from the viewpoint of the actual developers. A few of the more controversial decisions in the process will be discussed sketchily but the focus will, as far as possible, be kept on observed problems and actual answers to them.

1.   Preamble

The presentation in Dubrovnik underlying the present paper was never intended to be very academic and/or theoretical. On the contrary the focus was deliberately kept on empiricism from the viewpoint of a practician developing language technology from within an administrative system not affiliated with a university or any other academic institution.

The present paper will adhere to the same principles. Accordingly, very limited space will be dedicated to methodological considerations and theoretical discussions while the focus, as far as possible, will be kept on observed problems and concrete answers to them.

2.   A short introduction to Greenland and Greenlandic

From 1721 to 1953, when it became an integrated part of the Danish kingdom, Greenland was a Danish colony. In 1979 Greenland obtained home-rule, followed by self-government in 2009. On October 1st 2021 56,523 persons lived in Greenland out of whom 89.3 % were born there.[1]

---

[1]   Ethnicity is not recorded in Greenlandic statistics while birthplace is. In spite of the minor uncertainty caused by a small number of children being born to Danish parents in Greenland and a small number of children being born to Greenlandic parents in Denmark, it is comparatively safe to equate birthplace with ethnicity statistically.

Compared to most other small[2] languages, Greenlandic has always been strong and vital with

– linguistic rights never really challenged[3] and constitutionally recognized since the Home Rule Act of 1979. Since 2009, Greenland has been monolingual, with Greenlandic the only official language;
– a standard orthography accepted nationwide since 1861 based on the largest dialect but used in education and administration all over the country. It was replaced by the present (phonemic) standard orthography in 1973. The principle of one national orthography irrespective of dialectal varieties is thus well established in Greenland;
– language policy in local control and never tied to religious or political ideology.

Language is not recorded in Greenland's national register; neither has actual language use and competence been investigated scientifically, but for a rough estimate about half of the population are monolinguals in Greenlandic with no or limited command of Danish L2. About 25 % are believed to be more or less balanced Greenlandic-Danish bilinguals and the rest to have Danish L1 with no or limited command of Greenlandic L2. Greenlandic is thus by all standards a very vital language.

3.   Polysynthesis in practice

Greenlandic or Kalaallisut (kal) is the largest dialect in the family of Inuit languages formerly called the Esk-Aleut languages.

Typologically Greenlandic is part of the small group of polysynthetic languages, which, among other characteristic features, include a high level of inflection and a very rich morphology with hundreds of derivational morphemes that combine comparatively freely. A few Greenlandic neologisms will illustrate some of the principles of polysynthesis:

oqaaseq means 'word' – in the plural (*oqaatsit*) it means 'language';
+PAK is a noun-elaborating morpheme that means 'several N'. *oqaaserpaat* thus means 'several words';

---

[2]   The term "small languages" is used here in spite of the fact that it is considered politically incorrect by some. To me the alternatives are worse, such as the widely accepted term "lesser resourced languages". Greenlandic, no doubt, has a limited number of speakers and it is correct that the linguistic institutions in Greenland are very limited in size but in other respects, Greenlandic is much better resourced than maybe most other languages. As one example, public and political focus on the language should be mentioned as a very strong resource in Greenland.

[3]   The so-called Danification period from around 1950 to around 1975 undoubtedly put quite some pressure on the language, however.

+SUAQ is a noun-elaborating morpheme that means 'big N'. *oqaasersuaq* thus means 'a big word' and *oqaaserpassuit* (oqaaseq+PAK+SUAQ) 'very many words';

-LIRI is a verbalizing morpheme that means 'deal with N'. *oqaasileri-* thus means 'work with language', *oqaaserpaleri-* (oqaaseq+PAK+LIRI) means 'deal with a number of words', *oqaasersualeri-* (oqaaseq+SUAQ+LIRI) means 'deal with a big word' and *oqaaserpassualeri-* (oqaaseq+PAK+SUAQ+LIRI) means 'do language technology';

+NIQ is a nominalizing morpheme that forms abstract verbal nouns. *oqaasilerineq* thus means 'linguistics' and *oqaaserpassualerineq* means 'language technology'.[4]

As can be seen, one stem combined with four out of several hundred derivational morphemes generates 12 new stems. If we include inflectional morphology these 12 stems alone will produce more than 3,000 wordforms that all combine freely with about 50 enclitic morphemes generating more than 150,000 individual wordforms.

The rich morphology is a challenge for Greenlandic language technology but, as a matter of fact, a minor problem compared to the syntax problems caused by inderivation[5] and the fact that a number of features like gender, definiteness and tense have no immediate morphological manifestations.

Polysynthesis is a challenge for Greenlandic language technology but not an unsurmountable one as a concrete parsing example will demonstrate. In three different sentences, the same wordform *kusanartumik* (beautiful) has three different syntactic functions as (1) an adnominal argument to an inderived object, (2) an adverbal argument to a main verb and (3) an adverbal argument to an inderived verb inside a noun:[6],[7]

"<kusanartumik nuliaqarpoq >" *He has a beautiful wife*
"kusanar" TUQ N Ins Sg @i->N #1->2
"nuliaq" QAR V Ind 3Sg @PRED #2->0

---

[4]   Note that also *oqaaserpalerineq* and *oqaasersualerineq* are well-formed words.

[5]   The process when a stem after derivation forms part of a new stem of another word class but maintains its original syntactic features. See Langgård (2002) for a thorough introduction to this issue.

[6]   A number of secondary tags for use with higher level analyses have been stripped from the examples for clarity.

[7]   The tags in the example: TUQ is a nominalizing derivational morpheme 'one who Vb'; N is a 'noun'; Ins is the 'oblique case instrumentalis'; Sg is 'singular'; @i->N '"adjective" to an inderived object'; QAR is a verbalizing derivational morpheme meaning 'have N'; V is a 'verb'; Ind is 'indicative mood'; 3Sg is 'subject's person is 3. sing.'; @PRED is 'main verb'; @ADVL> is 'adverbial pointing right'; @i-ADVL> is 'adverbial to inderived verb pointing right'; Abs is the 'case absolutive'; 1Sg is 'subject 1. sing.'; 3SgO is a 'verb inflected for 3. sing. object in the transitive verb'; #n->n are dependencies.

"<kusanartumik oqaluppoq >" *He talks beautifully*
"kusanar" TUQ N Ins Sg @ADVL> #1->2
"oqalup" V Ind 3Sg @PRED #2->0

"<kusanartumik oqaluttoq naapippara >" *I met somebody talking beautifully*
"kusanar" TUQ N Ins Sg @i-ADVL> #1->2
"oqalup" TUQ N Abs Sg @OBJ> #2->3
"naapip" V Ind 1Sg 3SgO @PRED #3->0

As can be seen, the Greenlandic parser has the capacity to automatically distinguish between the different grammatical structures and tag all words adequately.

4. The Greenlandic language technology project – preconditions

Deliberate language planning has always been part of language policy in Greenland. Before 1959, when Landsrådets sprog- og retskrivningsudvalg[8] (the first government institution for language) was introduced, language policy was not explicitly set out in the colonial and early post-colonial administration of Greenland but there can be no doubt that the laissez-faire attitude towards Greenlandic clearly included much respect for the native language of the colony. For instance Greenland's first nationwide newspaper, Atuagagdliutit[9] founded in 1861, was monolingual in Greenlandic. It was printed in Nuuk and distributed free of charge explicitly in order to strengthen the orthographical standard of 1851 (Oldendow 1957).

From around 1990, when grammar and spell checkers started to be used regularly in Danish and English word processing programs, requests for comparable tools for Greenlandic were occasionally heard and a few attempts were actually made to construct Greenlandic spell checkers based on word lists around the turn of the century. With a detection rate as low as 20-25 % they were useless but the wish for language technology to support the vulnerable Greenlandic language slowly started to grow, although it was generally considered an impossible endeavour for a small language. It should be noted that such attitudes were normal among laypeople and language professionals alike.

This discourse began to change in 1999, the beginning of Greenlandic language policy and language planning in its current form, when an academic secretariat

---

[8] This can be translated ad hoc by "The local parliament's committee on language and orthography".

[9] *Atuagagdliutit* literally means 'reading matter given away [for free]'. As a curious but interesting side note, *Atuagagdliutit* was the world's first newspaper with colour illustrations.

for the parliament's three[10] standing committees on language was established with a staff of two. The new institution was later given its present name, Oqaasileriffik/ The Language Secretariat, and has since grown to its present staff of eight.

Already before 1999 there was public and political awareness of language technology as a support for the vulnerable Greenlandic language and there were a few attempts to produce concrete technology. Especially Henrik Aagesen's morphological parser, Qimawin (Aagesen 2004), should be mentioned as a fine example of mature language technology provided by an independent researcher at an early stage. Unfortunately, Qimawin never got the attention it deserved academically and never came into widespread use.

As soon as Oqaasileriffik had been set up, it focused on compiling basic resources and adapting an existing grammar of Greenlandic to prepare it for machine readability. By 2005 the lexical resources and grammatical description had reached a level that made it possible to start up the language technology project on a more ambitious scale.

5.   The Greenlandic language technology project – expected and observed obstacles in the run up to the project's launch in 2005

Oqaasileriffik almost immediately realized that the real problems facing the establishment of a language technology project in a minority society with rather traditional and conservative values were very different from the ones one could expect to have to face. While Oqaasileriffik expected typological questions and technological problems to be the main challenges, it soon became clear that a number of attitudinal problems were much more severe and had to be faced and addressed before the project could be launched:

– In Inuit societies, the primary opinion formers in relation to traditional culture including language are the elders. In their opinion language technology was unnecessary outsiders' technology;
– Polysynthetic Greenlandic deviates far too much from languages traditionally associated with language technology. In addition, the scarcity of training data is expected to render any Greenlandic projects undoable;
– Language technology presupposes a staff of specialized computational engineers and an advanced state of IT infrastructure. Neither exist in Greenland and outsiders' support is not an option since almost no non-Greenlanders speak Greenlandic;
– Language technology is prohibitively expensive;

---

[10] To be precise only the language board and the place names' committee were actual committees while the decision-making body concerning personal names was a rather independent work group affiliated with the bishop's office.

– Should Greenland succeed – against all odds – it will be of no use anyway since the tech giants will never add local tools to their applications for economic reasons.

So while Oqaasileriffik established the process of compiling basic resources, at the same time it had to invest a considerable amount of energy in dialogues with society and in public debates about modern language planning. Fortunately, access to the media is rather open in the small society as are possibilities to deliver public presentations. Both were extensively exploited at the same time as Oqaasileriffik, with the help of small grants for limited projects, was able to publish small applications paving the way for the funding of future, more ambitious projects while attempting to design them in such a way that they could be economically acceptable for funding by Greenlandic public means.

Apart from struggling with the inveterate belief that language technology for Greenlandic is impossible for typological reasons, one other attitude drew much energy from creative work. Conservatism and the high level of respect for elders well known almost everywhere in first-nation societies proved to be major obstacles for a qualified dialogue with society. For Oqaasileriffik to be taken seriously and to pave the way for future funding, the fact that "real" Greenlandic is much more than the elders' sociolect as well as their purism had to be addressed directly. For Greenlandic to survive in the modern world, society had to learn to accept the fact that any language must be able to adapt to hitherto unknown registers and domains. A language used exclusively for local affairs in the past will not survive long.

After a few years as outlined above, the compilation of basic resources had reached a certain size and a new public discourse ready for a language technology project seemed to have emerged so a project constructing the first Greenlandic finite-state transducer was launched in 2005, when Oqaasileriffik received a small grant to relieve one staff member of other duties and got a head start because of generous start-up support from several Nordic universities. Especially Giellatekno in Tromsø directly facilitated the project, including extensive, private teaching of Oqaasileriffik's staff. Without Giellatekno's support in the project's early days, Greenlandic language technology would not have been anyway near its present status.

After a year's work, the first finite-state automaton was mature enough for a spell checker and a few small online tools to be constructed. The spell checker had a modest detection rate of around 80 % and the tools were rather primitive but they were enthusiastically received by society.

They concretely proved that Greenlandic language technology made by local staff is doable, which, looking back, might have been its most important impact.

Over the next few years, the automaton was debugged and expanded with the help of small grants interspersed with periods without funding. This changed

dramatically in 2011 when a €400,000 grant from the Danish Velux foundation enabled Oqaasileriffik to expand the tagger project into a parser project and hire two BA students for in-house training.

From the very beginning it was obvious that training was the key to success and had to be an integral part of the project since there was never the option to pick qualified staff "off the shelf". The study of language technology was not offered anywhere in Denmark in those days.[11] Furthermore, it was next to impossible to raise interest in language technology in the younger generation and to attract students. During a nation-building era, cultural studies, history and other academic disciplines that can be immediately related to a reborn identity as a non-European Inuk were in very high esteem while it proved difficult for the newly established university to rouse students' interest in "European" studies like formal linguistics and computer science.

The training aspect is crucial and to a high degree explains why the 2011-grant turned out to be the paradigm shift it actually was. So to secure the project's future we had to accept in-house training although it was very time consuming for senior staff.

6.   Summing up the challenges and actions taken to answer them

The Greenlandic language technology project is believed to have achieved much better results than almost all other LT projects for very small languages.[12] In Greenland politicians and language administrators are convinced that this is explained to a large extent by the fact that Greenlandic language policy has been consistent, also in situations where the public has been critical of elements of the policy. For instance Greenland has always had only one robust national orthography in spite of rather deviating dialects. Especially among the 3,000 speakers of East Greenlandic this policy is resented by many but the political demand for only one standard has never been seriously challenged. In many minority societies with a more permissive view on dialects, Greenland's one orthography policy is often questioned but the parliament considers standard orthography to be an important tool in preserving Greenlandic.[13]

---

[11]   There were options outside Denmark and a few Greenlanders were actually involved in such programs but the challenges for a Greenlandic speaking student with limited Danish L2 and less English L2 proved to be too prohibitive for us to exploit that option. Also the economic aspect should be mentioned. It is extremely costly to send Greenlandic students to universities outside Denmark.

[12]   The Giellatekno project for Samic languages is one important exception.

[13]   Whether this is the case or not shall not be debated here but it should be pointed out that neither East Greenlandic nor Inuktun in the northernmost part of West Greenland is critically endangered after almost 150 years without local orthographies.

Another factor that is believed to be important in keeping Greenlandic strong is the fact that Greenland, unlike Canada and Alaska for instance, did not put cultural control in the hands of Elders' Councils or the like. Instead comparatively young ministers of culture and directors at Oqaasileriffik have counterbalanced the elders' purist agenda and broadened public opinion about "correct" language.

Another question should be addressed in this context, namely the degree of ambition. In most small languages the criterion for success is keeping the local language alive in relation to local matters while leaving all non-traditional matters like technical terminology, higher education, foreign trade, etc. to be handled by the nearest majority language. In Greenland this is not an option. Neither inside nor outside parliament are voices to be heard advocating diglossic approaches to technical terminology, for instance. Even that must be localized.

Language policy is explicitly set out in the Self Government Act of 2009 to be unrestrictedly monolingual in Greenlandic. A language policy as ambitious as outlined here is, of course, strenuous everywhere in language administration and education but still the policy is believed to have contributed considerably to the healthy state of the Greenlandic language over the years.

7.   Conclusions

Greenlandic is extremely vital in comparison with other small languages. At Oqaasileriffik, it is our firm belief that the rather restrictive, albeit not puristic, language administration pursued over many years has played an important role in ensuring that Greenlandic remains alive and healthy.

The Greenlandic language technology project is an important part of the overall picture as its success depends to a large extent on the fact that it evolved on the basis of a robust standard variety and that language technology in turn reinforces said standards.

Once this starting point of limited permissiveness in both status and corpus planning in Greenland has been established, a few principles and experiences should be mentioned that are believed to have been important in keeping the project alive and growing for so many years.

The project has to be anchored locally. As mentioned earlier, we received much support from Nordic colleagues in the early stages of the project. Fortunately, this support never came in the shape of ready-to-use programs developed outside Greenland. Instead it had the shape of helping to help oneself. Therefore the overall project design as well as all of the tools has been produced locally in Greenland. It should be observed that this does not imply an unwillingness to reach out for help from outside. On the contrary, the small and fragile milieu of Greenlandic language technologists is almost constantly in need of much help – and is lucky enough to get most of what is asked for. But there are important preconditions

to the nature of the help asked for. Only solutions that can be maintained and updated locally by local staff are welcome.[14] This includes the necessity of importing only know-how at a level of abstraction that is viable for the local competence and local education of the local workforce.

The project has to be more ambitious than probably any other language technology project for a language with resources comparable to the Greenlandic ones. The unavoidable fact that a language is a language no matter how few speakers it has is not a question of degree. Accordingly, attempts to develop resources for a variety of any language exclusively for local use in connection with local affairs is not enough. In our globalized world, even small languages need to address unknown topics and unfamiliar domains as much as major languages do. Therefore the Greenlandic language technology project deliberately included "difficult stuff" like technical terminology, neologisms and the like almost from the very beginning.

Only technology that is multifunctional and versatile is viable. Greenland has very limited resources both in terms of manpower and money. One such non-existent resource is a manned institution for NLP using mainstream techniques like machine learning, AI and the like. Accordingly, Greenland must rely on other technologies. Rule-driven technology is an approach Greenlanders can depend on without relying on a foreign workforce because the technology puts limited demands on computational know-how and because Greenlanders are the ones who know the language intimately. It is also a very versatile technology. Once basic lexical resources have been compiled and a tagger and a parser developed, this one set of resources will suffice to construct a number of applications and tools including spell checkers, grammar checkers, and L2 material. It will also take an MT project far if paired with a glossing device.

Permissiveness is a luxury most minority languages cannot afford. This postulate is extremely controversial but Greenlandic decision makers are convinced that there is no alternative if a vulnerable language like Greenlandic should survive for future generations.

As noted repeatedly above, human and economic resources for the Greenlandic language are extremely limited. After the introduction of Home Rule in 1979, the overall situation for the administration of the Greenlandic language obviously improved a lot. Funding has improved dramatically and after establishing an institute for Greenlandic language when Ilisimatusarfik/Greenland's University was

---

[14] There is one important exception. A number of years ago Oqaasileriffik bought a larger application from outside that has proven to be too technical and complex for local competences and has tied Oqaasileriffik to some legal restrictions which cannot be controlled locally. That application is still running and will do so for a number of years until an alternative developed and controlled by Greenlandic human and economic resources can be established.

founded,[15] a small group of Greenlandic-speaking language professionals has emerged making it possible for Oqaasileriffik and the university to fill a dozen or so positions for the administration of Greenlandic and teaching Greenlandic at university level.

Still, although this recent development is very positive, the fact remains that the needs are many and extensive, leaving Greenland in the sad position of efforts put into activities outside a narrow core of daily obligations and immediate political demands for new tools and facilities will inevitably drain resources from the core activities.

So out of necessity rather than inclination, Oqaasileriffik has only rudimentarily included dialects, dialectisms, and varieties such as Facebook-Greenlandic in the basic resources and applications which have been developed recently.

In terms of controversiality error correction is in a league of its own. To most fellow language technologists, applications in general should not always expect correct input. Instead, the programs should deal with typos and other inaccuracies, including dialectisms, in a clever way and process input seamlessly as if the input was given in the expected standard. Greenlandic politicians have explicitly asked Oqaasileriffik not to include error correction to any large degree in our language technology project for pedagogical reasons. A high level of L1 language awareness is, namely, considered important for future vitality and error correction is believed to be counterproductive to this political aim.

Accordingly, the Greenlandic language technology project is basically prescriptive apart from neologisms and morphological reductions of a certain frequency as well as grammaticalization at all descriptive levels. Such natural developments are considered in the present work.

The staff at Oqaasileriffik does not, a priori, see such a political demand as an unjust restriction to their work. The bottom line is that the standardization and prescription that have prevailed in Greenland as far back as records go appear not to have been harmful to the vitality of the language. On the contrary, the dramatic decline in fellow Inuit languages in Canada and Alaska that in many respects are comparable with Greenlandic but where dialectal diversity has been a priority in language policy definitely does not go unnoticed.

We know, of course, that no causality can be postulated exclusively on the basis of this observation but the fact remains that Greenlandic is vital and healthy in the realm of present language policy and that this is a fact we feel we have to consider.

---

[15] Ilisimatusarfik officially became a university in the parliamentary act of May 9th, 1989, but before that a BA in Greenlandic culture including some focus on language had been an option at the university's predecessor, the Inuit Institute in Nuuk, since 1984.

8.   The future for the Greenlandic language technology project

Oqaasileriffik expects the present development to continue and expand in the near future. The basic resources generally have standards which are high enough to develop a wide range of tools and applications as well as to improve existing ones. Likewise funding seems to be secure at least at the present level in the immediately foreseeable future and hopefully beyond.

Funding has actually improved in 2022 with a new chair for a terminologist created in this year's Finance Act and a substantial grant received from the Danish parliament for a private entrepreneur to improve and expand a language technology based Greenlandic L2 system. It is expected that the Danish grant will create much synergy with the projects at Oqaasileriffik.

Apart from matured resources and improved applications, the years to come will see new ones especially in the fields of technical terminology and pedagogical materials for Greenlandic L2. Furthermore high priority will be given to English in Greenland. English resources are scarce and the need for adequate teaching materials at school which do not presuppose Danish as a bridge to English is great as is the general need for modern dictionaries between Greenlandic and English.

One aspect, though, of English in Greenland calls for special attention, namely the great impact of English on Greenlandic via the tech giants that is rapidly increasing everywhere in Greenland after the sea cable laid in 2008 made general access to the internet better and cheaper.

No valid information on the phenomenon is available but quite a number of personal observations and calls from worried parents about Greenlandic children communicating with other Greenlandic children in pidgin-style English suggests that the problem is growing. The primary sources for this kind of English are allegedly YouTube and gaming but extensively used non-localized applications like Google, MS Office and the major operating systems by the adult population are expected to add to the picture.

This present development might be the biggest threat to Greenlandic ever experienced but no one knows what can be done about it. Extensive localizing might reduce the dangers but nothing like that seems to be on the tech giants' cards.

Oqaasileriffik has tried hard to get into contact with the tech giants about the problem but nothing approaching a dialogue has come out of that. Most of the correspondence is simply ignored and on other occasions, we get what seems to be robot-generated reactions that do not address the problems written about at all.

At the moment several initiatives for a working group under the auspices of the Nordic Council of Ministers are in the making but it is not yet possible to predict whether they will be more successful than earlier attempts by Oqaasileriffik.

Still, the clock is ticking and reports on English affecting children's Greenlandic mother tongue are growing more numerous all the time so idling is not an option

for Greenland. The impact of English on Greenlandic is already a fact and nothing implies that this will diminish in the foreseeable future. To prepare for such unavoidable bilingualism in cyberspace lots of work must be done soon. This includes the production of Greenlandic-English MT to render localization a possibility[16] as well as serious refinements of Greenlandic writing aids and lexical resources to make the mother tongue competitive towards English L2 also in technical domains, just to mention two of the many achievements needed

That is all very far away but standing still is going backwards so something must be done. Added to this is the fact that all achievements on route for that goal will improve the quality of tools and lexical resources that all Greenlanders will need access to for Greenlandic to survive in the shadow of English L2.

References

Aagesen, H. (2004): *Qimawin: en grønlandsk ordkløver og samlet ordbog : program til PC med Windows95 eller nyere*. Nuuk: Atuagkat.

Langgård, K. (2002): Inderivation in Greenlandic. In: Nedergaard Thomsen, O./Herslund, M. (eds.): *Travaux du Cercle Linguistique de Copenhague Vol XXXII. Complex Predicates and Incorporation a Functional Perspective*. Copenhagen: C.A. Reitzel, 67–119.

Oldendow, K. (1957): *Bogtrykkerkunsten i Grønland og mændene bag den: en boghistorisk oversigt*. Copenhagen: F.E.Bording.

Statistics Greenland: stat.gl/dialog/main.asp?lang=da&version=202104&sc=BE&colcode=o

---

[16] This assumes that the tech giants, in time, will open up their applications for third-party software like Greenlandic MT.