

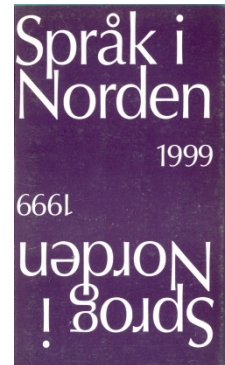
Sprog i Norden

Titel: Føresetnader for å ta vare på dei nordiske språka i språkteknologien

Forfatter: Arne Kjell Foldvik

Kilde: Sprog i Norden, 1999, s. 5-8

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



© Nordisk språkråd

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.

Føresetnader for å ta vare på dei nordiske språka i språkteknologien

Arne Kjell Foldvik

I dette innlegget tar jeg for meg norsk språk i forhold til språkteknologi og mener at det jeg sier om norsk, også har relevans for de øvrige nordiske språka.

En opplagt føresetnad for å ta vare på norsk språk innafor taleteknologi er at vi bruker norsk i dette arbeidet og utvikler norske språkteknologiske produkter.

Jeg vil starte med et lite taleteknologisk tilbakeblikk. Det er i år 30 år siden den første syntetiske ytringa blei laga på norsk. Jeg laga den på PAT i Edinburgh da jeg studerte der. (PAT står for 'parametric artificial talking device'.) Ytringa som jeg sendte som julehilsen på lydband til fonetikerne ved universitetet i Bergen, lød: "PAT i Edinburgh ønsker dere god jul og godt nyttår". Det tok ei hel uke med intenst arbeid og endeløst mye prøving og feiling for å få denne ytringa til for 30 år siden. Dessverre viste det seg at mottakerne i Bergen ikke forstod noe som helst av den.

I dag kan vi skrive en hvilken som helst norsk setning på tastaturet på datamaskinen, trykke linjeskift og få setningen lest opp i neste øyeblikk. I tillegg er kvaliteten så god at vi forstår det som blir sagt første gangen vi hører det.

Er norsk språk trua i taleteknologen i dag? Svaret mitt er et rungende Ja!, og trusselen kommer, som vi alle veit, fra engelsk. Det er viktig å fastslå at det ikke er språkimperalistiske briter eller amerikanere som representerer trusselen. Det er derimot nordmenns egen holdning til norsk og til engelsk som gjør at engelsk språk er en trussel for norsk språk.

Engelsk har her i Norge en høg prestisje, er omgitt av velvilje og har positive konnotasjoner. Viss jeg skulle sette det litt på spissen, ville jeg si at oppfatninga er denne: Dersom en er god i engelsk, er en intelligent, er en dårlig i engelsk, så er en dum.

Svært mange nordmenn har – kanskje nettopp derfor – ei helt urealistisk positiv bedømming av sine egne engelskkunnskaper. I ei rundspørring nylig på universitetet her i Trondheim, svarte ca. 85 % av sivilingeniørstudentene at de beherska engelsk særdeles godt. De samme studentenes egne skriftlige ferdigheter i engelsk, slik de viser seg i diplomarbeidene deres, er langt fra særdeles gode. De er tvertimot ofte fulle av helt elementære ortografiske og grammatiske feil. I en kronikk i Aftenposten i forrige uke påviste en av de som jobber ved Britisk institutt ved universitetet i Oslo at situasjonen var den samme der. Flere av de som underviser i engelsk her i Trondheim, forteller at de har vansker med å få studenter som er svake i engelsk, til å akseptere at de faktisk er det.

Det er et tankekors at vi som har sittet og dere som sitter i Norsk Språkråd, har brukt og bruker tid og energi på lange ordskifter om talemålsnormering, og at Norsk Rikskringkasting har tilsatt en språkkonsulent på bokmålssida og en på nynorsk-sida som skal gi ris og ros om norskkuttalen til NRK-medarbeiderne i radio og TV. Disse to konsulentene tar på alvor klager fra lyttere som hevder at det er ord og uttale i NRK som gjør at de ikke forstår det som blir sagt. Alle disse orda ville klagerne finne betydninga på dersom de slo dem opp i ei større norsk ordbok. Men samtidig, og tilsynelatende uten at noen protesterer i det hele tatt, kommer det nå innslag i nyhetssendingene NRK radio på engelsk, ofte rask, utydelig engelsk med dårlig signal/støyforhold. Disse innslaga blir ofte ikke oversatt i det hele tatt. I løpet av slike innslag vil det garantert være en mengde ord som ikke blir forstått, men da er det ingen som protesterer overfor NRK. Nei, da slår vel lytterne helt frivillig opp orda i ei engelsk-norsk ordbok så de kan bli bedre i engelsk.

Det er vi nordmenn med vår underdanige holdning til engelsk som er den største trusselen mot norsk språk i vår IT-tid. Når engelsk får komme ukritisk inn bl. a. på den måten jeg har nevnt, er det at vi står overfor en reell trussel om domenetap for norsk. Det er alvorlig at f. eks. manualer for nye versjoner av dataprogrammer ofte først foreligger på norsk lang tid etter at en

har fått installert programmet på datamaskinen og fått den engelske manualen, for når den norske utgaven endelig foreligger, har vi ikke behov for manual lenger, men har kanskje så smått begynt å se fram til neste programversjon.

En enda alvorligere trussel mot norsk er det at vi for noen hundrelapper nå kan skaffe oss dikteringsprogrammer på engelsk for PC der vi kan diktere og redigere dokumenter helt uten tastatur og mus. Vi kan snakke i normalt taletempo med 100–160 ord i minuttet og få teksten opp på skjermen med en gang med mindre enn 5 % ordfeil. Siden vi ikke engang er i nærheten av kunne tilby noe slikt for norsk, kunne det være fristende raskt å diktere et brev på engelsk i stedet for å skrive det langt langsommere for hand på norsk, spesielt dersom vi veit at den norske mottakeren av brevet skjønner engelsk.

En annen sjølsagt forutsetninga for at norsk skal overleve i IT-alderen, er at Stortinget bevilger penger til utvikling av norsk taleteknologi, for taleteknologisk forskning er dyrt. Det er ikke riktig som det står i en artikkel i Språknytt nr. 3-1998, at det har blitt brukt lite penger på språkteknologi i Norge. Innafor taleteknologi er det faktisk i løpet av de siste åra blitt brukt mye penger. Men vi kan enes om at dette er utilstrekkelig, og at det er langt mindre enn det f. eks. svenskene har brukt på taleteknologisk forskning.

For å lykkes med taleteknologi er det helt avgjørende å ha store taledatabaser av mange talere fra alle deler av landet, barn, kvinner menn, unge og gamle, tatt opp under ulike forhold, over telenettet, i kontorlandskap, i stille og i støyende omgivelser som en kan trene talegjenkjenningsprogrammene på. Disse taleopptaka må transkriberes og segmenteres. Det tar mye tid og må gjøres av folk som kan det. Det er viktig å få politikerne som bevilger pengene, til å forstå at det er helt urealistisk at enkeltfirmaer innafor taleteknologi som f. eks. Telenor skulle bygge opp disse basene aleine. Her må det penger både fra næringslivet og fra staten til, også fordi det å ta vare på språket vårt, er en viktig kulturoppgave.

Forutsetninga for at de nordiske språka skal ha sin sjølsagte førsteplass foran engelsk innafor språkteknologi her i Norden, er at vi har alle de språkteknologiske verkøya og kundetilbuda på våre egne språk. Men ikke bare det: Det vi må kunne tilby av taleteknologi på nordiske språk, må helst være bedre enn det som vi blir tilbydd på engelsk og dessuten ikke vesentlig dyrere heller.

Her er det viktig å ta med en siste ting: Et dikteringsprogram er ikke godt nok dersom det bare aksepterer og forstår bare én form for talt norsk. De engelske dikteringsprogrammene som er i handelen i dag, klarer like godt Glasgow-engelsk, overklasse britisk-engelsk, amerikansk-engelsk, indisk-engelsk som min egen form for norsk-engelsk. Jeg vil gjerne få gi den rosen til forskerne Jon Emil Natvig og Knut Kvale ved Telenor FoU, at i det arbeidet de har gjort med norsk taleteknologi hittil, har forutsetninga deres hele tida vært at alle dialekter og sosiolekter skal kunne brukes med det samme gode resultatet. Jeg håper at også andre firmaer som nå kommer med i taleteknologi for norsk, kommer til å ha denne målsettinga: Teknologien skal tilpasses brukernes egen måte å snakke på. Ei forutsetning for statlig støtte til oppbygging av taledatabaser må sjølsagt være at ulike varianter av norsk blir representert. Det er viktig å være klar over at taleteknologi ikke nødvendigvis må gi ensretting i uttale: Telenor har alt nå ei telefontjeneste der en kan få greie på om tog er i rute eller forsinka. Gjenkjenningsprogrammet i denne tjenesta "forstår" uttale med tungespiss-r og skarre-r akkurat like bra, og tjukk l og vanlig l går like greitt.

Dersom de nordiske språka inntar førsteplassen foran engelsk i språkteknologi, kan det ha ei gunstig virkning på det punktet som jeg starta med: Knefallet overfor engelsk som et "finere" språk enn f. eks. norsk. Dersom vi får nordiske språkverktøy til å fungere like godt eller til og med bedre enn de engelske, øker vi statusen for våre egne språk og kan få folk til å forstå at de små språka våre til og med duger innafor høgstatusdomener som informasjonsteknologi.