

Kære gæster.

Jeg vil begynde med at ønske os alle hjertelig til lykke med åbningen af dette enestående ordbogsprojekt, ISLEX. Særlig vil jeg gratulere Halldóra og Þórdís med det fantastiske job som de har gjort, og alle medarbejdere i Island, Danmark, Norge og Sverige, samt projektets styregruppe. Da arbejdet med ISLEX begyndte for alvor på Leksikografisk Institut, Orðabók Háskólans, for seks-syv år siden var jeg formand for bestyrelsen af instituttet, og deltog i en uformel gruppe som var med til at planlægge projektets første skridt. Efter at instituttet blev til en afdeling i Árni Magnússon Instituttet, Stofnun Árna Magnússonar í íslenskum fræðum, for fem år siden har jeg ikke haft en formel kontakt til projektet. Trods det har jeg fulgt det med megen interesse hele tiden og det er mig en stor glæde at blive vidne til åbningen af ISLEX her i dag.

Jeg regner med at jeg i fremtiden vil have stor nytte af ISLEX som en almindelig bruger, men i dag står jeg her først og fremmest som repræsentant for sprogteknologien. Jeg deltager for tiden i et stort europæisk projekt som kaldes META-NORD. Alle de nordiske og baltiske lande er med i dette projekt som finansieres af EUs ICT PSP, dvs. Information and Communication Technology Policy Support Programme. Projektet startede den første februar i år og skal vare i to år. Det har som mål at udvikle og dokumentere effektive metoder til at bygge sprogresurser for sprogene i de baltiske og nordiske lande, med fokus på semi-automatisk og automatisk resurseudvikling. Meningen er ikke at bygge eller udvikle nye sprogresurser fra grunden indenfor dette projekt, men at færdiggøre og standardisere eksisterende resurser og gøre dem tilgængelige gennem ressourcesamlingen META-SHARE.

META-NORD har to søsterprojekter; CESAR i Sydøsteuropa og METANET4U i Sydvesteuropa. Over disse tre projekter er paraplyen META-NET som er et Network of Excellence som har til formål at skabe det teknologiske fundament som er forudsætningen for et flersproget europæisk informationssamfund. META-NETs langsigtede mål er at fremme sprogteknologi af høj kvalitet for alle Europas sprog for dermed at opnå en højere grad af politisk og økonomisk enhed gennem kulturel diversitet. Teknologien skal fremme nedbrydningen af eksisterende barrierer mellem Europas sprog og man håber at kombineret med intelligente applikationer og løsninger vil sprogteknologien i fremtiden betyde at europæere nemmere kan tale sammen og lave forretninger med hinanden selv om de ikke taler samme sprog.

Hvorfor taler jeg om sprogteknologi her og nu, på åbningsfesten til ISLEX? ISLEX er jo et leksikografisk projekt, men ikke sprogteknologisk. Men sprogteknologi er utænkelig uden gode leksikalske resurser, og META-NET lægger netop særlig vægt på at oprette og tilreddegøre flersproglige resurser som kan bruges til brobygning mellem forskellige sprog. Så vidt jeg ved er ISLEX en helt enestående resurse i dette henseende. Jeg kender ikke nogen anden ordbog af denne størrelse som kombinerer fem sprog. Jeg tror det er uhyre vigtigt at vi benytter os af dette til gavn for sprogteknologien. Vores sprog er små, i den forstand at de har relativt få talere, men de kan styrkes betydelig ved at samarbejde og udvikle fælles leksikalske resurser.

Hvilken nytte kan sprogteknologien have af flersproglige resurser som ISLEX?

Maskinoversættelse er naturligvis det første man tænker på. I det sidste par år har oversættelsessystemer som baseres på statistiske modeller, især Google Translate, været de mest populære. Disse systemer bruger ikke tosproglige eller flersproglige ordlister men trænes på parallelle tekster, dvs. tekster som findes på to eller flere sprog. Denne metode er derfor afhængig af at der eksisterer store mængder digitale tekster, hvilket som regel ikke er tilfældet for sprog som har relativt få talere. Jo mindre tekstmængde som systemet trænes på, desto lavere bliver kvaliteten på oversættelsen. Når man oversætter mellem tæt beslægtede sprog, kan direkte udskiftning af ordene være en fornuftig metode. I de sidste år er Apertium-plattformen som er udviklet ved Universitetet i Alicante i Spanien blevet brugt for at bygge maskinoversættelsessystemer mellem mange beslægtede sprog, blandt andet mellem svensk og dansk, og mellem norsk bokmål og nynorsk.

Om den regelbaserede metode giver succes, afhænger helt af om der er store ordbøger med morfologisk, syntaktisk og semantisk information til disposition. Det ville være helt fantastisk for udvikling af maskinoversættelse mellem nordiske sprog at have adgang til en database som ISLEX, med oversættelser eller lænker mellem fem sprog. Dette vil særlig være af nytte for maskinoversættelse mellem islandsk og andre sprog. Der eksisterer langt flere digitale ordbøger mellem dansk, norsk og svensk og andre sprog end mellem islandsk og andre sprog. Derfor kunne man tænke sig at bruge en af de fire andre sprog som en slags bro mellem islandsk og fx finsk.

En anden vigtig nytte man kunne have af ISLEX er i tværsproglig søgning, hvor man indtaster et søgeord på et sprog og får søgeresultater på flere sprog. Det er helt klart at sådanne søgemuligheder ville være vældig nyttige for mange i Norden som godt kan læse søgeresultater på hinandens sprog selvom de ikke ville kunne bruge søgetermer på andre nordiske sprog end deres eget. Søgningssystemet kan også muligvis sluttes til et maskinoversættelsessystem så at søgeresultaterne bliver automatisk oversat til søgningssproget.

For nogle år siden satte Nordisk ministerråd penge til projektet Nordisk Netordbog og i forbindelse med det satte Nordplus Sprog penge til projektet Tværsproglig søgning på nordiske sprog. Desværre kom der ikke ret meget ud af disse projekter, især fordi det viste sig at der var mangel på nødvendige resurser for flere af de nordiske sprog. Man kan ikke udvikle tværsproglig søgning uden at have flersproglige ordlister på de relevante sprog med forbindelser mellem beslægtede ord. Men med ordlister fra ISLEX ville det være muligt at udvikle et meget godt søgningssystem mellem de fem deltagende sprog. Det er også meget vigtigt at færøsk nu også kommer til at blive en del af ISLEX. Det vil muliggøre udarbejdelse af mange slags sprogteknologiske resurser og applikationer for færøsk.

Maskinoversættelse og tværsproglig søgning er kun to af flere områder indenfor sprogteknologien hvor ISLEX kan være af nytte. Det er klart at den også vil være særlig vigtig for computer-assisted language learning, CALL, og dermed for internordisk sprogforståelse. Hvis man kigger længere ind i fremtiden kan man forestille sig avancerede flersproglige dialogsystemer, direkte real-time oversættelse af samtaler hvor deltagerne taler

hver sit sprog, osv. Dette er vistnok stadig fremtidsmusik, men fremtiden er måske ikke så langt væk som vi tror og nærmer sig måske hurtigere end vi havde ventet.

Men hvorfor bryder vi os egentlig om sprogteknologi? META-NETs mission siger at sprogteknologien vil muliggøre kommunikation og samarbejde på tværs af sprog, sikre brugere af alle sprog lige adgang til information og viden, og videreudvikle den netværksbaserede informationsteknologi. Men den vigtigste grund er måske at sprogteknologien spiller en afgørende rolle i bestræbelserne på at beskytte og fremme vores nordiske sprog. Vi har desværre en god grund til at tro at vores sprog er i fare for tiden. I sammenligning med sprog som engelsk, fransk, tysk, og spansk er alle de nordiske sprog vældig små. Nu for tiden tales der ofte om underforsynede sprog eller "less-resourced languages", dvs. sprog som ikke har store sproglige resurser som gør det muligt at udvikle sprogteknologiske værktøjer for disse sprog.

Man kan godt sige at alle de nordiske sprog er underforsynede i sprogteknologisk henseende. Sprogteknologien spiller allerede en stor rolle i vores daglige liv, og dens rolle bliver stadigvæk større. Det er klart at hvis der ikke udvikles samme slags sprogteknologiske værktøjer for vores sprog som for de store sprog med mange talere så er vores sprog dømt til at tabe i kampen mod engelsk. Dette sker ikke i år, og ikke i løbet af de næste par årtier, men hvis vi ikke reagerer på faren nu, så kan det blive for sent. Kun ved at udvikle flersproglige resurser og sprogteknologi som baseres på dem kan vi sikre vores sprog en plads i fremtidens informationssamfund og derved bevare Europas kulturelle og sproglige diversitet.

I sidste instans er det de sprogteknologiske løsninger der skal danne bro mellem Europas sprog. META-NET har derfor taget initiative for udarbejdelse af 30 sprogteknologirapporter, en for hvert sprog. Disse rapporter er næsten færdige og bliver udgivet i begyndelsen af næste år, hver rapport både på engelsk og det vedkommende sprog. Alle disse sprograpporter har den samme struktur. De begynder med et kapitel som hedder "En risiko for sproget og en udfordring for sprogteknologi" og er fælles for alle rapporterne. Derefter kommer et kapitel om vedkommende sprog i det europæiske informationssamfund, sprogets særlige karakteristika, den seneste udvikling i sproget, sprogsamfundet, sprogpleje osv. Det tredje og sidste kapitel handler om sprogteknologisk støtte til sproget med hensyn til sprogkontrol, Internetsøgning, taleteknologi, maskinoversættelse osv.

I sprograpportene er de 30 europæiske sprog som behandles i rapportserien klassificeret ud fra fire forskellige kriterier: Sprogteknologisk støtte for taleteknologi, for maskinoversættelse, for tekstanalyse, og for sprogresurser. Indenfor hvert kriterium blev sprogene opdelt i fem forskellige grupper eller klynger. Sprogene i den første klynge har fuldstændig sprogteknologisk støtte, sprogene i den anden klynge har god støtte, sprogene i den tredje klynge har mellemgod støtte, sprogene i den fjerde klynge har ringe støtte, og sprogene i den femte og sidste klynge har næsten ingen eller slet ingen støtte. Det viser sig at den første klynge altid er tom – der eksisterer ikke noget sprog som man kan hævde at har fuldstændig sprogteknologisk støtte. Den anden klynge består i alle tilfælde kun af et enkelt sprog – engelsk.

I det store og hele taget er der en god korrelation mellem antal talere af hvert sprog og sprogets placering i klynger. Det er derfor ikke uventet at sammenligning tværs på sprog viser at de nordiske sprogs situation ikke er særlig god. Islandsk sidder i den nederste klynge med hensyn til alle fire kriterier. Det samme gælder for tre andre europæiske sprog som har relativt få talere – lettisk, litauisk og maltesisk. Dansk og norsk sidder i fjerde klynge med hensyn til tre af kriterierne men i den femte og sidste med hensyn til maskinoversættelse; det samme går for svensk med undtagelse af at svensk er i tredje klynge med hensyn til resurser. Det er altså klart at alle disse sprog virkelig behøver et mere solidt grundlag for deres sprogteknologi. ISLEX kan blive en vigtig del af dette grundlag.

Som jeg sagde før deltager repræsentanter for alle de nordiske og baltiske lande i META-NORD. Tilfældigvis er Universitetet i Bergen den norske repræsentant og Göteborgs Universitet den svenske – altså de samme institutioner som deltager i ISLEX, selvom der er lidt forskellige personer involveret. Den danske repræsentant er Center for Sprogteknologi ved Københavns Universitet, som arbejder tæt sammen med DSL, den danske deltager i ISLEX. Derfor skulle det ikke blive så svært at oprette et godt samarbejde mellem ISLEX-gruppen og META-NORD, så at ISLEX resurserne indgår i META-NORDs resurser, enten i sin helhed eller en begrænset del af dem. Disse resurser er så værdifulde for nordisk sprogteknologi og for bevarelsen af de nordiske sprog i den digitale verden at det er vores pligt at udnytte dem så godt som muligt.

Det er et velkendt faktum at det er svært at få penge til tidsbegrænsede projekter, men det er næsten umuligt at få penge til at vedligeholde dem. Men som alle sproglige resurser må ISLEX stadig revideres. Dette gælder især nu i det en og tyvende århundrede, hvor ny teknologi, nye produkter, nye begreber, og nye ord dukker op i større mængder end nogensinde før. Derfor er det nødvendigt at tilføje nye ord og nye betydninger, forbedre den semantiske beskrivelse, fjerne forældede ord og betydninger, revidere oversættelser, osv. Hvis dette ikke bliver gjort tager det ikke lang tid før resurserne bliver forældede og ubrugelige indenfor mange områder, især sprogteknologien. Desværre ser det ikke ud til at arbejdet med ISLEX her i Island kommer til at fortsætte i nær fremtid. Det er uhyre vigtigt at man søger midler til forbedring, revidering og vedligehold af ISLEX så den fortsætter med at være den værdifulde resurse som den er nu.

Til slut vil jeg takke for at få lejlighed til at tale ved denne glædelige begivenhed og gentage mine lykønskninger til alle som har været med til at lave ISLEX, og faktisk til alle i Norden, med denne flotte ordbog.