

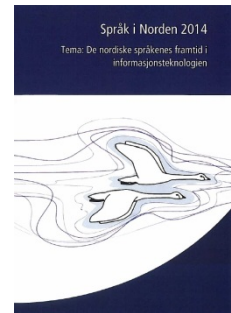
Sprog i Norden

Titel: Selvfølgelig snakker vaskemaskiner grønlandsk i fremtiden

Forfatter: Per Laggård

Kilde: Sprog i Norden, 2014, s. 110-125

URL: <http://ojs.statsbiblioteket.dk/index.php/sin/issue/archive>



© Forfatterne og Netværket for sprognavnene i Norden

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre numre af Sprog i Norden (1970-2004) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Selvfølgelig snakker vaskemaskiner grønlandsk i fremtiden

Per Langgård

Det grønlandske sprog er efter indførelsen af Selvstyret Grønlands eneste officielle sprog. Grønland er således formelt ensproget. Dermed er grønlandsk blandt sprogene med de allerfærreste brugere overhovedet i gruppen af komplette og samfundsbærende sprog. I en tidsalder, hvor afvikling af minoritetssprog langt mere er normen end udvikling af dem, stiller situationen selvsagt enorme udfordringer til det grønlandske sprog, udfordringer, der ikke vil kunne mødes uden udvikling og anvendelse af en for minoritetssprog ualmindeligt avanceret suite af sprogteknologiske hjælpemidler. Artiklen beskriver forudsætninger for og status på projektet og opridser nogle af de særlige problematikker, grønlandsk sprogteknologi må forholde sig til.

Indledning

Det fremgår entydigt af Selvstyrelovens § 20, at grønlandsk er det eneste officielle sprog i det selvstyrede Grønland. Grønlandsk er altså - eller skal bringes til at være - et komplet og samfundsbærende sprog.

Implikationerne af denne simple iagttagelse er mange og utvivlsomt langt mere indgribende end de fleste har forestillet sig det. Herunder ikke mindst de grønlandske politikere. Før indførelsen af Hjemmestyre i 1979 var det dansk, der håndterede stort set alle domæner bortset fra intimsfæren og de traditionelle erhverv. Forskning, udenrigshandel, lovgivning og de mange andre lignende „nye“ emner foregik på dansk, men blev nogenlunde konsekvent oversat til grønlandsk.

Hjemmestyret betød bl.a. en kraftigt øget fokus på det grønlandske sprog. Sproget var i en årrække et såkaldt prioriteret område i Hjemmestyrets politik, og perioden så da også en del initiativer til styrkelse af det grønlandske sprog og til modernisering af sprogrøgt og sprogpolitik; men der skete alligevel forholdsvis lidt konkret i forhold til den meget store plads, sprogspørgsmålet havde i den offentlige debat, og hvor fremtrædende en plads det havde symbolpolitisk.

Blandt de nyskabelser, som perioden trods alt så, bør dog nævnes etab-

leringen af Grønlands universitet med eksplicit fokus på den nationale kulturarv¹ og det grønlandske sprog i 1983 og etableringen af Oqaasileriffik/Sprogsekretariatet i 1999².

I sommeren 2000 nedsatte landsstyremedlemmet for bl.a. kulturområdet en lovforberedende arbejdsgruppe, der bl.a. skulle „*bane vejen for en klar og langsigtet sprogpolitik i Grønland. ...*“. Af arbejdsgruppens rekommandationer³ fremgår det, at det for et sprog med kun ca. 50.000 talere er absolut vitalt at tænke i sprogteknologiske baner, hvis det grønlandske sprog ikke alene skal overleve, men også i praksis overtage de mange domæner, hvor sproget stadig i al væsentligt er dansk.

Rekommandationerne blev positivt modtaget af et enigt landsting, men der fulgte ingen økonomiske midler med de gode viljer.

Det grønlandske sprogteknologiske projekt i historisk overblik

En dårlig start for 10 år siden

Et af de ønsker, der stod øverst på ønskelisten i 2001, var grønlandsk skrivehjælp til MSOffice. I mangelen af tilstrækkelige projektmidler forsøgte Oqaasileriffik derfor at løbe grønlandsk sprogteknologi i gang ved at „springe over, hvor gærdet er lavest“ og fremstille en listebaseret stavekontrol på basis af en ordliste omfattende ca. 350.000 enkeltord, som inspektør Erik Fleischer på privat initiativ havde kompileret over en år-række.

Listen blev korrekturlæst manuelt og rensset for de fleste fejl, hvorefter den blev kompileret som stavekontrol sammen med et automatisk orddelelingsprogram af finske Lingsoft Oy; men stavekontrollen kom aldrig til at virke efter hensigten. Trods listens mange ord viste det sig, at den genererede falske alarmer i ca. 3 af 4 ord i løbende avistekst. En ‘falsk alarm’ betyder i sprogteknologisk fagjargon, at programmet ikke genkender et korrekt stavet ord og derfor markerer det som fejl.

Det var et fejlskøn at starte projektet, og det drænede i flere år ressourcer, der kunne have været anvendt langt bedre, men beviste dog med al ønskelig tydelighed, at der ikke findes nemme løsninger. Det polysynteti-

1 Landstingslov nr. 8 af 16. oktober 1981

2 Etableret efter en forelæggelsesdebat i efteråret 1998 uden selvstændigt lovgrundlag

3 “*...men ordet*” 2001

ske grønlandske sprog behøver en særdeles slagkraftig teknologi før det lader sig behandle maskinelt.

Fra starten i 2005 til nu

Startpunktet for grønlandsk sprogteknologi i sin nuværende form kan ganske præcist fastslås til den 22. april 2005 efter den nordiske konference om skrivnehjælp i Pargas. Det var nemlig dagen, da førsteamanuensis Trond Trosterud fra Tromsø med sin person og sit universitet i ryggen gik aktivt ind og sammen med undertegnede helt konkret skabte strukturen til den grønlandske Finite State Transducer (fst), der har vokset og vokset lige siden, og som stadig er hjertet i det grønlandske projekt. Og selv om indholdssiden for længst er hjemtaget til Grønland, ligger vedligeholdelsen af formsiden stadigvæk i Tromsø. Grønland har ikke selv noget sprogteknologisk miljø hverken datalingvistisk eller i de mere datalogiske aspekter, så uden hjælpen fra Tromsø ville projektet næppe nogensinde være kommet ordentligt i gang.

Derefter gik det stærkt i en periode. Fra efteråret 2005 var jeg frikøbt fra anden tjeneste af en NORDPLUS-bevilling, der fra årsskiftet blev fulgt op af en mindre tillægsbevilling på den hjemlige finanslov. På den basis kunne vi allerede i efteråret 2006 præsentere den grønlandske stavekontrol, *Kukkuniiaat*⁴ i første version. Med en dækning i løbende tekst på kun ca. 80 % var den langt fra „voksen“, men den virkede endda og fik en begejstret modtagelse og hurtigt en meget stor udbredelse især i Grønland.

I de 4 år fra 1.1.2007 til ultimo 2010 måtte det sprogteknologiske projekt nedlukkes i flere perioder på grund af manglende finansiering; men der var dog perioder, hvor det lykkedes at rejse midler til frikøb, så processen aldrig gik helt i stå. I løbet af de år blev *Kukkuniiaat* opgraderet til version 2.0, der med en dækning på ca. 90 % nåede op på Microsoft's de facto industristandard for voksne⁵ stavekontroller, og der blev taget hul på det, der har vist sig at være vores hidtil største udfordring, nemlig den

4 *Kukkuniiaat* er en neologisme, der direkte oversat betyder 'fejlfejneren'. Det illustrerer fint et par polysyntetiske principper for orddannelse. Ordets morfologiske struktur kan gengives som:
kukku- 'at fejle'

NIQ+vn 'verbalnomen'

IAR+nv 'at fjerne N'

UTE+vn 'midlet til at Vb'

5 En danisering af engelsk fagjargon hvor termen *mature* anvendes om fuldt udbyggede stavekontroller

for grønlandsk så afgørende disambiguering. Hermed menes et program, der ud fra en semantisk eller syntaktisk kontekst kan vælge den korrekte ud af flere mulige analyser af de enkelte ord. Grønlandsk er karakteriseret af nogle avancerede assimilationsprocesser, der skaber et meget stort antal homonymer. Dette var tidligere et lidt mindre problem, idet retskrivningen før 1973 i alt væsentligt byggede på samhørighedsprincippet⁶, således at lydlige sammenfald umiddelbart kunne holdes adskilt ortografisk, men med indførelsen af det stort set gennemførte lydlige princip i den nye retstavning af 1973 er antallet af ens stavede ord med helt forskellige betydninger eksploderet.

Samtidig blev den gamle leksikale database radikalt omstruktureret til i al væsentligt at være den aktuelle *Katersat*, som kan besøges på www.oqaaserpassualeriffik.org.

Fra 2011 til 2013 har projektet levet af en rundhåndet bevilling fra den danske Velux-fond. Bevillingen har dels gjort det muligt at frikøbe under tegnede næsten helt fra øvrig tjeneste, og dels har den finansieret en uddannelsesstilling for den første grønlander, der bliver uddannet inden for det sprogteknologiske emne. Endelig har bevillingen muliggjort en ganske effektiv „yngelpleje“, idet vi for første gang nogen sinde har kunnet ansætte gode studenter, der har ytret ønske om at blive tilknyttet projektet, som studentermedhjælpere. Lige nu har vi universitetets absolutte karaktertop i form af 4 studenter, som vi håber at kunne gøre interesse ret nok i sprogteknologi til at de vil videreuddanne sig i faget. Der skal måske understreges et grundvilkår for det grønlandske projekt på dette sted, nemlig at rekrutteringen er et dramatisk problem. Grønland har absolut ingen uddannede sprogfolk, der ikke i forvejen er overbelastet med arbejde, og for næste generation gælder det, at alle talenter formelig flås væk til bedre betalte jobs især i undervisningen og i informationsbranchen.

Samtidig har vi netop modtaget den fantastiske nyhed, at det grønlandske sprogteknologiske projekt fra årsskiftet 13/14 er kommet på finansloven med to nyoprettede stillinger, så for første gang har vi for alvor tillid til at vi kan arbejde i ro fremover.

⁶ Som det i nogen grad kendes fra dansk i ord som *havde* og *vådt*, der fastholder samhørigheden med hhv. *have* og *våd* selv om *-v-* og *-d-* ikke længere har en lydlig realitet.

Indhold og målsætninger

Skriftsprog og sproglig overlevelse

Grønlandsk er sandsynligvis det mest vitale minoritetssprog i hele verden. Grønland er således langt hinsides det normale succeskriterium for hovedparten af verdens minoritetssprog, nemlig tosprogethed med et formelt anerkendt og vitalt minoritetssprog som andetsprog til et større majoritetssprog også blandt børn.

Grønlandsk har - i kortere perioder ganske vist lidt problematiseret - altid været fuldt anerkendt som SPROGET i Grønland også i uddannelserne og i administrationen, og ingen, der har besøgt Grønland, er et øjeblik i tvivl om, at grønlandsk er ekstremt vitalt. Sproget tales overalt af alle, og når man specielt uden for hovedstaden hører dansk, er der tydeligvis tale om ikke fuldt behersket fremmedsprog.

Alligevel er der al mulig grund til at være opmærksom, for trods talesprogets stærke placering er der store problemer med skriftsproget i mange henseender. Den enkeltes skriftlige performans lader oftest meget tilbage at ønske, skriftligt grønlandsk i offentligheden inkl. de skrevne medier er næsten udelukkende oversættersprog efter dansksprogede forlæg, og i en lang række domæner er manglen på terminologi foruroligende.

Dette er selvsagt kritisk i en globaliseret verden, hvor skriftlig information overtager en stigende del af kommunikationen både på det individuelle plan i fx de sociale medier og mellem samfund og borger i form af bl.a. serviceportaler og andre gør-det-selv tjenester.

Skriftsproget, herunder teknisk terminologi, må og skal udvikles og støttes, hvis grønlandsk ikke skal risikere sprogdød allerede på mellem-langt sigt. Der findes næppe noget alternativ til udstrakt anvendelse af avanceret teknologi til virkeliggørelse af dette.

(Fraværet af) grønlandsk L2⁷ som symptom

Grønlandsk talesprog er som nævnt uhyre vitalt med en vigtig undtagelse: Det er kun grønlændere, der taler grønlandsk endda i en sådan grad, at det i den store gruppe af blandinger snarere er reglen end undtagelsen, at dansk er førstesproget og grønlandsk enten ikke-eksisterende

⁷ L2 er en hyppigt anvendt forkortelse for 'andetsprog eller fremmedsprog'. Forkortelsen kommer af engelsk 'second language'.

eller svagt. For etniske danskere gælder det, at stort set slet ingen taler grønlandsk.

Også dette er farligt for det grønlandske sprogs overlevelse. Det er ikke sundt for statusbalancen mellem dansk og grønlandsk, at forpligtelsen til sprogskift udelukkende er grønlandsk, så for etableringen af en sund balance mellem sprogene spiller også grønlandsk som fremmed- og andetsprog ubetinget en rolle.

Det er bare ikke så enkelt at få en kvalificeret grønlandskundervisning etableret alle steder. Bosætningen er som bekendt meget spredt, og der er meget få undervisere, der er uddannet til at undervise i grønlandsk L2.

Så også på dette felt er behovet for teknologiske løsninger påtrængende. Det er derfor ingen større overraskelse, at DVD'en *Grønlandsk for voksne*⁸, der er det første forsøg på at udvikle teknologibaseret undervisning, er blevet en meget stor succes.

Majoritet og minoritet

I forhold til Danmark og resten af verden foregår kommunikationen endtydigt på de andres præmisser. Der er ingen uden for Grønland, der forstår grønlandsk, så det kan ikke være anderledes. Samtidig er forholdet umådeligt farlig for det grønlandske sprogs overlevelse, for det betyder, at enhver form for virksomhed, der forudsætter varer og tjenesteydelser udefra - hvilket gælder næsten overalt i et moderne samfund - nødvendigvis kommer til at foregå på fremmedsproget med den konsekvens, at det grønlandske sprog ikke bliver udbygget og fornyet i den grad, sproget ellers burde udvikle sig.

Hvis grønlandsk altså skal udvikle sig til at være et komplet og samfundsbærende sprog i praksis, skal grønlandsk nødvendigvis gøres til det naturlige konciperingsprog for den nye generation grønlandske professionelle, hvilket igen forudsætter, at der udvikles metoder til at formidle de grønlandsksprogede budskaber på de vigtigste fremmedsprog. Pt. vil det sige dansk og engelsk. Uden maskinoversættelse og udstrakt brug af oversættelseshukommelse kan dette ønske næppe nogensinde realiseres.

Bemærk altså, at det af sprogpoltiske årsager må være oversættelse fra grønlandsk til dansk, der først skal søges fremmet, ikke den modsatte vej⁹.

8 Forlaget LearnGreenlandic 2011. Se www.learngreenlandic.com for yderligere information.

9 Faktisk vil oversættelse den anden vej løbe ind i alvorlige problemer, fordi de eksisterende beskrivelser af grønlandsk - især syntaks - langt fra er præcise nok til at oversættelser fra fremmedsprog kan baseres på dem. Se videre nedenfor i afsnit 3.1.

De nødvendige værktøjer og grundlæggende ressourcer

Et sprog er som bekendt et helt sprog uagtet antallet af talere er stort eller lille, så grønlandsk behøver principielt lige så meget og lige så raffineret sprogteknologi som ethvert andet sprog. Dvs.

- Korpora, både et nationalt korpus og et parallelkorpus
- Ordbøger
- Taggere og parsere
- Skriveværktøjer
- Termbank
- Maskinoversættelse og oversættelseshukommelse
- Talesyntese og talegenkendelse

Bemærk ordet ‘principielt’ ovenfor, for selv om loven om, at et sprog er et sprog, selvsagt er gældende, er det samtidig et faktum, at ikke alle sprog har de samme menneskelige og økonomiske ressourcer at udvikle teknologierne for. Det er derfor ret sandsynligt, at vi i Grønland aldrig vil nå frem til at have samtlige elementer i listen ovenfor repræsenteret; men det forhindrer ikke, at det er den fulde suite af teknologier, der er slutdestinationen, omend målet er uopnåeligt. Bestræbelsen må være at komme så langt som muligt.

Hvordan vi gør det

Teknisk og metodisk er der ikke mange nyheder i det grønlandske projekt. Det bygger udelukkende på kendte teknologier med de nødvendige tilpasninger, som de meget lange ord og den meget komplekse morfologi tilsiger.

Finit state transducer (fst) for grønlandsk

Hjertet i grønlandsk sprogteknologi er en traditionel finite state transducer baseret på two-level tænkningen iflg. Koskenniemi og Karlsson¹⁰.

Det mest karakteristiske særkende ved den grønlandske fst er den enorme størrelse derivationsleksika¹¹ har. I de allerførste forsøg på at etablere den grønlandske automat blev derivationsmorfemerne tilføjet enkeltvis med option for kombination med andre derivationsmorfemer; men det viste sig hurtigt, at tilgangen ikke fungerede. Performansen gik ned, og overgenereringen var ustyrlig allerede ved få snese morfemer. Vi har derfor valgt den noget tungere vej kun at medtage dokumenterede morfemkombinationer, hvilket i sig selv er et omfattende projekt, idet der hidtil er registreret omkr. 5.000 tilhængskombinationer pr. bøjningsklasse. Det betyder at den samlede automat alene for derivationsproblematikken omfatter ca. 130.000 kombinationer.

Parsing

Grønlandsk har en ualmindeligt stram fonotaks og samtidig en strengt fonematisk ortografi, der udelukkende baseres på overfladestrukturerne. Dertil kommer, at der er masser af både assimilation og trunkeringer med tilhørende lydskift, når tilhæng og endelser sættes til stammerne. Som kort omtalt ovenfor giver disse forhold i kombination et meget højt ambiguitetsniveau. Disambigueringen er derfor en uomgængelig forudsætning for at bruge grønlandsk sprogteknologi til andet end stavekontrol og simple ordformsgeneratorer mv. Og samtidig er disambigueringen en kompliceret proces på grønlandsk.

Leksikal database

Der er i løbet af de sidste små 10 år opbygget en ganske betydelig leksikal database i MySQL. Den indeholder i øjeblikket omkr. ¼ mio. leksemer på hhv. grønlandsk, dansk og engelsk.

På grund af de mange homonymer har vi måttet redefinere lemmabegrebet en smule således, at et lemma ikke alene er defineret ved et enkelt nøglefelt - normalt selve opslagsordet - men ved opslagsordet + ordklasse + semantisk tag + (om muligt) opslagsordet i den gamle retstavning.

¹⁰ Af tidlige værker kan nævnes Koskenniemi 1983 og Karlsson 1990

¹¹ Hermed menes listerne med morfemer og regler, der håndterer de tusindvis af kombinationer af afledningsmorfemer (derivativer), der findes på grønlandsk.

Talesyntese

Det stod i 2013 endelig fast, at der vil blive udviklet en syntetisk grønlandsk stemme. Det er danske Mikroværkstedet a/s, der kommer til at udvikle den på basis af prototypen, *Martha*, som Mikroværkstedet for et par år siden etablerede i samarbejde med Oqaasileriffik.

Martha er i en række henseender nyskabende, fordi hun er udviklet stort set uden redundans på basis af meget få, men fonetisk, morfologisk og syntaktisk gennemanalyserede prompter. Det har været nødvendigt at gøre det på den måde, fordi der ikke findes anvendelige korpora på grønlandsk, fordi ordfrekvensen på grønlandsk må være en af de laveste blandt Verdens sprog, og fordi det tonale system på grønlandsk er uhyre labilt. Der findes ikke egentlige faste tryk, men derimod en konstant bevægelse i tonehøjderne, efterhånden som der tilføjes nye tilhæng, endelser og enklitis.

Den „voksne“ *Martha* vil være klar i beta-version i efteråret 2014 og være fuldt udbygget i 2015.

- Og resten?

Næste delprojekt nu, hvor de grundlæggende ressourcer begynder at være på plads, vil være maskinoversættelse til dansk. Det er vi så småt ved at tage hul på og har et håb om at kunne præsentere en alfaversion inden for de næste ca. 5 år.

Af andre projekter i støbeskeen kan nævnes, at vi straks talesyntesen er klar påregner at begynde at (videre)udvikle interaktive, „intelligente“ programmer til grønlandskundervisningen.

For bare at nævne nogle få af de mange mulige muligheder, der er kommet inden for rækkevidde efterhånden som infrastrukturen er blevet bedre og mere stabil.

Udfordringer nu og udfordringer fremover

Vi tør nok med en vis ret hævde, at det grønlandske sprogteknologiske projekt har været en succes, men det hører med til forståelsen af denne succes at nævne nogle af de problemer, projektet har måttet overvinde eller har måttet acceptere trods de forsinkelser og andre ubehageligheder, de har medført.

Jeg skal i det følgende kort diskutere (i) det grønlandske sprogs særlige historie og deraf følgende store huller i institutionel baggrund og i beskrivelsen af sproget (ii) minisamfundets vanetænkning (iii) mangel på

ressourcer af enhver art og (iv) de mørke skyer, der trækker op globalt, ikke bare for grønlandsk, men for alle små sprog.

Behov for grammatisk grundforskning

Af historiske årsager er næsten al vores viden om grønlandsk modersmål tilvejebragt af ikke-grønlandere, hvoraf mange har talt sproget yderst mangelfuldt. Dertil kommer, at den grønlandske sprogbeskrivelse næsten udelukkende har haft et komparativt sigte. Den er i alt væsentligt en hjælp til at finde vejen ind i grønlandsk for ikke-grønlandere. Grønlandsk mangler helt enkelt en autoritativ modersmålsgrammatik.

Man må desuden være opmærksom på, at stort set al grønlandsk sprogbeskrivelse med kun små modifikationer baserer sig Kleinsmidts grammatiske tænkning fra 1851 og langt hen ad vejen på det sprog, Kleinsmidt beskrev dengang.

Denne konstante tænkning i komparative baner betyder, at „sprogvidenskab“ langt ind i de professionelles rækker er ensbetydende med ordlister med grønlandske oversættelser af danske udtryk. Komplerede immanente analyser på mange beskrivelsesniveauer uden umiddelbart komparativt sigte har ikke rigtigt nogen plads i denne tænkning.

Det betyder endvidere, at beskrivelser af grønlandsk på niveauer over enkeltordsniveau kun eksisterer rudimentært eller som korte, generelle fremstillinger, der langt fra går i dybden. Syntaks og semantik er helt enkelt meget vanskeligere at forholde sig til end fonologi og morfologi for forskere, der ikke helt behersker sproget selv.

På den baggrund bør det være let at forstå, hvorfor den for sprogteknologien nødvendige beskæftigelse med fx disambiguering og semantiske frames af mange opfattes temmelig virkelighedsfjernt. Vi kunne jo i stedet bruge kræfterne på noget „fornuftigt“ nemlig at udvikle metoder til at producere tosprogede ordlister hurtigt og effektivt, lyder logikken.

Der var engang, hvor denne komparative tænkning gav en slags mening, fordi de fleste af beslutningerne alligevel blev truffet på de nordiske magtsprog og fordi ingen aktivitet uden for lokalområderne foregik på lokalsprogene.

Det er ikke længere tilfældet. Autonomien fandt vej til Grønland i 1979. Nu tænkes tankerne på grønlandsk, men de formuleres kun delvist på grønlandsk. Reelt er der stadig en hel del dansk – og dermed en hel del oversættelse – tilbage i beslutningsprocesserne og i implementeringen af dem.

Det er ikke længere godt nok, men det kræver meget for alvor at ændre denne tingenes tilstand, herunder at den grønlandske grammatik nyskrives som reel modersmålsgrammatik. Det er et stort projekt, men der er efter min opfattelse ingen vej udenom. Uden en dyb og detaljeret beskrivelse af det moderne sprog, som det tales i dag, vil vi ikke kunne tage teknologien - eller pædagogikken og kulturen for den sags skyld - til næste niveau.

Det samfunds bærende grønlandske sprog i Grønland skal selvfølgelig kunne udtrykke lige så mange forhold og gøre det lige så præcist som fx det samfunds bærende danske sprog i Danmark. Det er så selvfølgelig, at det ikke burde være nødvendigt at gentage, men implikationerne for grønlandsk sprogforskning og sprogrøgt er mildt sagt svimlende. Det betyder nemlig, at den lille håndfuld forskere, der findes ved sprogsekretariatet, i den højere uddannelse og på forlaget skal kunne præstere det samme på grønlandsk som de tusinder af specialister, der varetager de samme funktioner fx på dansk i Danmark.

Så selv om vi jo næppe kan forvente grønlandske pendanter til *Svenska Akademiens ordbok* eller *Det danske sprogs historie* betyder det ikke, at forskning i immanent sprogbeskrivelse kan undværes. Nogen må gøre noget, hvilket vil sige så meget som muligt. Det vil ud over at styrke modersmålsfaget gøre det meget enklere at lave god grønlandsk sprogteknologi.

Vanetænkning både internt og eksternt

På den ene side er det let at være sprogteknolog i Grønland, for interessen i vore aktiviteter er stor, og vi nyder tydeligvis også ganske stor respekt i samfundet. På den anden side er det svært at få øje på konkrete tiltag og handlingsplaner, hvor vore resultater er omsat til praksis fx i undervisningen, i sprogrøgten eller i forlagene.

Jeg ved af personlig erfaring fra gæstelærervirksomhed på universitetet og i fremmedsprogsundervisningen, at de nye programmer ret enkelt kan bringes til at virke også pædagogisk, og at eleverne tager rigtigt vel imod dem og ønsker sig meget mere i samme stil, men alligevel bliver de nye muligheder for randomiserede øvelser, selvkontrollerende hjemme- og gruppearbejde og meget mere ikke rigtigt brugt af andre end os selv endnu.

Noget principielt tilsvarende opleves også i sprogrøgten og i forlags-

virksomheden inkl. ordbogsarbejdet, hvor de nye teknologier endnu ikke er taget i anvendelse.

Det kan selvsagt ikke udelukkes, at det er os selv, der har været for dårlige til at formidle budskabet, men det kan også skyldes, at det kræver mange år og måske endda helt nye generationer sprogarbejdere, at få den nye tænkning, som praktisk anvendelse af teknologien forudsætter, til at slå igennem. Og at der altså efter arbejdet med at skabe mulighederne følger en pædagogisk opgave, der ikke må undervurderes.

Dette er selvfølgelig en kendt situation også for alle andre end os, men jeg vil alligevel vove den påstand, at problemerne er en del tungere i minoritetssprogene end i de større sprog. Dels er den gruppe sproglærere, forlagsredaktører mv. der forventes at opfange teknologierne og implementere dem til praktisk anvendelse meget mindre end i lidt større samfund, således at der ikke rigtigt eksisterer et niveau anvendte forskere som bindeled mellem grundforskningen og de konkrete anvendelser, og dels er fokus i modersmålsundervisningen i minoritetssprog af mange årsager næsten undtagelsesløst retrospektivt.

Der er for mig slet ingen tvivl om, at det retrospektive fokus i kombination med problemerne med at implementere nye teknologier tilstrækkeligt hurtigt er dybt problematiske for minoritetssprogene. Praksis har tydeligt vist, at det ikke er tilstrækkeligt at sikre reel lokal kontrol med lokalt sprog og sikre formel og reel anerkendelse af alle sproglige rettigheder¹². Uden samtidig indtænkning af lokalsproget vs. det globale samfund vil processen aldrig lykkes.

Der findes adskillige minoritetssprog, der har det meget dårligt, selv om de har høj status i lokalsamfundet og formel anerkendelse i det omgivende samfund. Inuit-sprogene i Canada er et tydeligt eksempel. Der har der været sproglig selvbestemmelse i hænderne på the Elders i 14 år. Der er tale om en konstruktion, som tilsyneladende alle også uden for Canada respekterer. Men intet tyder på, at dette har været nok til at knække de nedadgående vitalitetskurver derovre. Tværtimod synes de nedadgående tendenser at være blevet forstærket¹³.

Den slags erfaringer sætter nogle af os i et slemt dilemma, for når man som forsker er overbevist om, at det eneste der – måske! - findes mellem

12 Ofte kaldet *linguistic rights*. Se fx Skutnabb-Kangas et al. 1988

13 Det skal dog understreges, at meget af den forskning, der vurderer tendenserne på den måde, opfattes stærkt kontroversielt af mange i Nunavut.

et minoritetssprogs overlevelse eller ej, er rationel planlægning og struktureret arbejde, tvinges vi ikke sjældent til at forlade den nemme politisk korrekte hovedvej og begive os ud i det modsatte med en til vished grænsende sandsynlighed for at blive skudt grimme ting som kulturimperialisme eller akademikerhovmod i skoene.

Bundlinien er altså den samme for både minoritetssamfundets egne medlemmer og for eksterne specialister, nemlig at det er rationel handling, der er behov for, ikke politisk korrekthed.

Manglende ressourcer

Det grønlandske modersmålsfag har i mange år både i grundskolen og i gymnasiet været bedrevet som et dannelsesfag næsten uden fokus på fagets instrumentelle aspekter af lærere, der har lært deres elever at holde af og respektere modersmålet, men som ikke samtidig har lært dem et godt, varieret grønlandsk og objektiv afstand til eget sprog endsige lært dem et anvendeligt metasprog.

Dette hænger i høj grad sammen med tankerne ovenfor i kapitel 3.1. Modersmålsfaget mangler helt enkelt de solide beskrivelser, som lærerne kunne basere en relevant og motiverende undervisning på.

Det er godt, at de unge har lært at holde af deres modersmål, men det er ikke nogen ideel baggrund for næste generations sprogteknologer. Computeren har brug for operationaliserbare generaliseringer og kan ikke bruge varme følelser til ret meget.

Så både for næste generations sprogteknologi og sprogrøgt er det på høje tid, at nogen med passende politisk gennemslagskraft får gjort noget også ved denne problematik.

Sprogteknologi tilbyder en enorm støtte for et minoritetssprogs overlevelse, men samtidig må man gøre sig det helt selvfølgelige forhold helt klart, at der lige så lidt findes særlige mikrometoder som der findes "små" sprog. Sprogteknologi til et givet sprog må nødvendigvis beskrive hele sproget, så teknologierne i sig selv er principielt lige store uagtet antallet af talere af det sprog teknologien retter sig imod.

Det ER svært, men der kan gøres meget ved at dele ressourcer og ved så vidt muligt at koordinere så mange arbejdsgange som muligt for at facilitere ressourcodelingen. Folkene i Tromsø har gjort en imponerende indsats for at skabe sådanne åbne ressourcemiljøer til umiddelbar gavn for mange minoritetssprog rundt om i verden. Uden dem ville det grøn-

landske projekt under ingen omstændigheder være, hvor det er. Det ville måske aldrig være kommet i gang.

Small is NOT beautiful

Der er nogle afgørende forskelle mellem at arbejde med de morfologisk relativt ukomplekse nordiske sprog og morfologisk tunge sprog som finsk og samisk for slet ikke at snakke om den absolutte morfologiske sværvægter, grønlandsk.

Dette er selvfølgelig ikke en overraskelse for de sprogligt indviede, men for ikke-fagfolk er det bestemt ikke nogen selvfølge, så de har svært ved at forstå, at vi ikke bare hjælper fx Google med at udvikle noget Google Translate i stedet for at bruge masser af år på at gå de lange omveje via disambiguering, dependenser, semantiske frames og andre „uforståeligheder“ eller går den lige vej og kontakter Mac for at få *Siri* til at snakke grønlandsk.

Når vi forsøger at forklare, at Google's og andres teknologi ikke rigtigt kan bringes til at fungere på sprog med en typologi som grønlandsk, opfattes det oftest som bortforklaringer eller akademisk perfektionisme eller mødes med en lettere hovedrysten.

Det er nogenlunde enkelt at se bort fra den slags reaktioner, så længe synspunktet fremføres af lægmænd uden særlige magtmidler; men når synspunktet høres fra et medlem af finansudvalget eller et medlem af forskningsrådet er det somme tider meget svært at sidde overhørigt.

I forhold til fagets rigtigt tunge spillere såsom Microsoft eller Nuance skaber netop dette forhold - som det ser ud i øjeblikket - uoverstigelige problemer. For firmaer som de nævnte eksisterer Grønland og Sápmi knap og nap på verdenskortet, og det har vist sig umuligt at opnå en blot nogenlunde tilfredsstillende adgang til at lægge plugins ind i giganternes software fx MSOffice. I stedet henvises vi til at benytte de indbyggede standardprogrammer baseret på statistisk teknologi. Typisk får vi slet ingen svar, når vi påpeger, at den tilbudte teknologi er komplet uanvendelig på sprog så morfologisk rige som grønlandsk og at regelstyret teknologi er sine qua non for os. Sprog med meget få talere er simpelthen for økonomisk uinteressante til at sådanne firmaer tager sig tid til overhovedet at besvare vore henvendelser, uden at vi foretager os noget ekstraordinært som eksempelvis at lade en minister underskrive henvendelsen. Så får vi måske et svar, men tit bare i form af en maskingenereret respons, der alligevel ikke besvarer, hvad henvendelsen egentlig handlede om.

Pt. ser det sort ud. Til og med Windows XP var grænsefladerne nogenlunde acceptable, i Windows 7 og 8 er de klart utilfredsstillende, men det er dog muligt at indlægge tilføjelsesprogrammer til batch-kørsler. I WindowsRT er ikke engang den mulighed åben, og vi har al mulig grund til at antage, at det kun bliver værre i kommende versioner af operativsystemet.

En humlebi kan ikke flyve, men bare vingeslaget er hurtigt nok, gør den det

Vi ER oppe mod stærke kræfter vel at mærke både globalt og lokalt, men vi er ikke klar til at acceptere 'mission impossible'. Både folkeligt og politisk er 'glokalisering' - dvs. kravet om et konkurrencedygtigt lokalsprog i eget samfund midt i den globaliserede virkelighed - helt reel. Bemærk ordet 'konkurrencedygtigt'. Et lokalt hjemmesprog, der i praksis kun bruges i privatsfæren om kendte, indenlandske forhold kvalificerer ikke til betegnelsen. Små samfundsbærende sprog må nødvendigvis være lige så funktionelle som de store sprog, der omgiver dem. Som historien tydeligt nok har vist det, overlever de ikke ret mange generationer, hvis de ikke lever op til dette krav.

Der er som påvist mange forhold, der skal være på plads og spille sammen, hvis det grønlandske projekt skal lykkes og Nordens sproglige diversitet fortsat være en af kronjuvelerne i vores kulturarv. Vi vover som sagt den påstand, at avanceret sprogteknologi til sprogene med de færreste brugere er helt central i denne forbindelse, for uden den kan så få mennesker, som der her er tale om, helt enkelt ikke følge med til at udvikle de mange sproglige hjælpemidler og tjenester, som et globaliseret samfund behøver.

Så vi arbejder videre i erkendelsen af, at vi jo nok ikke har kræfterne til at udvikle alle de snesevis af hjælpemidler, vi kunne ønske os, men også i erkendelsen af, at grønlandsk under alle omstændigheder står stærkere, hvis vi kan hale de første 10 teknologier fra ønskelisten i hus, end hvis vi kun kan levere de første 5!

Per Langgård er chefkonsulent ved Oqaasileriffik/ Sprogsekretariatet og har fra 2005 haft ansvar for udviklingen af grønlandsk sprogteknologi.

Summary

With the introduction of Self Government in Greenland in 2009 Greenlandic was installed as the only official language in Greenland. The implications of this for a language with less than 50,000 speakers are overwhelming. In the article a case is made to show that this policy cannot be achieved without the development of advanced language technology. Furthermore, it is demonstrated that normal (stochastic) approaches will not work with polysynthetic Greenlandic whereas rule-driven technology already has proven to be efficient. Finally, the article identifies a number of problems specific to Greenland and Greenlandic that the language technology project must address.

Litteratur

- Karlsson, Fred., 1990: Constraint Grammar as a Framework for Parsing Unrestricted Text. In: H. Karlgren, ed., *Proceedings of the 13th International Conference of Computational Linguistics*, Vol. 3. Helsinki 1990, 168-173.
- Koskenniemi, Kimmo, 1983: Two-level Morphology: A General Computational Model for Word-Form Recognition and Production, *Publications*, p. 160, University of Helsinki, Department of General Linguistics.
- Landstingslov nr. 8 af 16. oktober 1981 om Ilisimatusarfik (Inuit instituttet)
- „...men ordet“. *Rekommandationer fra Arbejdsgruppen for sprogpolitisk redegørelse*. Oqaasileriffik, Nuuk, 2001
- Skutnabb-Kangas, Tove and Jim Cummins (eds.), 1988: *Minority education: from shame to struggle*. Clevedon, Avon: Multilingual Matters.