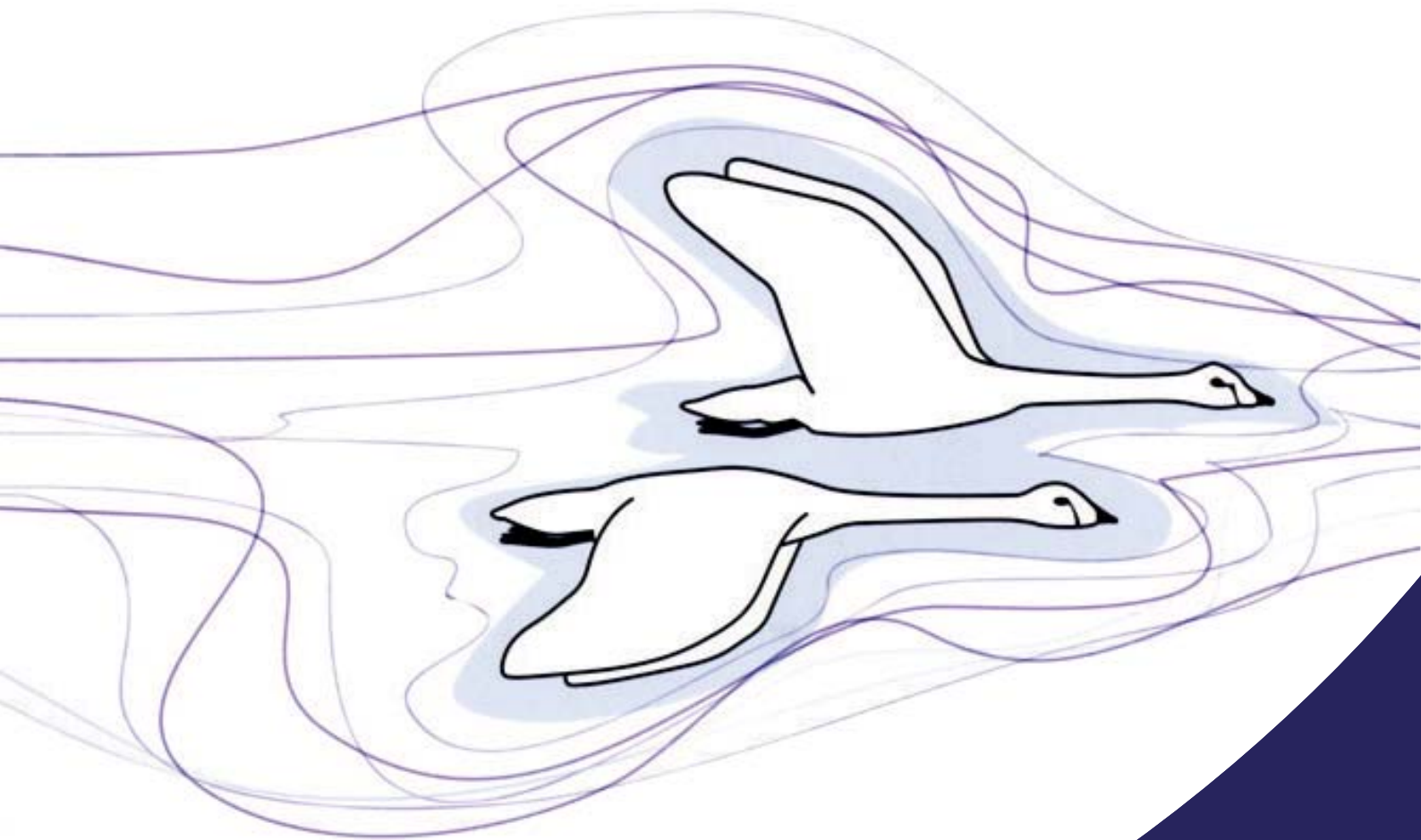


Språk i Norden 2014

Tema: De nordiske språkenes framtid i
informasjonsteknologien



Språk i Norden 2014

Tema: De nordiske språkenes framtid i informasjonsteknologien

Språk i Norden 2014

Tema: De nordiske språkenes framtid i informasjonsteknologien



norden

Nettverket for
språknemndene i Norden

Utgitt av Nettverket for språknemndene i Norden

Copyright: Forfatterne og Nettverket for språknemndene i Norden
ISSN: 2246-1701

Indholdsfortegnelse

Forord	7
Norsk språk i eit taleteknologisk perspektiv <i>Knut Kvale</i>	9
Tale er Guld	
Om talegenkendelse i de danske kommuners tjeneste <i>Peter Juel Henriksen</i>	24
Språkteknologi och språkresurser för språken i Sverige: En statusrapport	33
<i>Lars Borin och Rickard Domeij</i>	
Islex: eit skuleeksempel på godt språkteknologisk samarbeid i Norden	48
<i>Margunn Rauset</i>	
Sprogværktøjet Frasar.net	
Om fraser og fraseindlæring anskuet kontrastivt <i>Auður Hauksdóttir</i>	68
Det islandske ordklasseopmærkede korpus (MÍM) <i>Sigrún Helgadóttir</i>	83
Status for samisk språkteknologi <i>Sjur Nørstebø Moshagen</i>	95
Selvfølgelig snakker vaskemaskiner grønlandsk i fremtiden <i>Per Langgård</i>	110
FIN-CLARIN – en humanistisk forskningsinfrastruktur med betoning på språk	126
<i>Krister Lindén</i>	

Sprogteknologi og sproginstitutioner. Hvilken rolle kan sprognævnene spille i forhold til sprogteknologi?	133
<i>Sabine Kirchmeier-Andersen</i>	
Sprog i fokus: Islandsk	151
<i>Jóhannes B. Sigtryggsson</i>	
Det nordiske sprogsamarbejde i 2013.....	159
<i>Anne Kjærsgaard og Sabine Kircheimer-Andersen</i>	
Det nordiske sprogsamarbejde og Nordisk Sprogkoordination 2014–2018	173
<i>Hulda Zober Holm</i>	
Ny sproglitteratur 2013	181
<i>Danmark</i>	
<i>Finland</i>	
<i>Færøerne</i>	
<i>Island</i>	
<i>Norge</i>	
<i>Sápmi</i>	
<i>Sverige</i>	
Adresser språknemndene 2014.....	226

Forord

Språk i Norden er en årbok for språknemndene i Norden, og publiseres nå bare elektronisk på nettsidene til Nettverket for språknemndene. Lenken er sprogkoordinationen.org. Årboka har kommet ut og hatt samme navn fra 1970. Den blir redigert av de nordiske sekretærene i språknemndene i Danmark, Finland, Island, Norge og Sverige. Hovedredaktøren for dette nummeret er fra Norge.

De fleste artiklene i årboka er basert på innlegg som ble holdt på det årlige nordiske språkmøtet. I tillegg har redaksjonen bedt om noen artikler som utfyller årbokas tema.

Det 60. nordiske språkmøtet ble arrangert i Akureyri 28.–29. august 2013 av Nordisk språkkoordinasjon i samarbeid med Stofnun Árna Magnússonar í íslenskum fræðum. Temaet for årets språkmøte var de nordiske språkernes framtid i informasjonsteknologien. Noen innlegg presenterte status for språkteknologien for sitt språk, mens andre presenterte konkrete prosjekter og resultater. Noen innlegg presenterte arbeidet med tilrettelegging av språkressurser ("rådata" for å lage produkter og tjenester), og andre presenterte hvilke erfaringer de hadde med innføring av språkteknologiske produkter og tjenester.

Knut Kvale gir et oversyn over historikk og nåværende status for norsk taleteknologi. Peter Juel Henriksen presenterer erfaringer med innføring og bruk av talegjenkjenning i danske kommuner. Sverige var tidlig ute med språkteknologisk forskning og innsamling av språkressurser, og Lars Borin og Rickard Domeij gir oss en statusrapport om hvor langt de har kommet pr. 2013.

Språkteknologien gir muligheter for å kombinere tekst og tale i leksikon, ordbøker osv. Margunn Rauset viser hvordan det nordiske ordbokverket ISLEX har utnyttet språkteknologien til beste for brukerne ved å kombinere ordinære ordbokartikler med tale og bilde. Auður Hauksdóttir går et skritt videre ved å presentere språkverktøyet frasar.net, som er en type verktøy de fleste vil ha nytte av når de skal lære seg et nytt språk. Dette verktøyet kan brukes mellom dansk og islandsk.

Det islandske ordklasseoppmerkede korpuset som Sigrún Helgadóttir

skriver om, er en type språkressurs som er viktig underlag for utvikling av språkteknologiske produkter og tjenester.

Sjur Nørstebø Moshagen og Per Langgård gir status for henholdsvis samisk (Moshagen) og grønlandsk språkteknologi (Langgård). Krister Lindén presenterer den finske delen av det europeiske infrastrukturprosjektet CLARIN (Common Language Resources and Technology Infrastructure). CLARIN har som mål å gjøre digitale språkressurser tilgjengelige for forskere innen humaniora og samfunnsvitenskap.

Sabine Kirchmeier-Andersen skriver om hvilken rolle språknemndene kan, og bør, spille når det gjelder utvikling av språkteknologien, og hvor viktig det er at nettopp språknemndene deltar i dette arbeidet for å sikre produkter og tjenester på våre respektive språk.

I flere numre har vi hatt en artikkel som blir kalt "Språk i fokus", og i dette nummeret skriver Jóhannes B. Sigtryggsson om islandsk.

I 2013 hadde Dansk Sprognævn oppgaven som nordisk språkkoordinator etter at Bodil Aurstad tok permisjon fra stillingen for å tiltre en annen stilling. Anne Kjærgaard og Sabine Kirchmeier-Andersen oppsummerer det nordiske språksamarbeidet i 2013.

Hulda Zober Holm tiltrådte som nordisk språkkoordinator mot slutten av 2013, og redaksjonen har bedt henne presentere Nordisk ministerråds rammer og planer for det nordiske språksamarbeidet og Nordisk språkkoordinasjon for perioden 2014 til 2018.

Som alltid publiseres en bibliografi over ny språklitteratur fra de nordiske landene, og i år har vi utvidet med en oversikt fra Færøyene og en fra det samiske språkområdet.

Redaksjonen for Språk i Norden 2014 har bestått av: Jóhannes B. Sigtryggsson (Stofnum Árna Magnússonar í íslenskum fræðum), Pia Jarvad (Dansk Sprognævn), Rickard Domeij (Språkrådet, Sverige), Torbjørg Breivik (Språkrådet, Norge) og Anna Maria Gustafsson (Institutet för de inhemska språken (Focis)).

Oslo i juli 2014
Torbjørg Breivik
hovedredaktør

Norsk språk i eit taleteknologisk perspektiv

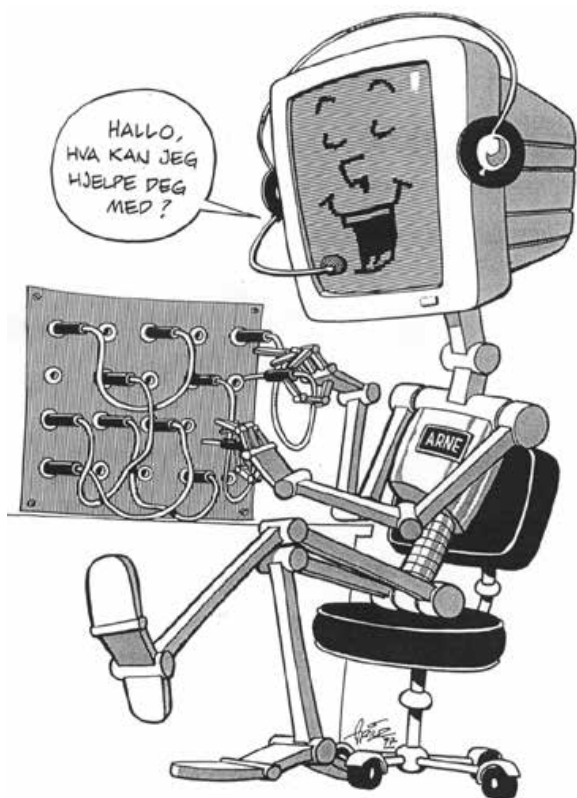
Knut Kvale

Denne artikkelen skildrar fyrst korleis talesyntese og automatisk talegjenkjenning er bygd opp, og drøftar deretter kva for verknader slik taleteknologi kan få for norsk språk. Automatisk talegjenkjenning for generell diktering («tale-til-tekst») er den delen av taleteknologien som kan påverke språket mest. Men programvare for generell norsk diktering finst ikkje enno. Ein av årsakene til dette har vore manglande språkteknologisk infrastruktur i form av ein norsk språkbank, som vart etablert fyrst i 2010. Håpet er at me får samla inn og systematisert språkdata frå heile landet slik at norske talegjenkjenningar kan lærast opp til å kunne tolke dei ulike uttalevariantane i norsk.

Innleiing

Tale er den mest naturlege måten menneska kommuniserer på. Alle er «ekspertar» i å snakke og lytte. Me formidlar tankar og kjensler via eit talt språk utan hjelpemiddel og med liten fysisk innsats. Me kan rusle omkring medan me snakkar, og både augo og hendene er frie til å gjera andre ting. Men når me skal kommunisere med ei datamaskin (PC, nettbrett og andre mobile terminalar), må me halde oss i ro og bruke tastatur og mus eller trykke på skjermen. I mange samanhengar er det difor ynskjeleg å kunne tale til datamaskina (talegjenkjenning) og få talt respons (talesyntese).

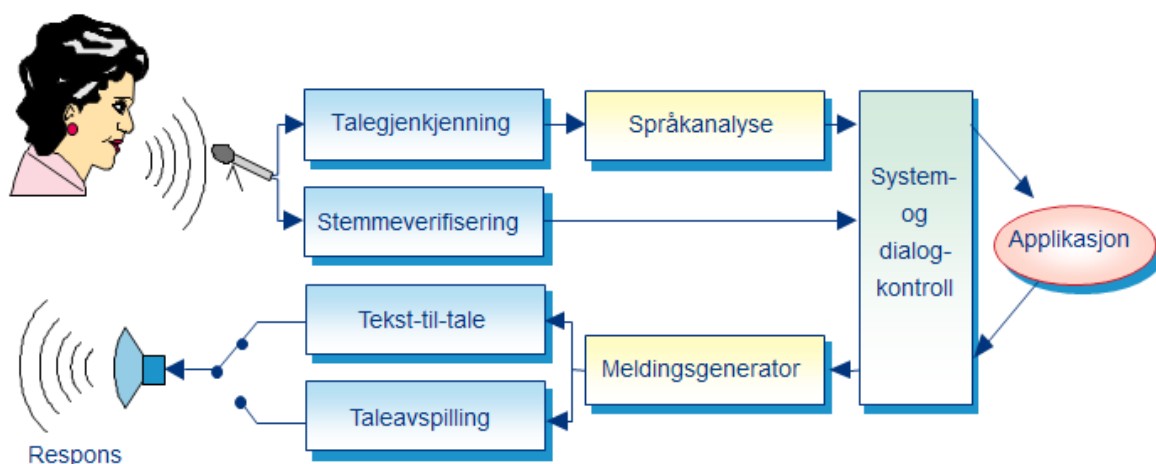
Heilt sidan dei fyrste datamaskinene vart laga, har det vore forska på teknikkar som gjer det mogleg å snakke med datamaskiner, og no har taleteknologien for ein del språk kome så langt at det er mogleg å diktere brev og andre dokument på PC-en, nettbrettet eller mobiltelefonen, eller ein kan spørja om alt frå togruter til aksjekursar. Me har fått *taleføre* datamaskiner som kan hjelpe oss, som illustrert i Figur 1.



Figur 1 Taleteknologi: Ei talefør og hjelpsam datamaskin! (Teikninga sto fyrste gong på trykk i Aftenposten 1. juni 1997, side 30. Teiknaren Arild Midthun har gjeve Språk i Norden løyve til å bruke teikninga).

Kva er taleteknologi?

For å kunne diskutere korleis taleteknologi kan påverke bruken av norsk språk, er det viktig å ha ei felles forståing av kva taleteknologi er. Med taleteknologi meiner me teknikkar for talekoding (talekompresjon), talesyntese, talegjenkjenning og stemmeverifisering. Figur 2 viser typiske komponentar i ei datamaskin som kan snakke og forstå tale. Me seier at datamaskina har eit talebasert brukargrensesnitt.



Figur 2 Talebasert brukargrensesnitt: Datamaskin som forstår tale og som gjev talt respons

Når me snakkar til ei datamaskin med eit talebasert brukargrensesnitt, blir trykkbølgjene i lufta fanga opp av ein mikrofon og klassifiserte som språklydar og ord ved hjelp av automatisk talegjenkjenning. Viss maskina skal kunne gjenkjenne og utføre meir enn ein kommando om gongen, må ein språkanalyse finne grammatisk rette og meiningsfulle setningar frå dei gjenkjende orda. System- og dialog-kontrollen sjekkar om dei gjenkjende orda og setningane gjev eit eintydig oppslag i databasen, og spør eventuelt om att for å sikre rett søk. Responsen frå systemet kan enten spelast av ved å setje saman innlesne frasar (*taleavspeling*), eller det må genererast syntetisk tale direkte frå teksten som blir henta ut or databasen (*tekst-til-tale*).

Figur 2 viser også ein boks med *stemmeverifisering*. Dette er ein teknikk som utnyttar at alle menneske er unike og snakkar forskjellig. Stemmeverifisering kan brukast som ekstra sikkerheit i tilgangskontroll, spesielt ved oppkopling til personlege databasar via telenettet.

Teknikkane for talekoding og stemmeverifisering er ikkje så språkavhengige og vil difor ikkje påverke språkutviklinga. Men det vil kanskje *talesyntese* og *talegjenkjenning* gjere i framtida. Difor ser me nærmare på desse teknikkane nedanfor.

Talesyntese

Talesyntese betyr at datamaskina les opp ein tekst. Me skil mellom (1) talesynteser med eit lite ordforråd og faste meldingar, og (2) talesyntesar som kan lese opp vilkårlege, ukjende tekstar; eit såkalla «tekst-til-tale»-system.

Samanskøyting av faste meldingar

Ein del opplysningstenester nyttar berre nokre få ord og frasar, slik som f.eks.:

'Abonntenen har fått nytt nummer', 'Det nye nummeret er', '22' '44'
'66' '88'.

For slike tenester kan det løne seg å la ein person lese opp alle orda og dei faste frasane og så lagre dette. Opptaket må merkast slik at den syntetiske talen lett kan genererast ved å kombinere dei lagre orda og frasane. Syntetisk tale basert på innlesne setningar og ord er lett å forstå og høyrest ganske naturleg ut. Men slike system er ikkje fleksible. Viss ein ynskjer å lage ei ny melding eller å utvide ordforrådet, må det gjerast nye opptak med same stemma under same opptaksvilkår. Dette avgrensar mengda informasjon som kan leggjast inn og kor ofte informasjonen kan oppdaterast.

Tekst-til-tale

Ein *tekst-til-tale*-syntese kan lese opp ein vilkårleg tekst. Tekst-til-tale-syntesen er difor svært *fleksibel* og eignar seg spesielt godt for opplesing av ukjent tekst i elektronisk form, slik som elektronisk post, eller for opplesing av tekstlege meldingar som blir endra ofte, slik som nyheiter, vêr- og føremeldingar eller trafikkinformasjon som ligg på nettet. Me kan ringe til slike databasar og få den skriftlege informasjonen lese opp over telefon. Her var Norge tidleg ute: I mai 1998 lanserte Telenor Nextel opplesing av elektronisk post over telefon. Dette var ei av dei fyrste kommersielle tenestene i sitt slag i Europa.

Tekst-til-tale-syntese kan òg nyttast på PC-en, nettbrettet eller mobilen som lesehjelp for blinde og svaksynte eller for folk med ulike lesevanskar. Folk med talevanskar kan bruke tekst-til-tale-syntese som si eiga stemme, som ein "taleprotese".

Det finst mange ulike måtar ein kan syntetisere tale frå tekst. *Figur 3* viser nokre typiske funksjonar i eit tekst-til-tale-system. Teksten må vere i elektronisk form, og kan i prinsippet hentast frå databasar, lesast optisk

med ein skannar, merkast av i eit vindauge på skjermen eller skrivast inn direkte. Systemet må kunne handtere mange typar tekst, slik som forkortingar, tal, datoar, tidspunkt, formlar og tabellar. Dette kallar me *tekstnormalisering*, og det er slettes ikkje ei beintfram oppgåve i norsk.

Eitt eksempel er uttalereglar for tal: Skal talet 1996 lesast som «nitten nitti seks», som «eitt tusen ni hundre og nitti seks», eller som «ein ni ni seks»? Kva med talet "11.10"? Skal det lesast som «elleve komma ti», som «elleve og ein tidel», som klokkeslettet «ti minutt over elleve», eller som datoen «ellefte oktober»? Brøken $\frac{3}{4}$ skal som oftast lesast som «tre fjerdedelar», men i nokre høve er det ein dato, og då skal det lestast som «tredje i fjerde» eller «tredje april».

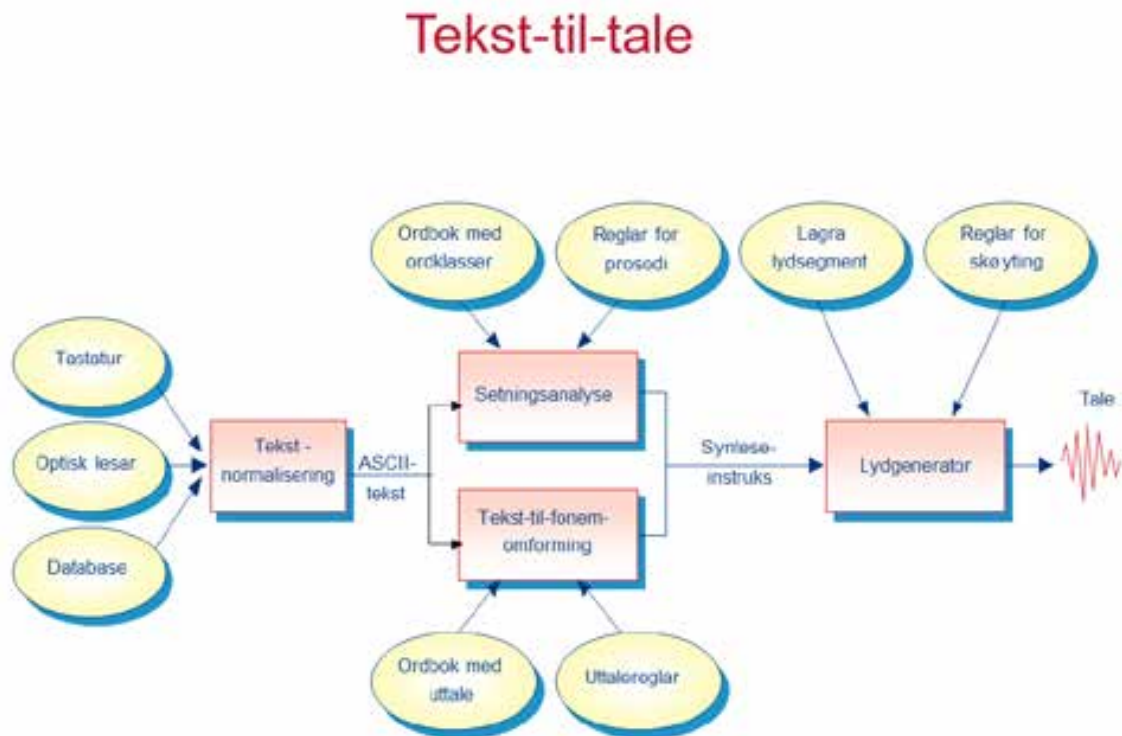
I tekstnormaliseringa kjem ein inn på grunnleggjande spørsmål som å definere kva ei setning er. For inndeling av tekst i setningar, seier me at ei setning sluttar ved punktum etterfølgd av mellomrom og stor bokstav. Regelen ser kanskje rimeleg ut, men vil få problem med setningar som «kvinnar var dr. E. Hals». Når det gjeld e-post og nettadresser, må ein bestemme om punktumteiknet skal lesast opp, og korleis det i så fall skal uttalast: som "punktum" eller som "dått"? Talet 11 skal uttalast «elleve», mens «11.» skal uttalast «ellefte». Men kva skal vi gjere når talet står i slutten av ei setning, rett før punktum, som i «Det kom berre 11.»?

Ein må òg lage reglar for korleis forkortingar skal handterast slik at det blir mest mogleg likt vanleg uttale i språket. Her er det fleire alternativ i norsk: Forsvarsalliansen NATO er ei forkorting for «North Atlantic Treaty Organization» og blir uttala som eitt ord, «nato». Forkortinga «NSB», som blir brukt om Noregs Statsbanar, blir uttala som staving: «N-S-B», medan forkortinga «m.a.o.» skal uttalast som «med andre ord».

For å gjere om ein tekst til lydskrift (*tekst-til-fonem-omforming*), slår synteseprogrammet opp i ei ordbok der kvart ord står saman med lydskrifta av ei vanleg uttale av ordet (slik ordet blir uttala isolert i ein valt dialekt). I tillegg trengst uttale av eigennamn, adresser, stadnamn, osb. Viss eit ord ikkje finst i ordboka, må ein ta i bruk generelle uttalereglar for å gjere om teksten til lydskrift. Uttalereglane kan for eksempel vise at bokstavgruppene «skj», «ski», «sj» og «rs» skal uttalast på same måten. I norsk går det an å lage svært lange ord ved å setje saman mindre ord, som for eksempel «gutteinternatskolebestyrar» eller endå verre: «fylkestrafikk-sikkerhetsutvalgssekretariatslederfunksjonene». Slike samansette ord må systemet prøve å dele opp i mindre ord ved å søkje i ordboka etter grunnforma av dei einskilde orda.

I motsetnad til oss menneske skjønner ikkje datamaskina kva ho sjølv seier (!). Systemet må difor estimere kva for ord det skal leggjast trykk på og korleis setningsmelodien bør vera. I setningsanalysen slår programmet opp i ei ordbok der ordklassa for kvart ord står, slik at for eksempel verb, adjektiv og substantiv blir identifisert. Dette hjelper til med å lage rett uttale og tonelag i setningar av typen «han rota på rommet» og «rota på treet». *Setningsanalysen* blir kombinert med reglar for kva for ordklassar som skal framheva i setninga og ein modell for setningsmelodi. Viss ein har tilgang til mykje taleopptak som er prosodisk merkt, kan ein utvikle meir avanserte prosodiske modellar og dermed kunne syntetisere meir naturleg tale.

Ut frå *tekst-til-fonem-omforminga* og *setningsanalysen* kjem ein synteseinstruks som inneheld ei rekkje med fonem, samt varigheit, intensitet og grunntone for kvart fonem. Så blir dei *lagra lydsegmenta* henta inn og modifisert med omsyn på synteseinstruksen, før dei blir skøyte saman til lengre lydsekvensar og sende til *lydgeneratoren* – og ut kjem den syntetiske talen.



Figur 3 Blokkskjema av eit tekst-til-tale-system.

Automatisk talegjenkjenning

Med *automatisk talegjenkjenning* meiner me at datamaskina enten skriv det me seier ("*tale-til-tekst*"), eller at ho gjer det me munnleg ber ho om ("*kommandostyring*"). Dette gjer at me kan kommunisere med ei datamaskin utan å bruke tastatur eller mus, noko som er spesielt nyttig for folk med nedsett funksjonsevne i armar og fingrar.

Med automatisk talegjenkjenning treng me ikkje lenger å vere fysisk til stades nær ei datamaskin for å styre ho, fordi me kan gje munnlege kommandoar til datamaskina tvers over rommet eller via telefon. Sidan me kan snakke med datamaskina, blir synet og hendene våre frigjorde til andre formål, noko f.eks. legar utnyttar når dei analyserer mange røntgenbilete og dikterer pasientjournalane samstundes.

Me skal no sjå nærmare på problema i automatisk talegjenkjenning, og korleis desse problema blir løyste. Det aller største problemet for automatisk talegjenkjenning er all variasjonen i talesignala som skal tolkast. Det er mykje lettare å lage god automatisk talegjenkjenning i ei personleg datamaskin der gjenkjennaren kan tilpassast kvar einskild brukar, enn å gjenkjenne kva ulike innringarar seier til ei offentleg telefonteneste.

Talegjenkjenning for PC – talartilpassa gjenkjenning

For nokre språk kan ein no kjøpe dikteringsprogram for PC for under tusen kroner. Slike program er utstyrte med programvare som prøver å lære seg den spesielle måten kvar einskild snakkar på. Dette kallast adaptasjon eller talartilpassa gjenkjenning. Programma lèt brukaren diktere og redigere dokument heilt utan tastatur og mus. Ein kan snakke utan pause mellom orda i normalt taletempo, det vil seie mellom 100 og 160 ord i minuttet. I tillegg til diktering kan brukaren redigere teksten ved å seie "uthev ordet", "flytt avsnittet til slutten av dokumentet" osv. Tekstredigering med talte kommandoar gjer at brukarane kan kome til å aktivere verktøy og funksjonar i tekstbehandlaren som dei elles aldri ville ha oppdaga at fanst!

Talegjenkjenning over telefon – talaruavhengig gjenkjenning

Talegjenkjenning i publikumstenester som skal kunne forstå alle folk, blir kalla talaruavhengig gjenkjenning, og er mykje vanskelegare enn talartilpassa gjenkjenning. I slike publikumstenester må systemet prøve å avgrense kva det er lov å snakke om. Dette blir løyst ved at systemet styrer dialogen og stiller innringaren konkrete spørsmål. I teletenester er difor

utforminga av dialogen svært viktig. Kommersielle system er ofte bygd opp rundt ein skjemastruktur der innringaren i praksis fyller ut dei blanke felt i dette skjemaet ved å svare på spørsmål frå systemet. Gjenkjenningarane kan i dag skilje ut viktige ord frå høflegsfraser og fyllord, slik at når ein innringar seier ”kan eg få æh togopplysning for Dovrebanen, takk”, blir orda ”togopplysning” og ”Dovrebanen” gjenkjent, og rett søk blir utført i databasen.

Variasjon i talesignala: Det store problemet for talegjenkjenning

Sjølv om teknologien for automatisk talegjenkjenning har kome langt, er det lenge til me kan snakke med ei datamaskin på same måten som me snakkar med kvarandre. Det største problemet for ein automatisk talegjenkjenner er all *variasjonen* i talesignala. Kvart menneske er unikt, og alle snakkar forskjellig. Ved nøye analyse av bølgeformene til lydar ser me at ein person ikkje greier å gjenta nøyaktig same lyden likt to gonger etter kvarandre. Så for datamaskina er kvar lyduttale unik. Dessutan blir realiseringa av språklydane påverka av samanhengen dei blir uttala i. I ei teneste med talebasert grensesnitt for styring av datamaskin over telefon vil det i tillegg vere mange slags bakgrunnsstøy, varierende kvalitet på transmisjonsmedia, ulik kvalitet på mikrofonane, og folk vil variere avstanden til mikrofonane. Den automatiske talegjenkjenneren må difor kunne redusere verknaden av alt som påverkar talesignalet og skilje talesignalet frå støyen.

Teknikkar i talegjenkjenning

Før talegjenkjenneren kan brukast til noko som helst, må han «trenast opp» til å gjenkjenne det vi ynskjer. Vi må bestemme om han skal gjenkjenne setningar, ord eller mindre einingar, ofte kalla «delord». Delord kan vere fonem eller stavingar. Det er svært mange ord i eit språk, men relativt få delord; det er for eksempel omlag 50 fonem i norsk. Talegjenkjenning basert på delord gjer systemet meir fleksibelt. Eit nytt ord kan enkelt leggjast inn i lista av ord som kan gjenkjennast ved å skrive lydskrifta av ordet. Ein slepp altså å trene opp gjenkjenneren på nytt for kvart nytt ord som skal leggjast inn i ordforrådet til gjenkjenneren. Dette kallast gjenkjenning med *fleksibelt vokabular*.

Figur 4 viser eit blokkskjema av ein delordbasert talegjenkjenner. Fyrst blir det digitaliserte talesignalet filtrert og omforma til ein meir kompakt

representasjon. Talesignalet endrar seg relativt seint slik at eit tidsintervall på 10 til 30 millisekund kan representerast med ein eigenskapsvektor med 20 til 50 parametrar.

Ved *mønstergjenkjenninga* blir desse eigenskapsvektorane samanlikna med modellane for dei lagra delorda. Det delordet som liknar mest, blir gjenkjent. Sidan realiseringa av ulike fonem kan likne på kvarandre, bør det takast vare på dei kandidatane som likna nest mest også. På denne måten blir det generert eit nettverk av delord. Det er vanleg å bruke statistiske modellar for delorda. Parametrane i desse modellane blir estimerte («trente») på grunnlag av eit kjent talemateriale («treningssett»).

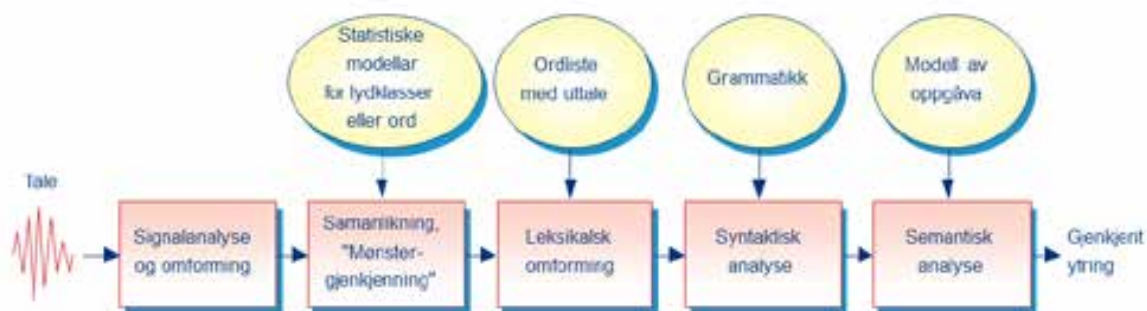
Ved *leksikalsk omforming* blir nettverket av gjenkjende delord sett saman til ord. Vokabularet til gjenkjennaren står i ei ordliste saman med lydskrift av forventta uttale av kvart ord (slik orda blir uttala isolert i ein vald dialekt). Ei slik ordliste bør kombinerast med reglar for koartikulasjonseffektar over ordgrenser. Ei statistisk ordliste inneheld sannsyn for dei ulike uttalevariantane av orda. Som oftast er det fleire moglege ordkandidatar. Den *syntaktiske analysen* kontrollerer om grammatikken i dei gjenkjende ordsekvensane er rette. Viss to ord har same uttale, men ulik skrivemåte, kan denne analysen hjelpe til med å velje det rette ordet.

Ein *deterministisk grammatikk* er bygd opp av reglar for kva ord som kan følgje etter kvarandre, eller han kan innehalde alle setningane som skal kunne gjenkjennast. Ein *statistisk grammatikk* inneheld sannsynet for at to ord følgjer etter kvarandre (kallast «bigram»), eller at tre ord kjem i rekkefølge (kallast «trigram»).

Viss datamaskina berre skal utføre enkle kommandoar, er syntaktisk analyse unødvendig.

Ei setning kan godt vere grammatisk rett, men likevel vere meningslaus. Med setninga «jenta var høy», meiner vi at jenta var lang, men «høy» kan òg tolkast som «tørt gras» eller at jenta var rusa på narkotika. For spesielle oppgåver som f.eks. reservering av billetter eller oppslag i databaser, kan det til ei viss grad modellerast kva som er meningsfullt (*semantisk analyse*).

Talegjenkjenning



Figur 4 Blokkskjema av eit automatisk talegjenkjenningssystem

For at det talebaserte brukargrensesnittet skal vere naturleg å bruke, må orda eller kommandoane som kan gjenkjennast, vere naturlege å bruke i den gjevne samanhengen. Sidan det er mange måtar å uttrykkje same meining på, må talegjenkjenneren ha mange synonym i vokabularet sitt.

Det er ikkje naturleg for oss å seie berre eitt ord eller éin kommando om gongen. Om systemet ber oss om å seie ein kommando, vil mange svare med ei heil setning, for eksempel slik: «Eg vil gjerne ha ... kommando ... takk». Talegjenkjenneren må då kunne overhøyre utanomsnakk og berre gjenkjenne den rette kommandoen. Denne teknikken kallar vi *ordspeiding*.

Viss brukaren ikkje ynskjer å høyre på ei lang utgreiing om moglege menyval, må brukaren kunne avbryte maskina og tale «i munnen» på ho. Systemet må difor kunne lytte og tale samtidig.

Utforming av dialogen er viktigast

Det er vanskeleg å samanlikne ytinga eller nøyaktigheita til ulike system for automatisk talegjenkjenning, fordi dei sjeldan blir testa på det same talematerialet eller på dei same oppgåvene. Kor godt ein automatisk talegjenkjenningar verkar er òg avhengig av kva for talemateriale han blir trent på. Det einaste som er sikkert, er at alle system for automatisk talegjenkjenning før eller seinare vil ta feil og misforstå det som er sagt. Det som er viktig for brukarane, er korleis dette systemet handterer gjenkjenningarfeil slik at brukarane likevel får utført tenestene dei ber om.

For ei bestemt teneste er det difor viktig å utforme ein god dialog slik at brukaren ikkje gjev opp med ein gong talegjenkjenningen gjer feil. Systemet må kunne spørje om att på ein naturleg måte og få stadfesta eller avkrefta ein hypotese om gjenkjent ord er rett. Viss det likevel blir feil, må det vere enkelt for brukaren å rette opp feilen ved å gå eit trinn tilbake i dialogen. Talegjenkjenningen tek difor vare på ord som liknar på det gjenkjende ordet, slik at systemet kan foreslå eit rimeleg alternativ når det tek feil. For publikumstenester i telenettet bør det også vere enkelt å bli sett over til ein kundebehandlar, for eksempel ved å seie «hjelp» eller «operatør» når som helst i dialogen.

Talebaserte brukargrensesnitt som har hatt suksess for publikumstenester i telenettet, er kjenneteikna ved at brukarane oppfattar tenestene som nyttige, dvs. at tenestene er betre enn liknande tenester eller at det ikkje finst alternativ. Tenestene er brukarvennlege ved at dei:

- *er naturleg å bruke*: Dei erstattar ofte tenester med taste-baserte menyar («for alternativ A, tast 1 ...»), og har intuitive namn på kommandoane som kan gjenkjennast.
- *er enkle å bruke*: Dei har innbydande grensesnitt og ein forståeleg dialog.
- *er nøyaktige*: Minst 98 % av orda blir rett gjenkjent.
- *har sanntidsrespons*: Brukaren får raskt svar på spørsmåla sine.
- *har gode dialogar*: Målretta dialogar som reduserer verknaden av gjenkjenningarfeil.

Ein god dialog bør styre samtala slik at kunden har få svaralternativ for kvart spørsmål, og kunden blir leia til å svare berre på det han blir spurd om. Dette avgrensar søkjerommet for talegjenkjennaren, og dermed blir risikoen for feil mindre.

Store mengder taleopptak trengst

Me har sett at ein statistisk basert talegjenkjennar må kunne modellere delorda, orda og språket:

- *Delorda*; dvs. modellere karakteristiske akustiske eigenskapar ved delorda og modellere variasjonen i realiseringa av dei.
- *Orda*; dvs. finne sannsynet for ulike uttalevariantar.
- *Språket*; dvs. modellere grammatikken ved å finne sannsynet for at ulike ord følgjer etter kvarandre.

For å kunne estimere parametrane i alle statistiske modellane må det gjerast opptak av mykje tale. Desse talesignala må merkast slik at det er lett å finne fram i talematerialet. Talarane må vere eit representativt utval personar med omsyn til kjønn, alder og dialektbakgrunn. I tillegg bør ein gjere opptak i same omgjevnader som ein har tenkt å bruke gjenkjennaren i. Eksempel: For å få eit realistisk treningsmateriale for ein talegjenkjennar som skal brukast til telefontenester, bør opptaka gjerast over telefonnettet med ulike typar telefonar og i ulike omgjevnader og støymiljø.

Utvikling av norsk taleteknologi har lenge lidd under mangelen av ein norsk språkbank. Store språkteknologiske forskingsprogram som for eksempel «KUNSTI- Kunnskapsutvikling for norsk språkteknologi» (2001–2006), hadde til føresetnad at relevante språkressursar for norsk fanst, eller skulle bli utvikla i form av ein norsk språkbank. Men dette skjedde ikkje, noko som blir oppsummert slik i «St.meld. nr. 35 (2007–2008) Mål og meining - Ein heilskapleg norsk språkpolitikk»:

«Men mangel på relevante språkressursar viste seg å vera eit problem for mange av prosjekta i KUNSTI. Til dels måtte det brukast tid og ressursar på å samla inn og utvikla nødvendig infrastruktur i form av språkressursar før ein kunne starta med sjølve forskingsarbeidet.»

...«Alt i alt vart resultatet i KUNSTI-programmet at ein hadde mindre ressursar å bruka til dei primære forskingsoppgåvene. Dette førte i sin tur at til uheldige justeringar av innretning og ambisjonsnivå i fleire av prosjekta. Samla sett måtte ambisjonane for heile programmet endrast. Bakgrunnen var at ein ved planlegging av programmet hadde lagt til grunn at arbeidet med å byggja opp språkbanken ville koma i gang samstundes med programmet»

Fyrst i 2010 vart Språkbanken etablert, sjå oppdatert informasjon på heimesidene:

<http://www.nb.no/Tilbud/Forske/Spraakbanken>

Taleteknologi og norsk språk

Domenetap

Det er viktig å vere klar over den trusselen om domenetap som norsk språk er utsett for, mellom anna innanfor taleteknologi i dag. Denne trusselen finn me fyrst og fremst på feltet automatisk talegjenkjenning, som er ressurskrevjande å utvikle.

Følgjande eksempel viser problemet: Eit talegjenkjenningsprogram for engelsk diktering på PC kostar eit par tusen kroner, men å «omsetje» programmet til norsk kjem kan hende på om lag 10–20 millionar kroner. Sidan me framleis ikkje kan tilby ei slik programvare på norsk, er det freistande å diktere e-post eller rapportar på engelsk (med norsk aksent) i staden for å skrive det, spesielt når ein veit at dei fleste norske og skandinaviske mottakarane forstår engelsk.

For å unngå at me i framtida må snakke engelsk til taleføre datamaskiner, er det viktig at Norge satsar meir på utvikling av norsk taleteknologi.

Normering

Vil så norsk talesyntese eller norsk talegjenkjenning ha normerende verknad for norsk uttale? Eit tekst-til-tale-system vil neppe påverke uttalen til lyttaren. Me er vande med at folk snakkar forskjellig, og difor forstår me det maskina seier utan at me sjølve endrar uttale. Men eit talegjenkjenningssystem som berre godtek ein spesiell uttale, for eksempel berre uttalen «sju» og ikkje «syv», kan ha innverknad på uttalen til dei som brukar dette systemet.

Ein skal vere klar over at det er produsentane av talegjenkjenningsutstyr som bestemmer kva for uttale datamaskina skal kunne gjenkjenne. Ved gjenkjenning av norske tal kan for eksempel gjenkjenneren lærast opp til å tolke berre ny teljemaåte («tjue-to»), berre gamal teljemaåte («to-og-tjue» eller «to-og-tyve»), eller kunne tolke både ny og gamal teljemaåte.

På same måten er det produsentane av talesyntese som bestemmer korleis datamaskina faktisk skal uttale orda og setningane. Produsentane kan for eksempel velje kva for dialekt som leggst til grunn for uttalen, dei kan bestemme kor trykket skal leggst i ord som «potet», om syntesen skal uttale tjukk l eller vanleg l i ord som «folk», og kva for r-uttale syntesen skal ha.

Idealet må vere å utvikle norsk taleteknologi som ikkje berre er basert på ein bestemt dialekt eller sosiolekt frå Oslo. Taleteknologien må kunne tilpassast måten folk faktisk snakkar. Prinsippet er at taleteknologien må tilpassast menneska – ikkje omvendt. Difor må me samle inn og systematisere tale frå heile landet, slik at den automatiske talegjenkjenneren kan lære seg både skarre-r og tungespiss-r, ulike trykkleggingar i ord, og uttalar med eller utan tjukk l.

Knut Kvale er seniorforskar ved Telenor Research og professor II ved Universitetet i Oslo, institutt for medier og kommunikasjon.

Summary

The paper describes speech technology and discusses the potential consequences this technology may have for the Norwegian language. Unfortunately, Norway has not yet got an automatic speech recognizer for general dictation. One of the reasons for this is the lack of spoken language resources. Automatic speech recognition (ASR) is based on statistical models, and to be able to estimate the parameters in these models a lot of transcribed speech data is needed. In 2010, “Språkbanken” - a language technology resource collection for Norwegian - was finally established. Hopefully the Språkbank will contribute to the development of ASR-systems that understand the variations in spoken Norwegian.

Tale er Guld

Om talegenkendelse i de danske kommuners tjeneste

Peter Juel Henriksen

Mange danske kommuner er parate til at indfase automatisk talegenkendelse, men er samtidig nervøse efter en lang række dårlige businesscases i den nærmere fortid. Der klages over høje licenspriser og lavt serviceniveau, den typiske virkning af et de facto monopol på leverandørsiden. Denne artikel beretter om en ny offentlig handlingsstrategi der har taget form i 2013. Nok er Danmark et lille sprogområde, men vores offentlige sektor er velorganiseret og købedygtig. Det seneste år har de danske kommuner organiseret sig omkring den taleteknologiske udfordring, man har i fællesskab formuleret en række spørgsmål og nye krav til leverandørerne og bevidst arbejdet frem mod et udbudsmateriale der åbner for fair konkurrence. I dette arbejde har også offentlige forskere, heriblandt denne artikels forfatter, deltaget som økonomisk neutrale rådgivere. Initiativet skal bane vej for den første rentable kontrakt på området - som vi håber at se i 2014 - en begivenhed som præcis kunne løsne den nuværende deadlock og åbne et milliardstort marked for taleteknologi i det offentliges tjeneste.

De danske hospitaler er erfarne brugere af talegenkendelse. De danske kommuner ville meget gerne være det også. Set fra en kommunalbestyrelses synspunkt har talegenkendelse en række indlysende fordele; den kan gøre sagsbehandleren mere produktiv, fjerne en forhadet skrivebyrde for de praktisk orienterede funktionærer, støtte de sprogligt udfordrede medarbejdere: ordblinde, nydanske, svagsynede, lavtuddannede. Med talegenkendelse på sin smartphone eller tablet kan hjemmehjælperen afslutte sin visitationsrapport undervejs til næste klient, og trafiktælleren føre sine observationer direkte ind i regnearket. Kort sagt, talegenkendelse er *lean*.

Baggrund

En lille snes kommuner - ud af Danmarks otteoghalvfems (98) - har sat penge i taleteknologiske produkter gennem de sidste fire-fem år. De har generelt været seriøst indstillet på at forberede deres personale på nye arbejdsrutiner hvor man dikterer frem for at taste; ingen har troet at fordelene skulle komme uden samvittighedsfuldt forarbejde. Alligevel har endnu ingen dansk kommune tjent sin investering hjem. I dag anvender kun ganske få kommunalt ansatte talegenkendelse på daglig basis, hvad der står i skarp kontrast til de titusindvis af læger i Skandinavien der benytter teknologien. Problemet har især ligget i tilpasningsfasen. De fleste kommuner giver udtryk for at deres leverandør lovede en høj service med uddannelse af brugere og tilpasning af talegenkendelsen *in situ*, men at virkeligheden blev en anden efter at kontrakten var indgået og leverandøren efter kort tid overlod driften til brugerne selv. En anden kilde til irritation er de meget lange perioder man må vente på gentrænede systemer (fx sprogmodeller og akustiske modeller optimeret til de aktuelle arbejdsforhold). Producenterne hævder, på deres side, at de ikke kan tjene penge på at levere talegenkendelse til den danske kommune i dag, og at man kun er gået ind i markedet af strategiske grunde.

På softwaresiden har de danske kommuner reelt kun haft én teknologi at vælge imellem, nemlig Nuance Communications. Om denne teknologi er formidlet og tilpasset af skandinaviske firmaer som KMD, IBM Denmark, Max Manus eller PDC/Dictus forandrer ikke det faktum at licensbetingelserne er dikteret fra USA. På grund af Nuance's de facto-monopol har kommunerne haft svært ved at lægge pres på deres direkte leverandører, da disses margin for konkurrence er uhyre smal og, som nævnt, styret af international kapital. At dømme efter Nuance's nyere PR satser koncernen i de kommende år på intelligente dialogløsninger rettet mod (primært) amerikanske firmaer og velhavende forbrugere, frem for at tilgodese de europæiske markeder og de mindre sprogområder. Denne strategi er for nyligt blevet accentueret da Nuance antog professor Ron Kaplan som øverste chef for sin forskningsafdeling, en verdenskendt forsker i formel grammatik og kunstig intelligens - og ikke i små sprog.

Nye initiativer - kommuner og universitet i samarbejde

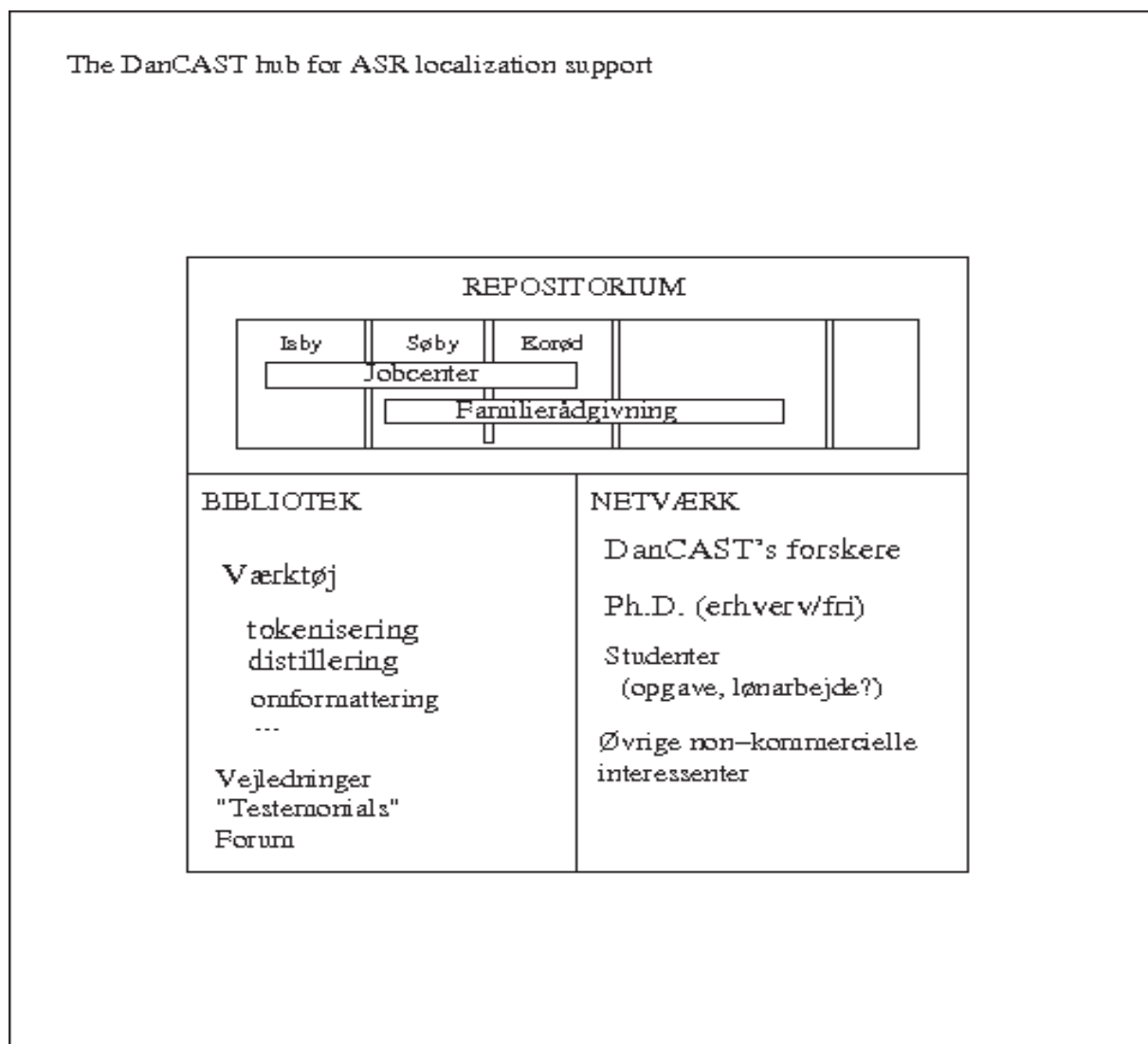
Hvordan bevæger de danske kommuner sig i dette minefelt? De kommunale budgetter er under pres i disse kriseår, og mange borgmestre skal vælge mellem fyringer og effektiviseringer. De spiller gerne bolden vi-

dere til deres IT-ansvarlige, med besked om at udrette et teknologisk mirakel som på én gang kan støtte de svageste ansatte og øge kommunens produktion. Principielt er dette jo ikke umuligt, men det kræver nye tekniske kompetencer hos de IT-ansvarlige, frigørelse fra gamle og tyngende industrikontrakter, vidensudveksling og genbrug af træningsmaterialer over kommunegrænser, *esprit mousquetaire*. Derfor er en lang række af de danske kommuner gået sammen i et frivilligt fællesskab, kaldet OS2, som mødes rundt i landet og udveksler erfaringer om indfasning af ny teknologi. OS2 har allerede haft succeser med yderst positive businesscases, og OS2-kredsen vokser derfor hastigt og tæller nu omkring halvdelen af de danske kommuner. Da taleteknologien er et særlig varmt emne, har de meste handlingsparate kommuner i OS2 dannet undergruppen *OS2Talk*, der gennem 2013 har taget initiativer til at bringe de centrale aktører sammen: De danske leverandører af taleteknologi (nuværende og potentielle), kommunerne selv, KOMBIT¹, Digitaliseringsstyrelsen – desuden kernebrugere i hospitalsvæsenet, Folketinget og Danmarks Radio.

De fleste af stormøderne har været afholdt på Copenhagen Business School (CBS) på Frederiksberg, organiseret af forskningscenteret DanCAST. Dette forskningscenter har i de seneste år samlet sine aktiviteter om to temaer, dels de særlige udfordringer som det danske (og de øvrige nordiske) sprog stiller til taleteknologien, og dels det anvendte perspektiv: hvordan dagens teknologiske løsninger tilpasses et mangearteret brugsmiljø, hvor stemmetyper, regiolekter, ordforråd, arbejdsvaner, støj og rumklang giver praktiske udfordringer som grundforskeren ofte abstraherer væk i sit laboratorium. For at gøre vejen så kort som muligt fra vores forskning til samfundets arbejdspladser har DanCAST åbnet et særligt forum som betjener kommuner med en dagsorden inden for taleteknologi. Dette forum kalder vi *The DanCAST hub for ASR Localization Support*.

Teknisk set er DanCASTs hub en stor server stillet til rådighed af CBS's IT-Afdeling, med garanti for vedligehold og opetid i alt fald fem år ud i fremtiden. Til serveren hører et lager med garanteret 500GB diskplads til hver af landets kommuner (på længere sigt er det IT-Afdelingens plan at lægge repositoret "i skyen").

1 KOMBIT er et selvejende (og derfor økonomisk uafhængigt) aktieselskab som rådgiver de danske kommuner om alle aspekter af IT-indkøb og anvendelse, lige fra udarbejdelsen af fair og åbne udbudsmaterialer, over efteruddannelse af personale, til juraen omkring misligholdelse af kontrakter.



DanCASTs HUB ("The DanCAST hub for ASR localization support")

Enhver dansk kommune med en ASR-dagsorden (automatic speech recognition) kan tegne en samarbejdsaftale med DanCAST. Aftalen forpligter CBS til at åbne og vedligeholde et repositorium hvortil kommunen har lov at uploade et stort materiale af tekster. Disse skal være udvalgt sådan at de støtter den træning som udviklingen af en ASR-løsning bygger på. Når de danske kommuner samler deres træningsdata på ét sted, giver det mulighed for at organisere dataene i adskillige fagområder, fx Jobcenter, Ældreområdet, Børn & Ungdom, uden at datamængden bliver for lille i de enkelte sektorer. Som datamængden vokser, bliver det lettere og lettere for den sidst tilkomne kommune at opnå et effektivt træningsgrundlag ved blot at tilføje sit eget mindre bidrag af dokumenter. Det er ingen

hemmelighed at den gode genkendelsesgrad i et ASR-system afhænger af et træningsmateriale der præcist dækker det ordforråd som applikationen skal genkende.

DanCASTs hub vedligeholder også et bibliotek af software, testemonials og andre dokumenter. Herfra kan kommunen downloade scripts og manualer og selv påbegynde arbejdet med at præparere sine tekster. Før et tekstkorpus kan anvendes til ASR-træning, skal det nemlig gennemgå en lang række processer til oprensning og normalisering.

- Formatering (dokumenter omlægges fra fx SQL, regneark, eller docx til klar tekst)
- Frasortering (redundant tekst, gentagelser, irrelevant markup, sidefod og sidehoved, ...)
- Tokenisering (ord frigøres fra omgivende tekstelementer såsom kommaer, parenteser, ...)
- Normalisering (forkortelser, numre, beløb, tider, etc. ændres til bogstaverede former)
- Anonymisering (følsomme informationer skjules: navne, cpr-numre, steder, etc.)
- Annotation (informationer af særlig betydning mærkes op)

Herunder er vist et stykke tekst fra en kommunal visitationsrapport. Fra den øverste form til den nederste form har teksten gennemgået en tokenisering, normalisering og anonymisering (forskellene er markeret med **fed** font; personnavn og dato er opdigtet). Teksten er derved blevet forberedt til træning af et ASR-systems sprogmodel.

* Prognose og behandlingsmuligheder **(1-3-12)**:

Der er klar **overenstemmelse** mellem de objektive fund og borgers fortælling..

- Foreligger der lægeskøn eller **udtalelse** fra egen læge **ang. pronose, arb evne mm?**

Ja. Egen læge ser ikke noget arbejdsmarkedsperspektiv for **Marianne**.

prognose og behandlingsmuligheder

første i tredje tolv

der er klar overensstemmelse mellem de objektive fund og borgers fortælling

foreligger der lægeskøn eller udtalelse fra egen læge angående prognose arbejdsevne med mere ja egen læge ser ikke noget arbejdsmarkedsperspektiv for _FirstName_

Til hvert af de nævnte trin i tekstforberedelsen stiller DanCASTs hub værktøj til rådighed (ikke-proprietært), sådan at kommunen får et reelt valg om den ønsker at overtage disse opgaver selv eller overlade dem til en totalentreprenør i en udbudssituation. Dette giver en ny valgfrihed, idet alle kontrakter mellem kommuner og leverandører til dato har placeret den samlede udviklingsopgave hos leverandøren. For nogle kommuner har det været en grim overraskelse at opdage at de derved har mistet rettighederne til egne data, som efter den nævnte forbehandling er overgået til ASR-leverandøren.

Endelig, som sit tredje indsatsområde, vedligeholder DanCASTs hub et kartotek af personer og organisationer med en særlig interesse i området (som for eksempel den centrale IT-ansvarlige i hver af kommunerne). Desuden giver vi små og mellemstore danske firmaer lov til at gøre opmærksom på sig selv, med en kort præsentation og kontaktinformation. På denne måde kan kommuner som måtte ønske at tage de første skridt i et ASR-projekt forud for en udbudsrunde, let finde hjælpende hænder med passende kompetencer, enten det er et lille nystartet firma eller en lingvistisk studerende på mindsteløn.

CBS tager ikke penge for sine ydelser, og DanCASTs tilbud til kommunerne er derfor gratis. Til gengæld giver kommunerne, gennem samarbejdsaftalen, CBS lov til at forske (not-for-profit) i de uploadede data, naturligvis sådan at CBS altid skal respektere kommunernes almene copyrights og den enkelte kommunes særlige anonymitetskrav. Desuden skal CBS slette alle kopier af lånte data, hvis en kommune vælger at afslutte sin aftale med DanCAST. Selv om DanCASTs hub kun har været åben for samarbejder i få uger, har allerede fem kommuner afleveret tekstdata til vores repository. Vi forventer lige så mange nye aftaler inden jul.

Høringen i november 2013

Om få uger indtræffer den store høring mellem kommuner og leverandører. Dybest set er høringens formål at sikre ASR-projekter der giver overskud for både køber og sælger, for hvis ikke vi snart opnår positive businesscases, risikerer vi at ASR ender helt uden for kommunernes interessefære.

Et vellykket udviklingsforløb kræver en detaljeret problemafklaring *før* kommunen går i konkret udbud. En stor del af afklaringsprocessen er fælles for alle kommuner, og tanken med høringen er at få tidlige svar fra industrien på så mange udfordringer som muligt. På den baggrund vil kommunerne stå bedre rustet til at udarbejde et generisk udbudsmateriale, som den enkelte kommune derefter kan tage som udgangspunkt.

For at få så megen nytte af høringen som muligt har en lønnet projektgruppe i månedsvis arbejdet med et oplæg, som nu er rundsendt til de potentielle ASR-leverandører. Oplægget konkretiserer de temaer som kommunerne i OS2 har kredset om i årevis. Hvilke krav kan og skal en kommune stille i sit udbudsmateriale? Hvilke dele af en ASR-løsning kan kommunen selv have ansvaret for at udvikle? Hvilke moduler i en samlet løsning kan kommunen bevare ejerskab til og genbruge i en senere udbudssituation (når der måske har vist sig en ny og stærkere taleteknologi i markedet)? Blandt de næsten hundrede punkter venter vi særligt spændt på industriens reaktioner på disse:

- Tilbudsgiver skal redegøre for hvor hyppigt den akustiske model opdateres.
- Tilbudsgiver skal redegøre for om det er muligt for ordregiver at få den akustiske model opdateret efter ønske.
- Det skal være muligt at anvende mere end en sprogmodel i den tilbudte løsning, for eksempel en sprogmodel til beskæftigelsesområdet og en sprogmodel til miljøområdet.
- Det skal være muligt at få gentrænet eksisterende sprogmodeller efter ønske.
- Tilbudsgiver skal redegøre for hvor lang tid det tager at gentræne en eksisterende sprogmodel.

- Tilbudsgiver skal redegøre for hvilke leverancer ordregiver i givet fald selv kan stå for, ved udarbejdelse af nye sprogmodeller.
- Tilbudsgiver skal redegøre for hvilke trin i forædlingsprocessen [tekstforberedelsen omtalt herover] kommunen selv kan udføre og hvilke tilbudsgiver skal udføre.
- Tilbudsgiver skal redegøre for om der er mulighed for at genbruge eventuelt udviklede databaser (for eksempel den fonetiske ordbog) ved senere inddragelse af nye taleteknologier som Text-to-Speech.

Det samlede høringsmateriale kan beses ved henvendelse til forfatteren.

Afsluttende bemærkninger

Den nyligt udkomne hvidbog, Europe's Languages in the Digital Age, konkluderer at de fleste europæiske sprog er dårligt forsynet med sprogteknologi.² Blandt de europæiske lande udmærker Norden sig ikke særligt. Vores sprog er for små til at hjemmemarkederne af sig selv genererer firmaer og produkter til sine egne sprog. Det gælder især de tungere teknologier som talesyntese (text-to-speech), automatiske oversættelse og talegenkendelse, hvor man (med få undtagelser) kun kan håbe på produkter fra internationale leverandører der i bund og grund er tillempede løsninger konciperet til andre sprog. Kvaliteten er næsten aldrig på niveau med den tilsvarende engelske, hvad der giver vores yngre generation endnu en grund til at svippe over i engelsk når de skal betjene deres hardware, browsere, GPS'er, biler, og så videre. Denne udvikling er blevet kaldt *den digitale sprogdød*.

Vi tror ikke at der er grund til dødsangst. Nok er vores sprog små, men af langt større vigtighed end det nøgne antal af talere er markedets grad af organisation og stabilitet. I Norden vil den offentlige sektor til alle tider være en af samfundets stærkeste og mest stabile købere. Når først de vellykkede projekter begynder at dominere, bliver der plads til et helt økosystem af leverandører. Et naivt håb? Næppe. Alene rygten om mere fair udbudsrunder har allerede fået to nye danske firmaer til at tjekke ind blandt tilbudsgiverne. Måske synger monopolet på sit sidste vers.

² <http://www.meta-net.eu/whitepapers/key-results-and-cross-language-comparison>

Peter Juel Henriksen er associate professor og principal investigator ved Danish Center for Applied Speech Technology (DanCAST), Copenhagen Business School, pjh.abc@cbs.dk

Organisationer og andre der er nævnt i artiklen:

OS2	http://www.os2web.dk
OS2Talk	http://www.os2web.dk/document/os2talk
KOMBIT	http://www.kombit.dk
Digitaliseringsstyrelsen	http://www.digst.dk
CBS/language-technology	http://bridge.cbs.dk
DanCAST	http://www.DanCAST.dk

Summary

Many Danish municipalities are quite ready to implement automatic speech recognition, however at the same time reluctant after a line of negative business cases in the recent past. Complaints have been heard of over-priced licenses and low service level, the typical effect of a de facto monopoly situation in the ICT market. This article presents a new public effectuation strategy installed in 2013. Denmark may well be a small country, but our public sector is highly organized and (for this reason) economically quite capable. During the last 12 months the municipalities have organized themselves, formulating together a list of requirements to be met by the suppliers as a part of the new tender documents. This work has been supervised by university based researchers (including this author) acting as commercially unbiased advisors. The initiative is intended as a significant move towards the first mutually profitable ASR contract in this area, an event we hope will occur in 2014 and which may be exactly what is needed to dissolve the current deadlock and open up a billion kroner market for ASR in the Danish public sector.

Språkteknologi och språkresurser för språken i Sverige: En statusrapport

Lars Borin och Rickard Domeij

Under senare år har behovet av språkteknologi och språkresurser för svenska och andra språk i Sverige kommit att uppmärksammas allt mer. Detta handlar dels om att på bästa sätt förvalta Sveriges under lång tid upparbetade spetskompetens inom akademisk forskning i språkteknologi, dels även om att se till att denna forsknings resultat omsätts i produkter och tjänster som kommer det svenska samhället och dess medborgare till gagn. I denna artikel lämnar vi en rapport om den akademiska språkteknologiforskningen i Sverige (avsnitt 1) och om ett färskt politiskt initiativ för att få till stånd en nationell språkresursinfrastruktur som ska kunna användas för produkt- och tjänsteutveckling (avsnitt 2).

Mot en svensk språkteknologi för forskning och utveckling – en rapport från fronten

När det gäller språkteknologi för svenska och dess användning i forskning och utveckling, har denna å ena sidan en lång historia i Sverige men å den andra ser man att utvecklingen i viss mån har stagnerat.

Den svenska forskningen inom området går tillbaka till 1950-talet för talteknologi och 1960-talet för textkorpusar och språkverktyg för det skrivna språket och har lett till att det idag finns ett antal aktiva forskargrupper i landet, som representerar de flesta av språkteknologins områden:¹

- svenska korpusar (Språkbanken vid Göteborgs universitet, Stockholm, Uppsala)

¹ Dessutom finns en rad företag som utvecklar språkteknologiska produkter och tjänster och som bedriver egen forskningsverksamhet, ofta i samarbete med akademiska forskningsmiljöer. Fokus i det här avsnittet ligger dock på den akademiska forskningen i snävare bemärkelse.

- flerspråkiga korpusar (Uppsala, Linköping, Språkbanken)
- taldatabaser (KTH)
- resurser för informationsåtkomst (SICS, KTH, Stockholm, Språkbanken)
- lexikondatabaser (Språkbanken, KTH, Språkrådet)
- många olika språkverktyg för text och tal (samtliga grupper)

Huvudsakligen bedrivs denna forskning i form av avgränsade korta forskningsprojekt och den är mer fragmenterad än vad som skulle vara önskvärt. De huvudsakliga finansieringsformerna täcker inte långsiktigt underhåll och tillgängliggörande av språkresurser. Språkbanken i Göteborg är den enda grupp som systematiskt arbetar med att säkerställa detta, men strikt på eget initiativ och med lokal finansiering från Göteborgs universitets humanistiska fakultet.

Det finns således ett stort behov av samordning av resursuppbyggnaden samt av harmonisering av språkresurser och språkverktyg med avseende på informationsstruktur, dataformat, programgränssnitt, licensvillkor, etc., så att en positiv spiraleffekt kan uppstå, där både ny forskning och kommersiell produktutveckling kan använda befintliga resurser fullt ut, utan att ständigt behöva återuppfinna hjulet eller hålla tillgodo med undermåliga lösningar.

I någon mån går utvecklingen i denna riktning, genom svensk medverkan i två breda europeiska initiativ, META-NORD och CLARIN, som beskrivs i de närmast följande avsnitten nedan.

META-NORD

META-NORD-projektet (2011–2013) var ett EU-finansierat samarbete mellan de fem nordiska och de tre baltiska staterna, med syftet att stärka och utveckla potentialen för språkteknologibaserad forskning och utveckling i det nordisk-baltiska området. META-NORD ingick tillsammans med två andra liknande projektsamarbeten under det vidare paraplyet META-NET,² ett europeiskt spetsforskningsnätverk med huvudsyftet att

² <<http://www.meta-net.eu/>>

främja den teknologiska grunden för ett flerspråkigt europeiskt informationssamhälle.

Den svenska delen av META-NORD åstadkom tre typer av resultat:

1. En vitbok om tillståndet för svenska språket med avseende på språkresurser och språkteknologi (Borin m.fl. 2012a), samt medvetandegörande om detta tillstånd hos forskare, politiker, forskningsfinansiärer, språkteknologiföretag, m.fl.
2. En kartläggning av språkresurser och språkverktyg i Sverige samt katalogisering av dessa i den gemensamma databasen META-SHARE³ (Skadina m.fl. 2011)
3. Anpassning av utvalda befintliga resurser till gemensamma standarder samt vidareutveckling av vissa resurser, i det svenska fallet länkning av ett svenskt ordnät till de danska, estniska och finska ordnäten via det engelska Princeton WordNet (Pedersen m.fl. 2012, 2013)

På grundval av den information som samlats in för de sammanlagt 32 META-NET-vitböckerna om 31 språk (bokmål och nynorska har varsin vitbok) har resursläget för dessa språk kunnat sammanfattas som i figur 1.

Sammanfattningsvis har META-NORD inom de relativt begränsade ramar som projektet förfogade över ändå kunnat lägga en god grund för vidare arbete, nationellt, nordiskt och europeiskt, och bidragit till att höja medvetenheten om behovet av fokuserade satsningar på svensk språkteknologi.

SWE-CLARIN

Termen *e-vetenskap* används ofta för att beteckna vetenskap som med hjälp av modern informationsteknik kan angripa problem av en omfattning och komplexitet som inte skulle kunna hanteras utan elektroniska hjälpmedel och möjligheten att dra fördel av geografiskt spridda resurser. Traditionellt har detta angreppssätt främst förknippats med extremt beräkningsintensiv forskning inom teknik- och naturvetenskap men i själva verket skulle man kunna säga att t.ex. korpuslingvistik (som med hjälp

3 <<http://spraakbanken.gu.se/metashare/>>

av språkteknologi bedriver forskning på stora insamlade textmassor) är ett mycket tidigt exempel på e-vetenskap. Detta är också ett område som utvecklats starkt över femtio år, och den teknologi och teoribildning som utvecklats i anslutning till detta forskningsfält har idag fått en rik flora av applikationer utanför den rent lingvistiska intressesfären.



Figur 1. Resursslåget för 31 av Europas språk enligt META-NETs vitbok (svenska inringat; Borin m.fl. 2012a: 33f).

Det stora europeiska infrastrukturprojektet CLARIN⁴ (Common Language Resources and Technology Infrastructure) syftar till att göra digitala språkreсурser (textsamlingar, inspelningar av ljud och bild, lexikon och så vidare), tillsammans med de språkteknologiska verktyg som behövs för att hantera dem, tillgängliga för forskare inom alla discipliner. Särskilt tonvikt läggs på forskare inom humaniora och samhällsvetenskap. Med andra ord: CLARIN erbjuder en möjlighet för alla forskare inom humaniora och samhällsvetenskap (och även andra discipliner som t.ex. medi-

4 <<http://www.clarin.eu/>>

cin och vårdvetenskap) att använda sig av e-vetenskapliga metoder i sin verksamhet.

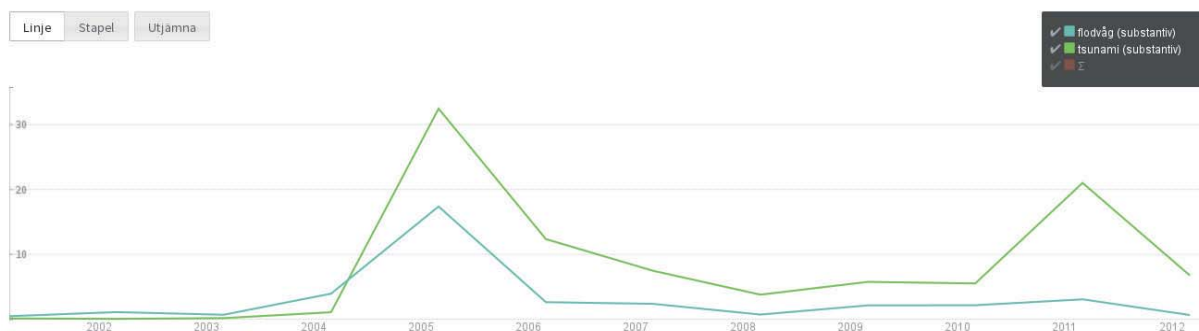
Efter en fyraårig förberedelsefas fick CLARIN under 2012 status av ERIC (European Research Infrastructure Consortium), en sammanslutning av stater och mellanstatliga organisationer. Vid inrättandet den 29/2 2012 hade CLARIN ERIC 9 medlemmar (8 EU-medlemsländer och en mellanstatlig organisation). Sverige var inte en av de initiala medlemmarna. En rad svenska forskargrupper hade dock deltagit i CLARINs förberedelsefas, och nio organisationer bildade därför ett konsortium för att förbereda en ansökan till Vetenskapsrådet om svenskt medlemskap i CLARIN ERIC samt fullt svenskt deltagande i CLARINs arbete, genom att:

1. definiera en nationell organisation för CLARIN-arbetet,
2. utforma en struktur och organisation för CLARIN-centra i Sverige,
3. utforma en teknisk infrastruktur och definiera de tjänster som ska vara tillgängliga via CLARIN,
4. utreda behovet av språkresurser och språkverktyg för Sveriges språk (svenska, minoritetsspråken och andra viktiga språk) i ljuset av CLARINs behov och målsättningar,
5. identifiera och mobilisera viktiga aktörer i forskningsvärlden, i kultur- arvsinstitutionerna och i relevanta digitaliseringsprojekt.

Ansökan beviljades av Vetenskapsrådet i november 2013, men exakt belopp och villkor är i skrivande stund inte kända utan ska bli föremål för förhandling under första kvartalet 2014. Den planerade nationella infrastrukturen inom området – SWE-CLARIN – kommer organisatoriskt att bestå av ett antal centra, baserade på väletablerade forsknings- och utvecklingsmiljöer. Dessa centra ska dels skapa, uppgradera och underhålla de digitala resurserna och verktygen, och dels fungera som utbildnings-, rådgivnings- och stödenheter för forskare utanför språkteknologigemenskapen.

Mot digital spetsforskningspotential med svensk språkteknologi: Kunskapsbaserad kulturomik

Som ett konkret exempel på den sorts aktiviteter som hör hemma inom CLARIN kan vi här anföra det nystartade VR-ramprogrammet *Mot kunskapsbaserad storskalig kunskapsutvinning ur svensk text* ('Towards a knowledge-based culturomics'), ett samarbete mellan Språkbanken i Göteborg, datavetare vid Chalmers tekniska högskola i Göteborg och språkteknologer vid Lunds universitet.



Figur 2. Förekomster av orden *flodvåg* (den undre blå kurvan) och *tsunami* (den övre gröna kurvan) i dagstidningar 2001–2012 (normaliserade lemmafrekvenser)

Nyligen har några forskare börjat utnyttja de enorma textmängder som resulterat ur Googles massiva bokdigitaliseringsprojekt för att i dessa textmassor försöka följa språklig och kulturell utveckling över de två senaste seklerna (Michel m.fl. 2011). Forskningsområdet har med buller och bång lanserats under namnet *culturomics*⁵ (analogt med *genomics*, *proteomics*, etc.), men de första studierna har med rätta kritiserats för att helt ignorera relevanta tidigare arbeten i språkteknologi och lingvistik, och t.ex. inte diskutera det inte alldeles enkla begreppet ”ord” i den här kontexten. Samtidigt är detta forskning som verkligen ligger i tiden. Det finns nu enorma mängder digital text att tillgå på svenska. Bara de svenska bloggarna uppgår till miljarder ord. Dessutom pågår ett antal kulturarvsdigitaliseringsprojekt, t.ex. har Kungliga biblioteket och Riksarkivet digitaliserat stora mängder svensk dagspress från de senaste 300 åren, sammanlagt miljarder ord.

Syftet med detta projekt är att lyfta *kulturomik* till kunskapsbaserad storskalig kunskapsutvinning ur stora mängder digitaliserad svensk text,

5 Se <<http://www.culturomics.org/>>

såväl modern som äldre (Borin och Johansson 2014). De inledande studierna i projektet behandlar politiska opinionsyttringar i sociala medier, utveckling av system som kan svara på frågor om textinnehåll, samt ordutveckling över tid. Sålunda visar figur 2 hur den tidigare huvudsakligen tekniska termen *tsunami* blev ett allmänord i samband med den tragiska naturkatastrofen i Sydostasien i slutet av 2004 och hur ordet sedan helt hade ersatt det tidigare *flodvåg* vid rapporteringen från Fukushimaolyckan 2011. Visualiseringen i figur 2 har gjorts med trenddiagramfunktionen i Språkbankens korpushanteringssystem Korp⁶ (Borin m.fl. 2012b) på basis av ett stort dagstidningsmaterial.

Förslag till nationell infrastruktur för det digitala samhället

Språkrådet i Sverige har på uppdrag av den svenska regeringen (Ku2011/860/KA) tagit fram ett förslag på hur en nationell infrastruktur för språken i Sverige kan skapas med utgångspunkt i behovet av språkresurser för utveckling av taligenkänning för tv-textning (*Infrastruktur för språken i Sverige* 2012). Som bakgrund till uppdraget skriver regeringen:

För att främja utvecklingen av teknik som ökar tillgängligheten till information för alla har flera berörda aktörer uttryckt behovet av att etablera en nationell språkdatabank med öppet tillgängliga språkdatabaser och tillhörande analysverktyg. En nationell språkdatabank avseende det svenska språket, de nationella minoritetsspråken och det svenska teckenspråket som omfattas av språklagen (2009:600), kan också utgöra grund för olika tillgänglighetstjänster, t.ex. på tv-området.

I det följande presenteras bakgrunden till förslaget, huvuddragen i det och hur regeringen tänker sig att gå vidare med det.

Ökade krav på tillgänglig information och kommunikation

Kraven på informationstillgänglighet har ökat i takt med att den digitala tekniken skapat nya möjligheter att göra information och service mer tillgängliga för alla. Enligt FN:s konvention om rättigheter för personer med funktionsnedsättning ska medlemsstaterna ”säkerställa att personer med

⁶ Se <<http://spraakbanken.gu.se/korp/>>

funktionsnedsättning får tillgång på samma villkor som andra till information och kommunikation” (artikel 9), även ”till tv-program, film, teater och annan kulturell verksamhet i tillgänglig form” (artikel 30).

Detta har lett till ökade krav i Radio- och tv-lagen (2010:696) och i sändningstillståndet för Sveriges television (SVT) att tillgängliggöra tv-utbudet via undertextning, uppläst textremsa och teckenspråkstolkning. Liknande krav ställs på public service-kanalerna i andra nordiska och europeiska länder (Nordens välfärdscenter 2011, ITU 2011). Det uppsatta målet i SVT:s sändningstillstånd är att samtliga program på svenska ska textas med hög kvalitet, även de direktsända. För att uppnå det målet behöver SVT utveckla ny teknik för tv-textning baserad på taligenkänning (SOU 2012:59), vilket uppmärksammades i den svenska Public service-utredningen som presenterades våren 2013 (SOU 2012:59).

Behovet av nationell språkinfrastruktur

För att utveckla sådan teknik krävs tillgång till språkresurser i form av tal-databaser, lexikon och andra språkdata som kräver stora arbetsinsatser och kostnader för att ta fram. Hittills har skapandet av sådana språkresurser främst inriktats på ett begränsat antal kommersiellt intressanta språk, i synnerhet engelska. Taligenkänningstekniken är därför i dag enbart förbehållen tv-företag som verkar inom stora språkområden, som exempelvis BBC. I Sverige och andra nordiska länder har marknadskrafterna inte varit starka nog att driva på utvecklingen i tillräcklig utsträckning (Domeij 2013). För de flesta europeiska språk saknas enligt den undersökning som redovisas i META-NETs vitböcker (Borin m.fl. 2012a; se avsnittet om META-NORD ovan) många av de grundläggande språkresurser som ses som nödvändiga för att stimulera forskningen och teknikutvecklingen på området. På sikt utgör det ett hot mot många av de europeiska språken, i synnerhet minoritetsspråken. I en rapport om tillgängliga medier från Nordens välfärdscenter (2011) nämns avsaknad av språkteknologisk utveckling som en av tre huvudorsaker till att alla inte har tillgång till tv i de nordiska länderna.

Tillgången till språkresurser är alltså en central språkpolitisk fråga med avgörande betydelse för språkutvecklingen och tillgängligheten till information och service i det digitala samhället (Domeij m.fl. 2011). När det kommersiella intresset inte är tillräckligt stort behöver samhället ta ansvar för att skapa den språkliga infrastruktur som krävs för att tillgodose behovet av språkresurser och driva på teknikutvecklingen, så att vi i Sverige

och andra nordiska länder kan ta del av de nya tekniska möjligheterna på någorlunda lika villkor som andra länder med mer kommersiellt gångbara språk. I Norge har man redan etablerat en nationell språkbank med uppgift att tillgodose behovet av språklig infrastruktur. I Sverige har frågan under senare år uppmärksammats av regeringen.

I *It i människans tjänst – en digital agenda för Sverige* (2011) presenterade den svenska regeringen en strategi med målet att göra Sverige bäst i världen på att utnyttja digitaliseringens möjligheter. Särskilt intressant ur ett språkligt perspektiv är betoningen på e-tillgänglighet och infrastruktur för att stimulera teknikutvecklingen. Där uppmärksammas också behovet av en nationell språkinfrastruktur för att stimulera utvecklingen av digitala produkter och tjänster. Regeringen skriver att ”det långsiktiga etablerandet av en nationell språkbank med språkdata-baser främjar utvecklingen av teknik vilket gagnar språken i Sverige och ökar tillgängligheten till information för alla” (s. 37). Därmed finns frågan om en nationell språkinfrastruktur på den politiska agendan.

Formuleringen rimmar väl med regeringens satsning på öppna data för att stimulera utvecklingen av digitala tjänster genom att uppmuntra myndigheter och andra organisationer att fritt dela med sig av befintliga data för utveckling av digitala tjänster och produkter. Öppna data kan förenklat sägas vara information som är tillgänglig utan inskränkningar i form av kostnader eller immaterialrättsliga hinder, vilket är av stor betydelse för utvecklingen av språkteknologi, inte minst för mindre resursstarka språk (Nørstebø Moshagen och Langgård 2011).

Förslag till nationell infrastruktur för språken i Sverige

I enlighet med den digitala agendan gav regeringen i uppdrag åt Institutet för språk och folkminnen, som Språkrådet är en del av, att ta fram ett beredningsunderlag för att utveckla formerna för drift och samordning för en nationell språkdata-bank (ku2011/860/KA). Det har ingått i uppdraget att ta fram en behovsanalys och kostnadsberäkning för projektet, undersöka vilka tillgängliga tjänster som kan användas samt lämna förslag till långsiktig finansiering av språkdata-banken. En viktig utgångspunkt är att data-banken bör vara öppen för aktörer som vill använda den som bas för produkt- och tjänsteutveckling.

I februari 2012 lämnades underlaget till kulturdepartementet i form av rapporten *Infrastruktur för språken i Sverige – Förslag till nationell språkinfrastruktur för det digitala samhället* (2012). Det förslag som läggs

fram i rapporten innebär i korthet att en nationell språkinfrastruktur för talbaserade tjänster stegvis etableras med utgångspunkt i behovet av direkttextning för tv. Rapporten hänvisar till att utvecklingen av taligenkänning nu står inför ett genombrott, vilket ger stora möjligheter att stimulera teknikutvecklingen och öka tillgängligheten till information och tjänster för alla, förutsatt att erforderlig infrastruktur snabbt kan utvecklas och göras tillgänglig för tjänsteutveckling. En första uppgift bör vara att ta fram taldata-baser för direkttextning av tv med siktet inställt på den långsiktiga uppgiften att se till att en bred och representativ taldata-bas för svenska görs tillgänglig för tjänsteutveckling och hålls uppdaterad över tid.

Att utveckla en taldata-bas för direkttextning av tv är ett konkret exempel på språkresursutveckling i samarbete med tjänsteutvecklare som öppnar möjligheter till ett långsiktigt samarbete, där SVT kan anpassa sin verksamhet till höjda tillgänglighetskrav samtidigt som SVT:s omfattande språkmaterial kan användas för utveckling av språkresurser och göras fritt tillgängliga för andra tjänsteutvecklare. Dörren bör öppnas även för andra intressenter som TV4. Ett sådant samarbete skapar också möjligheter till långsiktig finansiering av verksamheten.

Rapporten föreslog vidare att Språkrådet vid Institutet för språk och folkminnen får i uppdrag att samordna och följa upp arbetet med språklig infrastruktur på nationell nivå. I uppdraget ingår att ta fram en nationell strategi för arbetet med språklig infrastruktur, fortlöpande undersöka språkresursbehovet, inventera språkresurstillgången och förmedla befintliga språkresurser till teknikutvecklare.

Språkresurssituationen i Sverige

Som nämnts i avsnitt 1 är läget i Sverige nu det att språkresurssituationen är fragmenterad. Flera nödvändiga basresurser saknas, bl.a. representativa tal- och textdata-baser för svenska. En del språkresurser som tagits fram i olika forskningsprojekt finns tillgängliga för forskningsändamål på de forskningsinstitutioner som ansvarat för projektet. Dessa språkresurser underhålls vanligen inte efter att projektfinansieringen upphört.

Språkbanken i Göteborg bedriver som sagt en mer systematisk språkbankverksamhet med inriktning på textresurser. Däremot saknas motsvarande verksamhet med inriktning på talresurser. Terminologicentrum TNC tillgängliggör terminologi via Rikstermbanken. Det finns också flera myndigheter som sitter på viktiga språkresurser, t.ex. Myndigheten för

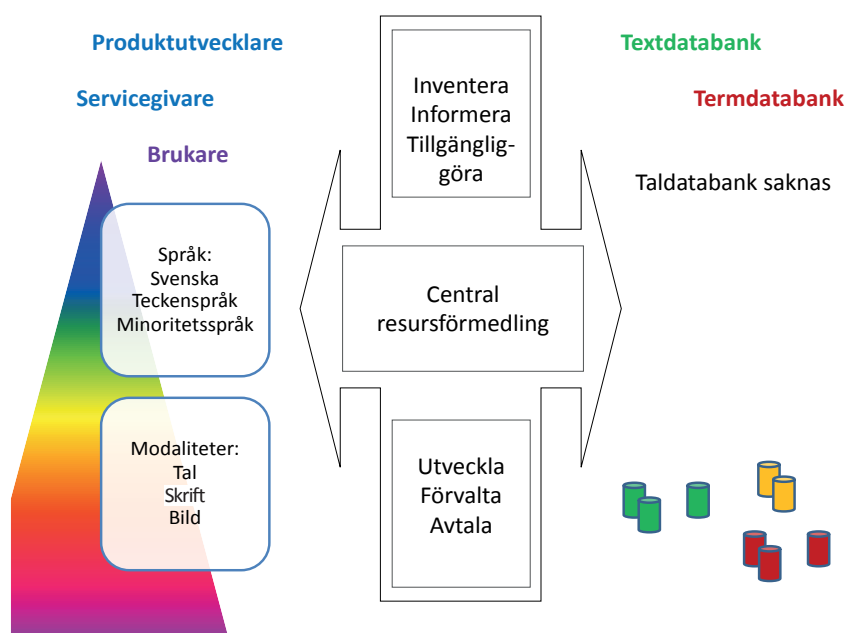
tillgängliga medier (MTM) och Institutet för språk och folkminnen dit Språkrådet hör.

Figur 3 ger en schematisk bild som utgångspunkt för överväganden om vad som skulle behöva göras för att förbättra språkressituationen i Sverige och stimulera produkt- och tjänsteutvecklingen. Högersidan i bilden illustrerar på ett förenklat sätt det faktum att det redan finns en del språkresurser i form av textdatabaser (grönt), termdatabaser (rött) och i taldatabaser (gult). För att stimulera utvecklingen av de tillgängliga tjänster (pyramiden på vänster sida) som ska komma brukare och servicegivare till del behöver efterfrågade språkresurser utvecklas och tillgängliggöras för produkt- och tjänsteutveckling.

Tre verksamheter behöver realiseras

Rapporten till regeringen föreslår att de tre verksamheter som skisseras i rektanglarna i mitten av figur 3 bör realiseras i syfte att utveckla och tillgängliggöra efterfrågade resurser. För det första behöver man göra förlöpande inventeringar av resursutbudet. Man behöver ta reda på vilka resurser som finns på olika organisationer, vilken kvalitet de har (med hänsyn till standarder m.m.), var man hittar dem och hur de är tillgängliga. Detta gäller inte bara för svenska utan också för de nationella minoritetsspråken och det svenska teckenspråket.

För det andra behövs en central resursförmedling som hjälper tjänsteutvecklarna att få tag på befintliga resurser på ett smidigt sätt. Detta kan göras genom ett samordnat arbete mellan olika språkresursinnehavare och -förvaltare som ställer sina språkresurser till förfogande via den centrala resursförmedlingen. I anslutning till förmedlingsverksamheten behöver det finnas information och service av olika slag, som att ge svar på frågor om språkresurserna från tjänsteutvecklare som vill använda dem och att ge stöd till personer eller organisationer som vill tillgängliggöra sina resurser via resursförmedlingen. Det kan handla om hjälp med att hantera juridiska frågor, med att göra metabeskrivningar, med att hitta lämplig förvaltare och liknande.



Figur 3. Schematisk bild med förslag till funktioner för bl.a. inventering, förmedling och utveckling av språkresurser.

För det tredje måste det finnas ekonomiska möjligheter att se till att efterfrågade resurser utvecklas och sedan också underhålls så att de kan hållas uppdaterade och standardanpassade. De kan också behöva anpassas på andra sätt med hänsyn till tjänsteutvecklarnas behov. I samband med detta bör man också kunna hantera licenser. Som alternativ till att utveckla nya resurser kan redan befintliga men inlåsta resurser frigöras så att de fritt kan användas för tjänsteutveckling, vilket innebär att frågor om upphovsrätt och personuppgifter måste kunna hanteras.

Utöver ovan beskrivna verksamheter behövs språkbanksverksamhet för lagring, underhåll och distribution av olika typer av språkresurser. Som framgår av figurens övre högra del finns i dag en språkbank för textresurser, dvs. en textbank, i form av Språkbanken vid Göteborgs universitet. Det finns dessutom en termbank i form av Rikstermbanken på Terminologicentrum TNC (Rikstermbanken). Däremot saknas en talbank för underhåll och tillgängliggörande av talresurser, vilket är ett stort hinder för utveckling av talbaserade tjänster. För att direkttextning av tv och andra talbaserade tjänster ska kunna utvecklas behöver därför en språkbanksverksamhet med inriktning på tal etableras, en taldatabank. Detta bör rimligen göras i nära samarbete med KTH som har internationellt gångbar kompetens på talteknologiområdet.

Slutord och framåtblick

Som framgått av ovanstående beskrivning behöver det på ett övergripande plan finnas en funktion för samordning av olika verksamheter och organisationer som deltar i ett nationellt samarbete för utveckling och tillgängliggörande av språkresurser. I anslutning till denna behöver en nationell strategi utarbetas för hur arbetet kring språkresurser ska bedrivas i samverkan mellan olika parter.

En nödvändig grundförutsättning för en god utveckling på språkteknologiområdet är att ett nödvändigt basutbud av språkresurser (en s.k. BLARK - *Basic Language Resource Kit*) tas fram och tillgängliggörs för forskning och utveckling av språkteknologi genom finansiering från Vetenskapsrådet och andra statliga finansiärer av grundläggande infrastruktur. Utöver det behövs riktade satsningar på utveckling av efterfrågade språkresurser för tjänsteutveckling. Eftersom behovet av språklig infrastruktur för forskningen till stora delar överlappar med den infrastruktur som behövs för produkt- och tjänsteutveckling är det viktigt att arbetet med språkinfrastruktur för forskning respektive produktutveckling samordnas.

Efter att under en tid ha behandlat det ovan beskrivna förslaget i kulturdepartementet aviserade den svenska regeringen i juni 2013 att den avser att starta ett pilotprojekt för att arbeta vidare på basis av förslaget. I december 2013 beviljade också Vetenskapsrådet en ansökan om svenskt deltagande i CLARIN vilket gör att förutsättningarna nu ser goda ut för att en mera organiserad och samordnad verksamhet kring en nationell språkinfrastruktur ska komma till stånd.

Nyckelord: språkteknologi, språkbank, språkinfrastruktur, språkresurser, META-NORD, SWE-CLARIN.

Lars Borin är professor i språkvetenskaplig databehandling, föreståndare för Språkbanken och Centre for Language Technology, Göteborgs universitet. Nationell samordnare för SWE-CLARIN, den svenska CLARIN-noden. Språkteknolog med bakgrund i jämförande språkvetenskap (slaviska och finsk-ugriska språk) och med ett aktivt forskningsintresse inom e-vetenskap, i skärningspunkten mellan språkteknologi och en rad humanistiska och samhällsvetenskapliga discipliner.

Rickard Domeij är doktor i datorlingvistik och arbetar som språkvårdare på Språkrådet i Sverige med språk och it som specialområde.

Litteratur

- Borin, Lars, M.D., Brandt, J. Edlund, J. Lindh, M. Parkvall, 2012a: *Svenska språket i den digitala tidsåldern. The Swedish Language in the Digital Age*. META-NETs vitboksserie. Berlin: Springer. <http://www.meta-net.eu/whitepapers/volumes/swedish-sv?set_language=sv>.
- Borin, Lars, M. Forsberg, J. Roxendal, 2012b: Korp – the corpus infrastructure of Språkbanken. *Proceedings of LREC 2012*, s. 474–478. Istanbul: ELRA.
- Borin, Lars och Richard Johansson, 2014: Kulturomik: Att spana efter språkliga och kulturella förändringar i digitala textarkiv. Populärvetenskaplig framställning för bloggen *Historia i en digital värld*. <<http://digihist.se/5-metoder-inom-digital-historia/fordjupning-kulturomik-att-spana-efter-sprakliga-och-kulturella-forandringar-i-digitala-textarkiv/>>
- Moshagen, Sjur N. and Per Langgård (eds.), 2011: *Visibility and Availability of LT Resources. Proceedings of the NODALIDA 2011 workshop*. NEALT Proceedings Series Vol 13. Riga: NEALT.
- Domeij, Rickard, T. Breivik, J. Halskov, S. Kirchmeier-Andersen, P. Langgård, S. Nørstebø Moshagen (red.), 2011: *Språkteknologi för ökad tillgänglighet – vilka möjligheter finns?* Utarbetad av Astin för Nätverket för de nordiska språknämnderna. Linköping: Linköping university electronic press. ISBN 978-91-7393-094-9.
- Domeij, Rickard, 2013: Behovet av språkinfrastruktur för direkttextad tv. I: Breivik m.fl. (red.), *Språk i Norden 2013*.
- Infrastruktur för språken i Sverige. Förslag till nationell språkinfrastruktur för det digitala samhället*, 2012: Beredningsunderlag för att utveckla formerna för en nationell språkdatbank enligt regeringsuppdrag Ku2011/860/KA.
- It i människans tjänst – en digital agenda för Sverige*, 2011: Rapport från Näringsdepartementet N2011.12.
- ITU, 2011: *Making television accessible*. International Telecommunication Union och G3ICT.
- Michel, Jean-Baptiste, Y.K. Shen, A.P. Aiden, A. Veres, M.K. Gray, The Google Books Team, J.P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M.A. Nowak, E. Lieberman Aiden, 2011: Quantitative analysis of culture using millions of digitized books. *Science*, 331.

- Nordens välfärdscenter, 2011: *Medier för alla? Tillgänglighet till tv i Danmark, Finland, Norge och Sverige*. Rapport från ett möte mellan tre nordiska samarbetsorgan för funktionshindersonsorganisationer.
- Pedersen, Bolette Sandford, L. Borin, M. Forsberg, K. Lindén, H. Orav, E. Rögnvaldsson, 2012: Linking and validating Nordic and Baltic wordnets. *Proceedings of 6th International Global Wordnet Conference*. s. 254-260.
- Pedersen, Bolette Sandford, L. Borin, M. Forsberg, N. Kahusk, K. Lindén, J. Niemi, N. Nisbeth, L. Nygaard, H. Orav, E. Rögnvaldsson, M. Seaton, K. Vider, K. Voionmaa, 2013: Nordic and Baltic wordnets aligned and compared through “WordTies”. *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013). NEALT Proceedings Series 16*, s. 147-162.
- Proposition 2012/13:164. *Bildning och tillgänglighet – radio och tv i allmänhetens tjänst 2014–2019*.
- Skadina, Inguna, A. Vasiljevs, L. Borin, K. De Smedt, K. Lindén, E. Rögnvaldsson, 2011: META-NORD: Towards sharing of language resources in Nordic and Baltic countries. *Proceedings of the Workshop on Language Resources, Technology and Services in the Sharing Paradigm*, s. 107-114. Chiang Mai, Thailand: Asian Federation of Natural Language Processing.
- SOU (2012:59) *Nya villkor för public service*. Kulturdepartementet, Public service-kommittén (Ku 2011:05).

Summary

Over the last few years there has been an increased appreciation of the need for high-quality language resources (LR) and language technology (LT) for Swedish and the other languages of Sweden. This issue has two facets: (1) Safeguarding the achievements of an old and strong Swedish LT research tradition; and (2) ensuring that the results of this research are realized in products and services that will benefit Swedish society and its denizens. This paper describes some connected efforts in this direction, which manifest themselves both in academic initiatives aiming at surveying, standardizing, upgrading and interlinking Swedish LRs and LT (the META-NORD and SWE-CLARIN R&D collaborations), and in a recent political initiative by the Swedish Ministry of Culture with the goal of establishing a national LR/LT infrastructure supporting the development of language-aware products and services.

***Islex*: eit skuleeksempel på godt språktechnologisk samarbeid i Norden**

Margunn Rauset

Islex er blant dei største nordiske samarbeidsprosjekta på språkfronten dei seinare åra. Det har resultert i eit fleirspråkleg digitalt ordboksverk som i dag består av seks ordbøker, som i ulik grad er ferdigstilte. Utgangsspråket i alle ordbøkene er islandsk, medan dansk, norsk (både bokmål og nynorsk), svensk og i seinare tid færøysk og finsk er språka ordartiklane blir omsette til. Både prosessen og produkta blir annleis om eit prosjekt er digitalt frå starten enn om ein utarbeider ei trykt ordbok som sidan blir digitalisert. Her tek eg i hovudsak føre meg korleis det digitale mediet har påverka redaksjonsarbeidet med norsk, dansk og svensk, men set òg det fellesnordiske prosjektet inn i ein større språktechnologisk samanheng.

***Islex* som prosjekt og nettordbøker**

Denne artikkelen tek utgangspunkt i erfaringar og refleksjonar vi hausta av arbeidet med den islandsk-norske *Islex*-ordboka ved Institutt for lingvistiske, litterære og estetiske studium ved Universitetet i Bergen og gjennom samarbeid med søsterredaksjonane våre ved Det Danske Sprog- og Litteraturselskab i København og Institutionen för svenska språket ved Göteborgs universitet. Den islandske samarbeidspartnaren vår held til ved Árni Magnússon-instituttet for islandske studiar i Reykjavik. Det var ved desse fire nordiske institusjonane arbeidet med *Islex* starta i 2006, og første versjon av nettordbøkene, som blei offisielt opna i november 2011, inneheldt islandsk på den eine sida og dansk, bokmål, nynorsk og svensk på den andre.

Fróðskaparsetur Føroya slutta seg til prosjektet i februar 2011, og planen er at den islandsk-færøyske ordboka skal opnast i løpet av 2014. Ein redaksjon knytt til Helsingfors universitet starta i januar 2013 arbeidet med ei islandsk-finsk ordbok. *Islex* var det første digitale ordboksverket som knytte saman dei nordiske språka, og det er både svært gledeleg og naturleg at «*Islex*-familien» har blitt utvida med færøysk og finsk dei seinare åra. I artiklane der alle språka ligg inne, ser dette slik ut:

vel ordbok:    

borgardómur subst m


 **uttale**


→ **BØYING**

borgar-dómur

jus

gammalt

 (domstol i Reykjavik som fram til 1989 behandla sivilsaker)

 byrett

Sjå den fulle artikkelen til [kría](#) i [Islex](#) her

Det digitale mediet gjer *Islex*-ordbøkene svært fleksible, dei kan på ulike måtar tilpassast brukarane sine behov og lyster. Brukar ein til dømes portalen [islex.is](#), er òg all tekst som ikkje er ein del av sjølve ordartiklane (òg kalla omteksten) på islandsk, og dansk er målspråk i artiklane. [islex.dk](#) er den danske sida, og her er omteksten på dansk. Brukar ein [islex.no](#), kjem ein inn i den nynorske delen (det kan enkelt endrast til bokmål ved å klikke på fana «norsk bokmål» rett over søkjefeltet), og [islex.se](#) er den svenske sida. Færøysk og finsk vil få eigne adresser, faner og flagg i brukargrensesnittet når medarbeidarane bak dei språka er komne litt lenger med sine prosjekt. I dag er einaste måten å få opp færøyske og finske omsetjingar på å trykkje på «alle»-knappen ved sida av dei skandinaviske flagga (sjå artikkelen til *kría* over), og sjølvsagt berre i dei tilfella redaksjonane har arbeidd med den aktuelle artikkelen.

Den islandske redaksjonen har vore den største og mest tonegjevande for utviklinga av både prosjektet og produkta. Han har hatt ansvar for det såkalla kjeldespråket (utgangsspråket i ei to- eller fleirspråkleg ordbok, i dette tilfellet islandsk) og for utforming og utvikling av databasen til *Islex* (Halldóra Jónsdóttir og Þórdís Úlfarsdóttir 2008). Redaksjonane i dei andre nordiske landa har ansvar for kvart sitt målspråk, altså språket som artiklane i ordboka blir omsette til (Aldís Sigurðardóttir o.a. 2008).

Plassering av *Islex*-ordbøkene i eit språkteknologiske landskap

Ordet *ordbok* gjev gjerne assosiasjonar til ei tradisjonell, trykt bok. *Islex* er derimot døme på ei samling ordbøker som er heildigitale; det har vore ein føresetnad for prosjektet at arbeidsmetodane og produkta skulle vere digitale, og heile ordboksverket er i dag gratis tilgjengeleg på Internett. In-

gen konkrete planar ligg føre om å trykkje nokon av *Islex*-ordbøkene som eit supplement til nettutgåvene. Det har vore diskutert og er teknisk absolutt mogleg, men det ville truleg vere lite økonomi i det. Alt tyder på at framtida for ordbøker er digital, og at brukarar i minkande grad er interesserte i å kjøpe trykte utgåver. I denne artikkelen blir *Islex*-ordbøkene omtalte både som nettordbøker og digitale ordbøker. I ein periode gav ein ut ordbøker i format som CD-ROM og liknande, men her tek vi utgangspunkt i at digitale ordbøker er å finne på nettet i dag, og vi skil derfor ikkje mellom nemningane.

Nøkkelen til å forstå skiljet mellom tradisjonelle og moderne ordbøker, trykte eller digitale, er ordboksbasen. Basen til ei ordbok har tradisjonelt vore ei eldre ordbok, og arbeidet har bestått av å fjerne ord ein meiner er forelda, og ein har lagt til materiale frå nyare ordbøker, eigne samlingar og så vidare. I ei ordbok som stettar dagens krav, består ordboksbasen av ei samling av skriftlege og/eller munnlege tekstar, eit såkalla (tekst) korpus. Dersom korpuset statistisk sett er eksemplarisk samansett og balansert, har korpusordboka den fordel at ho speglar den faktiske språkbruken betre enn ei tradisjonelt utarbeidd ordbok gjer (Bergenholtz o.a. 1997: 161). Ut frå dette kan vi slå fast at *Islex*-ordbøkene er korpusordbøker; dei byggjer ikkje på eksisterande islandske ordbøker, men spring ut frå eit korpus samla ved Árni Magnússon-instituttet for islandske studiar: [Íslenskt textasafn](#). I 2012 inneheldt dette korpuset rundt 67 millionar løpeord. Til saman 86 prosent av ordtilfanget i *Islex* er henta frå denne tekstsamlinga, medan redaktørane av ordboka har samla 14 prosent av oppslagsorda frå andre kjelder som avis- og tidsskriftkorpuset [Tímarit.is](#) og [Ritmálssafn Orðabókar Háskólans](#) (Þórdís Úlfarsdóttir 2013: 49).

Vidare er det vanleg å skilje mellom ordboksarbeid som er korpusbasert, og ordboksarbeid som er korpusdrive. Korpusdriven tilnæringsmåte er ein metode der korpus fungerer som ein empirisk base leksikografen hentar data frå, og der dei oppdagar lingvistiske fenomen utan å stille med førehandsdefinerte hypotesar og forventningar. Alle konklusjonar og påstandar blir utan unntak utarbeidde på grunnlag av korpusobservasjonar. På verdsbasis er det berre nokre heilt få ordbøker av kvalitet som er utarbeidde med korpusdriven tilnærming. Den andre varianten er korpusbasert tilnæringsmåte. Det er ein metode der korpus fungerer som ei samling av språkdata. Frå denne kjelda kan ein trekkje ut passende materiale for å støtte intuitiv kunnskap, verifisere forventningar og la lingvistiske fenomen bli talfesta. Ein kan finne bevis for eksisterande teoriar

og finne illustrative døme. Det er ein metode der ein undersøker korpus, og data blir brukte til å bekrefte lingvistiske forklaringar og forventningar ein hadde på førehand (Tognini-Bonelli 2001: 65ff). *Islex* er eit ordboksverk som utan tvil kan plasserast i kategorien ordbøker som er korpusbaserte; ein hentar ut informasjon ein treng, og får oversikt over tydingar, men samstundes byggjer ein på eksisterande teoriar og kunnskap.

Arbeidet i målspråksredaksjonane

Korleis påverkar ein heildigital redigeringsverkstad arbeidet med målspråka? Og kva har vi i dei skandinaviske redaksjonane opplevd som dei største fordelane og ulempene ved å vere med på å utvikle eit stort fellesprosjekt på ei digital plattform som i stor grad blei utforma etter kvart som vegen blei til? Synspunkta i denne artikkelen står for forfatta- ren si eiga rekning, sjølv om mange av momenta òg er ting vi har disku- tert på fellesmøte, i andre forum og i artiklar. Dersom ein ønskjer fleire vinklingar på dette, kan ein finne artiklar skrivne av fleire av redaktørane i prosjektet i løpet av perioden 2006–2014 i litteraturlista.

Språkteknologi og målspråksarbeid

I *Islex* er det fire hovudkategoriar med informasjon på islandsk:

- oppslagsord (i underkant av 50 000, ofte med fleire tydingar)
- bruksdøme (illustrasjon av typiske samanhengar eit ord kan opptre i)
- kollokasjonar (ikkje-metaforiske uttrykk som *ta seg ein røyk og gjere sjølv mord*)
- faste uttrykk (med overført tyding som *sitje med skjegget i postkassa og som fot i hose*)

I Skandinavia har vi utrusta desse kategoriane med det ein kallar ekviva- lentar – synonym på tvers av språk (Fjeld og Vikør 2008: 202) – parafraserande forklaringar og omsetjingar. Den islandske redaksjonen sette inn forklaringar til oppslagsorda og dei fleste faste uttrykka i databasen re- daksjonsarbeidet føregjekk i, og desse forklaringane danna grunnlag for arbeidet på målspråkssida. Så langt er det ingen store forskjellar mellom *Islex*-prosjektet og meir tradisjonelle prosjekt.

Fordi den islandske redaksjonen arbeider med ei korpusbasert tilnærming, er det fleire ord i *Islex* ein ikkje kan finne i andre ordbøker. Då må målspråksredaktørane gå til det same korpuset og sjå ordet i kontekst der, og ofte er det i tillegg nyttig å bruke søkjemotorar som Google for å sjekke tydingar og bruk på Internett. Ofte kan ein få litt andre kjelder, og gjerne med meir munnleg prega språkbruk, på nettet enn i spesialtilpassa korpus. Er ordet eit sjargonguttrykk eller har munnleg preg, er det sjølv sagt ein fordel. Når blogggarar og andre nettskribentar i alle aldrar og samfunnslag hamrar i veg på tastaturet, reflekterer dei truleg i liten grad over at det dei skriv, er ei rik kjelde for leksikografar! I andre tilfelle har andre språkressursar som det største islandske korpuset [Íslenskur Orðasjóður](#) med rundt 545 millionar løpeord vore gode kompletterande kjelder – òg for oss på målspråkssida – for å få betre kjennskap til eigenskapane til det islandske ordet.

Dette kan kanskje høyrast litt omstendeleg ut, og det er heller ikkje alltid nødvendig å gå så grundig til verks, trass alt er alle som har arbeidd på målspråkssida, kompetente brukarar av islandsk og har ofte ei intuitivt forståing av det islandske oppslagsordet. Samstundes kan det vere heilt nødvendig med ei ordentleg gransking av det islandske oppslagsordet for å oppdage at det gjerne har andre valørar og funksjonar enn ein kjende til i utgangspunktet. Slike kjem gjerne heller ikkje fram i ordboksdefinisjonane vi hadde å byggje på. Med slikt grundig arbeid kan vi unngå det vi i språkvitskapen kallar «falske vener». Det er ord og uttrykk som ser så like ut i to språk at dei lett blir tolka og brukte som ekvivalente, men eigentleg har dei ulike tydingar. I nærskyldige språk som dei nordiske har vi svært mange slike. Eit banalt døme er at ein ikkje må finne på seie at det islandske ordet *koppur* er ekvivalent med ein skandinavisk *kop(p)*, like opplagt er det ikkje at *almennilegur* ikkje er ekvivalent med nynorsk *alminneleg*.

I *LexicoNordica* 17 skriv Ruth V. Fjeld og Henrik Lorentzen (2010: 10) dette om sambandet mellom språkteknologi og utvikling av ordbøker:

Forholdet mellom språkteknologi og leksikografi er svært tett i moderne tid, det er nesten utenkelig å bedrive leksikografi i dag uten å nytte språkteknologiske ressurser. På den andre siden er språkteknologien blitt avhengig av gode leksikalske beskrivelser, langt på vei er disse to fagdisiplinene gått inn i en symbiose.

Denne skildringa er enkel å kjenne seg igjen i. Ein ting er å skaffe seg oversikt over det islandske oppslagsordet, funksjonane det har, og samanhengane det går inn i, men den mest krevjande oppgåva er oftast å finne dei presise målspråksekvivalentane, i dette tilfellet å gje dei islandske oppslagsorda med underkategoriar ei form på dei andre nordiske språka. Då må ein bruke alle tilgjengelege ressursar som korpus på kjelde- og målspråk, ein- og tospråklege ordbøker i trykt og digital form på alle aktuelle språk og supplere med anna stoff på nettet. I tillegg er nettet ei viktig kjelde for å finne fram til personar og institusjonar ein kan ta kontakt med for å stille spørsmål direkte dersom orda er høgt spesialiserte og utanfor det redaktørane sjølve kan finne fram til utan å bruke urimeleg lang tid. I dag er det nesten uforståeleg korleis redaktørar før den digitale revolusjon jobba med setlar og arkiv i staden for å enkelt og greitt kunne sveipe over fleire basar og kjelder med heilt spesifikke søkjemoglegheiter. Det kan ikkje vere tvil om at det er enklare i dag å oppfylle idealet om at ordbøker skal spegle språkbruken heller enn å ha ein normerande funksjon med det rike og breie kjeldegrunnlag ein har no. Ein unngår iallfall ideelt sett å vidareformidle feil og unøyaktige ekvivalentar som gjennom tidene ofte har blitt gjentekne frå ordbok til ordbok fordi metodikken har lagt opp til det.

Det er likevel viktig å få fram at sjølv om ordbøkene som frå før eksisterte mellom islandsk og dei skandinaviske språka, ikkje fyller krava vi stiller til ordbøker i dag, og/eller sjølv om dei er små i omfang, er det ikkje tvil om at det er mykje godt arbeid som ligg bak dei. Vi har sjølv sagt hatt god støtte og hjelp i eksisterande trykte og digitale ordbøker, men alltid parallelt med korpus og andre språkteknologiske hjelpemiddel. Ein bør ikkje kaste alt det gamle over bord i rein og skjær begeistring over dei nye hjelpemidla. Vi har foreløpig ikkje så gode språkteknologiske ressursar i Norden at det utan vidare kan tilrådest. Eit reint korpusdrive ordboksprosjekt er, slik eg ser det, ikkje realistisk med dei språkteknologiske ressursane vi har tilgjengeleg i Norden i dag (jamfør rapportar skrivne i regi av [META-NET](#)).

I redaksjonsarbeidet har vi nytta ein nettbasert database som er utvikla spesielt til dette prosjektet. Det sikrar at det går an å arbeide med ordbøkene i alle dei involverte landa samstundes, og det har vore store gevinstar å hente i å arbeide parallelt med dansk, norsk og svensk. Samstundes oppstår eit nytt problem, nemlig faren for «smitte» mellom målspråka, og den faren er større dess likare språka er. Det er flust av falske vener i dei

skandinaviske språka, ofte brukte dømme er *rar*, *artig* og *grine*. Redaktørane har under arbeidet fått opp det som allereie var gjort i databasen på dei andre språka, og om nokon først hadde valt ein lite presis ekvivalent, har det skjedd meir enn ein gong at dette har blitt gjenteke av redaktørar i dei andre redaksjonane. Ikkje minst er det fare for spreining av feil når tidspresset er stort, og det har det til tider vore. Ein del slike feil er retta opp, men det er fare for at det framleis står fleire igjen.

Oppdateringar i kjelde- og målspråka blir synlege både i basen og på nett med det same endringar blir utførte. Det er ikkje sjølvstykkt at det er slik. Redaksjonen bak *Oxford English Dictionary* ([OED](#)) har valt ei anna løysing der dei oppdaterer nettversjonen av ordboka si kvar tredje månad. Dei arbeider kontinuerleg med oppdateringar, men har lagt inn denne tidsbufferen slik at dei kan lese korrektur og sjekke i fleire rundar at det ikkje har sneike seg inn feil. Dette vitnar om ein litt annan ressurstilgang og kvardag enn den *Islex* har blitt til under.

I eit prosjekt som har vore i stadig utvikling, og der nye kategoriar har kome til etter kvart, har det vore vanskeleg å vere stringent i kategori-ane i dei nasjonale redaksjonane – og vi har nok heller ikkje klart å vere heilt konsekvente på tvers av landegrensene. Det har ført til at informasjon om det islandske oppslagsordet på dei skandinaviske språka av og til kjem før ekvivalentane (som i [jólasveinn](#)), men dessverre ikkje alltid (som i [föðurbróðir](#)). Ein ny kategori blei innført ganske seint i prosjektperioden, og i den norske redaksjonen resulterte det i ein regel om at dersom det var det islandske oppslagsordet som blei kommentert, skulle informasjonen kome før ekvivalentane. Var det dei norske ekvivalentane vi kommenterte, skulle informasjonen kome etterpå. Logikken og systematikken burde vere grei, men vi har ikkje hatt høve til å gå gjennom alle tidlegare artiklar for å endre og samkøyre dette. Det er nok ikkje i alle tilfelle heller krystallklart om det er oppslagsordet eller ekvivalenten som blir kommentert (jamfør [stærðfræðideild](#)). På tilsvarande måte er informasjon om dei skandinaviske ekvivalentane som er mest retta mot islandske brukarar, nokre gonger på islandsk, men ikkje alltid.

Utfordringa med å halde tunga beint i kategoriane blir forsterka av at det digitale mediet gjev oss høve til å bruke meir plass og kommentere meir enn trykte ordbøker tillèt. Å gå gjennom ordbøkene og sjå til at omsetjingar og løysingar er konsekvente er dømme på etterarbeid som vi gjerne skulle gjort meir av i dei skandinaviske redaksjonane. Samstundes er det usikkert kor stort problem dette faktisk utgjør for brukarane.

Det kan utan tvil vere forvirrande med ulike løysingar dersom ein har alle målspråka oppe samstundes (jamfør [allsherjargoði](#)), men det veit vi ut frå statistikk at dei færreste brukarar har. I tillegg må ein ikkje gløyme at dei fleste artiklane ikkje har slike markeringar, så truleg er dette ikkje eit på-trengjande problem for den jamne brukar.

Ulike føresetnadar og litt ulike produkt

Arbeidet i dei nasjonale redaksjonane blei i hovudsak finansiert av styresmaktene i dei respektive landa, og mandatet har vore å lage eit best mogleg produkt for eige språk. Ein slik måte å finansiere prosjektet på fører til ein tradisjonell arbeidsmåte der kvar og ein målspråksredaksjon i all hovudsak har arbeidd med og fokusert på å finne gode ekvivalentar og omsetjingar på eige språk. Alle *Islex*-ordbøkene er unike og med eige særpreg trass i god kommunikasjon og felles base i kjeldespråket.

Om ein deltek på internasjonale konferansar, oppdagar ein fort at vi i Norden er heldige som har styresmakter som er villige til å bruke pengar på slike fellesprosjekt fordi dei ser at ei ordbok ikkje berre er ei ordbok. Det er den direkte årsaka til at *Islex*-ordbøkene er gratis, men at ein likevel har kunne leggje seg på eit høgt ambisjonsnivå. På den andre sida har det vore store forskjellar i kor store løyvingar dei nasjonale redaksjonane har fått frå styresmaktene sine, og det har ført til at prosessane i dei ulike landa har vore litt ulike. Både talet på tilsette i dei ulike redaksjonane, økonomiske rammer og målsetjingar har derfor variert i dei skandinaviske landa. Som døme fekk den svenske redaksjonen fekk ei relativt stor løyving tidleg i prosjektperioden, men samstundes var det ein føresetnad at dei skulle drive leksikografisk utviklingsarbeid. Dette har ført til at den svenske redaksjonen har hatt eit fraseprosjekt gåande parallelt med det andre ordboksarbeidet (Anna Helga Hannesdóttir o.a. 2010), og i mange artiklar kan ein sjå at dei har med meir fraseologisk informasjon enn dei andre språka (jamfør [brauðsneið](#)).

I den norske redaksjonen har vi hatt nokre ekstra utfordringar og fordelar i og med at norsk har to målformer (bokmål og nynorsk). Dette kan vere krevjande når ein skal finne ekvivalentar til dei faste uttrykka. Dette er behandla i ei masteroppgåve frå Universitetet i Bergen (Rauset 2010a) og ein konferanseartikkel (Rauset 2010b). To målformer tyder ikkje at det er dobbelt opp med arbeid, kjeldespråksanalysen tek ein berre ein gong, men det seier seg sjølv at det har teke lengre tid enn om norsk berre blei skriva på ein måte. På den andre sida har det òg sine fordelar, mellom

anna var det best å ta dei nynorske ekvivalentane før bokmåsekvivalentane om ein ville prøve å unngå å bli påverka av dansk om artikkelen alleie var redigert på det språket (Rauset 2011).

Sjølv om den islandsk-danske, islandsk-norske og islandsk-svenske ordboka hadde offisiell fellesopning på Island i november 2011, heldt den norske redaksjonen på for fullt til januar 2013, og framleis står det att å omsetje ein god del bruksdøme til norsk, sjølv om alle oppslagsord, kollokasjonar og faste uttrykk er utrusta med ekvivalentar. Det digitale formatet gjorde at det ikkje representerte eit problem for dei andre redaksjonane at vi ikkje blei ferdige samstundes med dei, og dei norske brukarane fekk opp svenske og danske ekvivalentar der dei i ein periode mangla på norsk. Likeeins kom den danske redaksjonen ikkje heilt i mål, men der det er høl i det danske materialet, får danske brukarar opp ekvivalentar på dei andre målspåka. Både den danske og norske redaksjonen treng litt meir midlar til ferdigstilling, samkøyring, korrektur og finpuss, men det er ei viss trøyst at brukarane våre enkelt kan få supplerande informasjon om dei skiftar over til dei andre språka der det er høl.

Dei skandinaviske redaksjonane har valt litt ulike løysingar for å presentere dei grammatiske opplysningane knytte til ekvivalentane i dei ulike målspåka. Held ein musa over ein skandinavisk ekvivalent, skal ein få opplysningar om bøyingsmønsteret til ordet. Det ser litt ulikt ut i dei ulike *Islex*-ordbøkene, og det skuldast mellom anna at dei stammar frå ulike hjelperessursar.

Islex-ordbøkene skal stette krava til ulike brukargrupper, og dei ulike redaksjonane har hatt litt ulik tilnærming til denne problemstillinga. Kva behov ein brukar har for ekstraopplysningar, om det er dei islandske oppslagsorda eller dei skandinaviske ekvivalentane som treng forklaring, og kva språk det skal skje på, varierer alt etter om brukaren er islending eller skandinav, og kva nivå han er på. Den norske redaksjonen har måtte prioritere å få på plass dei to målformene, medan dei i Danmark og særleg i Sverige har gått lenger i å kommentere dei danske og svenske ekvivalentane for dei islandske brukarane (jamfør [fertugsafmæli](#)).

Samarbeid mellom kjelde- og målspåk

Den islandske *Islex*-redaksjonen la materialet til rette ved å dele oppslagsorda inn i ordklasser og igjen ned i mindre grupper der alle orda hadde same tematikk. Det redaksjonelle arbeidet med slike semantiske felt førte til auka oversikt over ord med sterk tematisk samanheng og gjorde det let-

tare å få større samsvar mellom artiklane. Oppgåva til ein ordboksredaktør er å analysere innhaldet i kjeldespråkseininga og opprette ein ekvivalensrelasjon mellom dette innhaldet og einingar i målspråket. Dette arbeidet kan resultere i ein eller fleire ekvivalentar eller ein parafrase (Anna Helga Hannesdóttir 2012: 39). Det semantiske feltet *rus* i *Islex* kan tene som døme. Det inneheldt alle stadium av *rus*, frå *brisen* og *på ein snurr* til *dritings* og *kanon full*. Det er 56 ord i dette semantiske feltet, dei fleste er adjektiv. Når ein kan arbeide med dei systematisk og i samanheng, blir det enklare å velje ekvivalentar som representerer likskapar mellom omgrepa på dei to språka.

Det har vore eit tett samarbeidet mellom kjelde- og målspråksredaksjonane i utviklinga av *Islex*-ordbøkene, og ein av fordelane med det, er at målspråksredaksjonane har fått kome med innspel til korleis kjeldespråket kan utviklast. Vi har kome med innspel til oppslagsord, bede om å få bruksdøme til ord og tydingar og kome med innspel til endringar i bruksdøma (Halldóra Jónsdóttir og Þórdís Úlfarsdóttir 2012: 356). Det har vore ein dynamisk prosess, til dømes slutta den islandske redaksjonen å bruke dei høgfrekvente islandske orda *frændi* og *frænka* i bruksdøma fordi dei skandinaviske språka ikkje har tilsvarende sekkeomgrep for mannlege og kvinnelege slektningar. I eit bruksdøme som eigentleg skal illustrere noko heilt anna, kan det vere forstyrrande med slike moment, og det er meir formålstenleg å bruke andre substantiv.

I *Islex* er kjeldespråksmaterialet tenkt å skulle kunne brukast til kva språk som er på målsida – sjølv om vi i Skandinavia altså har fått vere med og påverke ein del. Det vanlege i to-språklege ordboksprosjekt er at framstillinga av målspråket påverkar framstillinga av kjeldespråket, og at ein tek med og legg vekt på sider ved kjeldespråket som særleg valdar problem eller er annleis på målspråket. Når ein ikkje har det komparative blikket, men utviklar ein kjeldespråksbase som i teorien skal kunne brukast til kva språk som er på målsida, kan t.d. tydingsinndelingane stå fram som umotiverte (Rauset o.a. 2012: 514). Særleg problematisk blir dette ofte i dei kontekstlause bruksdøma, der ønskjer ein fokus på oppslagsordet, og kva andre ting som berre valdar «støy» i omsetjinga av dei, er heilt avhengig av målspråka (Hellerud 2012). Ein kan derfor tenkje seg at den færøyske, men kanskje særleg den finske redaksjonen kunne ønskje å endre på enkelte bruksdøme, men det har dei har ikkje høve til sidan dei skandinaviske redaksjonane i hovudsak har avslutta sine prosjekt, og kjeldespråksinnhaldet må vere likt i alle ordbøkene.

Det tek sjølvsagt tid å drive utviklingsarbeid. Tekniske løysingar og kategoriar har endra seg etter kvart, og det har ført til at måten målspråksredaksjonane har strukturert innhaldet sitt på, har endra seg gjennom prosjektperioden. Tidvis har ein måtte gå tilbake og endre på det ein har gjort tidlegare, men det har ikkje vore tid i nokon av målspråksredaksjonane til å gjennomføre dette konsekvent. Det fører til at målspråka kan ha litt ulik struktur alt etter kor tid redaksjonane har arbeidd med dei ulike artiklane. På den måten har dei som utarbeider dei færøyske og finske *Islex*-ordbøkene i dag eit meir oversiktleg og føresjåeleg arbeid, og dei har dermed eit betre utgangspunkt for å ha eit konsistent oppsett og ein stringent bruk av kategoriane i artiklane sine.

Ein positiv effekt av fellesprosjekt av denne typen der redaksjonane held til i fleire land, er at ein kan hente kompetanse frå dei ulike institusjonane redaksjonane har vore knytt til. Til dømes kom den språkteknologiske staben ved Det Danske Sprog- og Litteraturselskab med gode innspel til utforming av brukargrensesnittet, medan vi i heile prosjektperioden har drege vekslar på at redaktørane i dei ulike landa har ulik bakgrunn og har bidrege med innspel til andre der dei sit på spesialkompetanse. I Noreg fekk vi gjennomført eit delprosjekt der fagpersonar som kan både islandsk og norsk, las korrektur på orda i sine fagfelt. I *Islex* er fagtermar markerte som del av eit fagfelt som medisn, geologi, kjemi, matematikk, metrikk, filosofi og så vidare. Ein del av tilbakemeldingane etter denne runden kunne dei andre redaksjonane òg dra nytte av.

I ein heildigital kvardag må ein likevel ikkje undervurdere den direkte kontakten. Det har vore avgjerande for det gode samarbeidsklimaet at vi har blitt godt kjende med kvarandre gjennom to årlege møter for alle *Islex*-redaktørane i Skandinavia og Island. Styregruppa i prosjektet har vore med på eitt av to av desse årlege fellesmøta. Medlemmane der har då diskutert dei store linjene: spørsmål knytte til rettane til materialet, vidare utvikling og så vidare. Styregruppa har hatt ansvar den formelle kontakten mellom samarbeidsinstitusjonane, og medlemmane er òg offisielle representantar for sine institusjonar i prosjektet (Þórdís Úlfarsdóttir 2013: 44). I tillegg hadde dei skandinaviske redaksjonane fleire mindre arbeidsmøte i 2010 og 2011. Dette var gode høve til å reflektere rundt eigen praksis, sjå ting litt utanfrå og diskutere seg fram til fellesløysingar. Sjølv om ordbøkene er sjølvstendige produkt som dei nasjonale redaksjonane står til ansvar for, og redaksjonane til ei viss grad har valt dei løysingane dei meiner er mest tenlege for eigne språk, gjer det digitale mediet at brukarane

kan sjå dei i samanheng, og det har i stor grad ført til at ein likevel har søkt å finne fram til fellesløysingar. I *Islex*-databasen blei det òg utvikla eit internt meldingssystem som blei mykje brukt som supplement til e-post i produksjonsfasen. Der kunne ein spørje kvarandre til råds og peike på ting som andre gjerne burde sjå på igjen, både internt og på tvers av redaksjonane.

Brukargrupper og ulike forventningar til ei digital ordbok


Vi har brukt mykje tid på fellesmøta til å diskutere kven som er brukargruppene til *Islex*-ordbøkene, og korleis det påverkar innhald og tilrettelegging. Som islandsk-skandinaviske ordbøker skal dei oppfylle krava til svenske, norske og danske brukarar, mellom anna til omsetjing frå islandsk og når skandinavar skal lære islandsk. Dessutan vil islandske brukarar ha nytte av henne når dei skal uttrykkje seg på eit skandinavisk språk. Heimesida til prosjektet opnar for ulike søkjemåtar, mellom anna kan brukarane velje om dei vil få ekvivalentane på eitt eller fleire målpråk samstundes. Dermed kan ein både få innsyn i samanhengen mellom islandsk og eit særskilt målpråk og oversikt over den innbyrdes samanhengen mellom dei skandinaviske, og i framtida resterande nordiske språka.

Brukarar kan enkelt endre språk og innstillingar etter ønskje. Det gjev ein fleksibilitet som er heilt særeigen for det digitale mediet. Dette er i hovudsak positivt, men samstundes bør brukarane vere klare over at ordboka ikkje er utvikla og kvalitetssikra som ei t.d. svensk-dansk ordbok. Er det konkrete ord eller eintydige termar, kan *Islex* likevel fungere på denne måten. Ein norsk brukar kan slå opp i *Islex* for å finne ut at *tyttebær* er det same som [lingon](#) på svensk og at *blodpropp* er ekvivalent med [sveitalivur](#) på færøysk. I tillegg kan brukarane stille inn på fritekstsøk som kan gje eit endå meir utfyllande bilde av alle samanhengar ord og frasar førekjem i i *Islex*. Det digitale mediet kan føre til andre forventningar hos brukarane enn vi har arbeidd ut frå, nemleg at islandsk er kjeldespråk og at dei skandinaviske språka er målpråk. *Islex* kan opplevast ikkje berre som ei samling tospråklege ordbøker, men som ei samling tospråklege, bidireksjonale og multifunksjonelle ordbøker (Rauset o.a. 2012). Vi har ikkje kvalitetssikra på tvers av målpråka, men her ligg det samstundes eit stort og uutnytta potensial!

vel ordbok:     

kría subst f


 [uttale](#)


 [BØYING](#)


(lat. *Sterna paradisaea*)




 www.fauna.is


 [havterne](#)

 [rødnebbterne](#)

 [raudnebbterne](#)

 [silvertärna](#)

 [terna](#)

 [lapintiira](#)

Det kan iallfall diskuteras om ein har late seg freiste til å inkludere for mykje informasjon her. Det er godt mogleg at brukaren som slår opp dette ordet, kjenner fenomenet og berre treng ein ekvivalent, iallfall om han er islending, og i så fall må han skumme gjennom ei linje med lenkje til uttale, ei lenkje til fullt bøyingsparadigme, ei linje med informasjon om korleis ordet er samansett, informasjon om at dette er ein juridisk term som er gammal (dvs. utdatert) før det kjem ei norsk forklaring av det islandske fenomenet, før ekvivalenten til slutt kjem. Vi ønskjer sjølvsagt alt det beste for brukarane våre, og mange vil nok heilt sikkert setje pris på ein del slik tilleggsinformasjon, men kan dette også vere grunnen til at nokre brukarar tenkjer at dei enklare finn svaret på det dei lurar på, ved å bruke Google omsetjar (sjølv om resultatet der i dette konkrete dømet er det nokså håplause 'by dommen')?

På den andre sida såg ein nok ikkje føre seg ved oppstarten av prosjektet i 2006 at ein motstridande tendens kom til å gjere seg gjeldande: Bruken av smarttelefonar og lesebrett har eksplodert, og på dei plattformene er det gunstig med ei så komprimert form som mogleg, særleg på telefonane. *Islex* blei difor programmert for operativsystema Android og iPhone på vårparten i 2012, men innhaldet i artiklane er det same som når ein går inn på ei vanleg datamaskin, sjølv om mykje av omteksten er kuttet. Kostnaden ved å gjere dette var nokså høg, men dette er ikkje ein stad å spare dersom vi ønskjer å halde på brukarane våre. *Islex* er i utgangspunktet redigert for bruk på tradisjonelle datamaskiner, og det kan vere behov for vidare diskusjonar i styregruppa om korleis ein skal jobbe vid-

are med materialet for bruk på nye plattformer. Med farten på utviklinga i dag, er dette problemstillingar som må prioriterast.

Nettordbøker står i større fare enn tradisjonelle, trykte ordbøker for å bli raskt utdaterte og ubrukelege. Ei bok kan ein trekkje ut av hylla etter hundre år, innhaldet er der sjølv om det delvis er forelda, men det treng ikkje ta mange år før maskiner og programvare får problem med å „lese“ nettordbøker om dei om ikkje blir oppdaterte og vidareutvikla. Og i ein slik samanheng treng ein òg tenkje på innhaldet, ikkje berre innpakninga. Ein kan godt tenkje seg at brukarane forventar at nye ord og vendingar kjem raskare inn i digitale ordbøker enn trykte. Det er sjølvsagt ein av fordelane med nettordbøkene at dei stadig kan bli reviderte og forbetra. Er ein på den andre sida for rask med å ta inn nyord, risikerer ein at orda ikkje festar seg i språkbruken, og at ein på den måten òg kan verke utdatert.

Det har vore diskutert å lage ein *Islex*-app. Men skal ein utvikle ein slik, kan det hende ein lyt tenkje nytt i form av kva kategoriar som skal vere synlege for brukarane. Smarttelefonar blir ofte brukte når ein vil finne ut av noko i ein fart, og kanskje er det nok med sjølve tydingane, slik at ein kan kutte bruksdøme, kollokasjonar og faste uttrykk? Eventuelt at brukarane aktivt må velje å få opp meir informasjon dersom det er det dei er ute etter. Dette vil vere ei trekkspekløysing. Det kan vise seg at ulike plattformer krev ulikt innhald, og framtidige prosjekt må nok i endå større grad tenkje fleirplattformsløysingar parallelt. Det positive er at dess fleire plattformer ein er tilgjengeleg på, dess fleire brukarar og brukargrupper kan vi nå.

Ved alle ord som kan bøyast, kan brukarane finne fulle bøyingsparadigme for islandsk gjennom ei lenkje til prosjektet [Beygingarlýsing íslensks nútímamáls](#), som også blir drive hjå Árni Magnússon-instituttet for islandske studiar. Norske studentar som lærer islandsk, gjev tilbakemelding om at den tette koplinga mellom ordbok og grammatikk er svært viktig og nyttig – og det har sjølvsagt samanheng med den komplekse islandske grammatikken. Ein kan berre spekulere på om dette hadde vore ein like viktig funksjon i språk med mindre komplekse bøyingsmønster. Uansett er dette er eit døme på ein ekstrarfunksjon som er vanskeleg å få til utan det digitale mediet, og som er ekstra godt å ha på plass i eit språk som islandsk. At det går an å søkje etter islandske ord i andre kasus og former enn oppslagsforma, er også gull verdt. Det kan vere svært krevjande for ein skandinavisk brukar å analysere seg fram til kva som er oppslagsform av enkelte islandsk ord på grunn av bøyingsmangfaldet. Dette viser

at kva funksjonar som er vesentlege å inkludere i digitale ordbøker, er avhengig av språka sjølve.

Vidare bruk og utvikling av *Islex* i språkteknologisk samanheng

Ein av mange fordelar med den digitale plattformen er at *Islex* ikkje berre i teorien, men òg i praksis har blitt utvida som prosjekt dei siste åra i og med at færøysk og finsk har kome med på målspråkssida. Dette er sjølv sagt viktig for brukarane av dei aktuelle ordbøkene, men det er òg eit poeng i seg sjølv at ein får løfta fram heile breidda av dei nordiske språka. Det styrkjer *Islex* som ein utruleg spennande og viktig base for framtidig forskning på språk i Norden. Ordbøkene illustrerer godt kva som er likt og ulikt i dei ulike språka, ikkje minst på det leksikalske planet, og *Islex* burde kunne fungere godt som eit hjelpemiddel til å auke det språklege medvitet og språkforståinga i Norden.

Alle dei nasjonale redaksjonane har eigedomsrett over det dei sjølve har produsert, men både den islandske, danske, norske og svenske redaksjonen har takka ja til at materialet blir gjort tilgjengeleg for forskning. [META-NORD](#) var eit nordisk og baltisk prosjekt som fram til januar 2013 arbeidde med å «kartlegge og katalogisere språkteknologiske ressurser og verktøy med gjenbruksverdi for forskning og utvikling innan akademia og industri», og *Islex* er ein av mange ressursar som har blitt katalogisert gjennom [META-SHARE](#), ein fritt tilgjengeleg katalog med oversikt over rettar og eventuelt med lenkje til der ein kan laste ned ein ressurs. Den islandske META NORD-gruppa har gjort det mogleg å laste ned heimesida og databasen til *Islex* i LMF (Lexical Markup Framework)-format (Sigrún Helgadóttir og Eiríkur Rögnvaldsson 2013: 69). Det tilgjengelege materialet er då alle dei islandske kategoriane (oppslagsord, bruksdøme, kollokasjonar og faste uttrykk) ekvivalentar og omsetjingar av alt dette til dei skandinaviske språka. I tillegg kan ein finne grammatisk informasjon og lydspor med oppslagsorda. Ein slik gjenbruk av materialet er sjølv sagt ikkje noko ein tenkjer på til kvardags som redaktør, men alle redaktørane stilte seg sjølv sagt bak at materialet skal kunne brukast til forskingsformål. Det var elles ein av medlemmane av den islandske META NORD-gruppa, Sigrún Helgadóttir, som i eit møte med dei skandinaviske og baltiske kollegaene sine kalla *Islex* eit skuleeksempel på godt språkteknologisk samarbeid, jamfør tittelen på denne artikkelen.

Den norske redaksjonen har ei Facebook-side (som også heiter [islex](#)).

[no](#), gå gjerne inn og «lik» oss der!) der vi legg ut eit nytt ord frå *Islex* kvar veke. Dei som følgjer oss der, kjem frå alle dei nordiske landa, og det støttar oppfatninga vår om at mange brukarar av *Islex* ikkje berre er interesserte i sitt eige språk, men gjerne ser på samanhengen mellom dei. Det er òg eit forum som opnar opp for innspel frå publikum, men det har vi så langt ikkje sett mykje til. Det er vanlegare at brukarane gjev tilbakemelding gjennom «kontakt oss»-funksjonen på heimesidene til *Islex*.

Det er ikkje ideelt at *Islex* inneheld fleire nesten-ferdige produkt som manglar finansiering til å kome heilt i mål, kanskje er dette eit problem fleire reine digitale produkt støyter på. Det er utfordrande å få fullfinansiering til ein ordentleg korrekturrunde og ferdigstilling så lenge produkta i hovudsak ser greie ut. Det kan gjerne vere eit fenomen fleire kjenner til, at det er enklare å få midlar til å starte opp nye og spennande prosjekt enn å ferdigstille prosjekt som gjerne er sett i gang av andre. Manglande korrekturlesing er eit problem, men heldigvis blir iallfall ekvivalentane til oppslagsorda språkvaska når vi legg til morfologiske opplysningar til desse. Det er verre med bruksdøma, kolleksjonane og dei faste uttrykka, og faren for feilaktig påverknad frå dei andre skandinaviske språka er ikkje mindre der.

Det går føre seg spennande drøftingar om vidareutvikling av *Islex*-prosjektet, det hadde vore veldig spennande å prøve å «snu» ei eller fleire av ordbøkene, slik at dei blir tilrettelagde for bruk frå skandinavisk til islandsk. Det vil sjølvsagt krevje manuell gjennomgang og revisjon av alt materialet, men det som allereie er gjort, vil vere eit solid materiale til å byggje vidare på. Somme brukarar har nok i dag blitt skuffa over å ikkje å finne ord i *Islex* som ikkje er frekvente i islandsk, men som er vanlege i andre nordiske språk. Det ein må ha i mente er at ordtilfanget i *Islex* i dag er basert på det islandske språket og kva ord som blir brukte i ein islandsk kontekst. Ei ordbok som til dømes har eit av dei skandinaviske språka som kjeldespråk, vil delvis måtte ha med andre oppslagsord og andre uttrykk.

Eit parallellkorpus, ei samling av dei same tekstane på fleire språk, kunne ha vore ein svært god språkteknologisk ressurs i arbeidet vårt med *Islex*. Då kan ein tenkje seg ei samling av omsett skjønnlitteratur, tekstar frå Nordisk råd og Nordisk ministerråd som blir omsette til alle nordiske språk, brukarretteiingar og alle andre typar tekstar som blir utarbeidde for fleire eller alle dei nordiske språka i ulike samanhengar. Det har vore gjort spede forsøk på å utvikle eit nordisk parallellkorpus (Jansson 2012),

men ingen er ferdigutvikla. Dersom *Islex*-prosjektet skal vidareutviklast, er eit parallellkorpus etter mi meining noko ein burde vurdere om ein skulle laga.

Eit meir fundamentalt spørsmål er om det er behov for ordbøker når kvar og ein kan bruke Google som korpus (jamfør Hellerud 2012). For ein del svært kompetente språkbrukarar kan svaret vere nei. Ingen andre er i nærleiken av å byggje dei språkteknologiske ressursane sine på eit like stort materiale som Google, men det er ikkje tvil om at resultatet ofte blir både underleg og feilaktig når ein brukar Google omsetjar mellom språk med så få brukarar som islandsk og norsk. Tekstgrunnlaget for desse språka på nett blir for lite til at resultatata er til å stole på. Det er språka med mange brukarar, og då særleg engelsk, som er vinnarane i ein sjølvforsterkande prosess, òg i denne samanhengen. For dei fleste er ordbøkene gode og tidssparande hjelpemiddel til betre språkbruk og forståing. Dessutan er visuelt gode ordbøker som *Islex* ein underhaldande måte å utvide ordforrådet på, utforske både eige og andre språk og sjå nye samanhengar.

Konklusjon

Islex-ordbøkene er bra, og dei får gode tilbakemelding frå brukarane, men like fullt har ordbøkene og prosjektet eit stort potensial for vidareutvikling. Det står att ein del redigerings- og etterarbeid med dei allereie eksisterande ordbøkene, men i tillegg ligg det tallause moglegheiter for å utvikle ordbøkene på tvers av målspåka, eller å prøve å snu ei av ordbøkene frå eit skandinavisk språk til islandsk. Innhaldet i ordbøkene famnar allereie i dag så breitt at det burde vere rom for mange forskingsprosjekt og utvikling av tilleggsressursar.

Margunn Rauset er stipendiat ved Universitetet i Bergen og ordboksredaktør i den islandsk-norske *Islex*-ordboka. I tillegg har ho dei siste åra vore med å redigere dei einspråklege verka *Nynorsk ordliste* og *Ordbok for grunnskulen* og arbeidd som omsetjar frå islandsk til norsk.

Summary

Islex is a “born digital” project consisting of multiple online dictionaries gathered on one site. Icelandic is the source language and from the beginning the three Scandinavian languages – Swedish, Norwegian (with two

official standards) and Danish – were target languages. More recently Faroese and Finnish have been included as target languages. The site is publicly available on the Internet, free of charge. In this paper I discuss the different ways language technology and the digital medium have affected the editorial work and the online dictionaries.

Litteratur

- Aldís Sigurðardóttir, A. Helgadóttir, H. Jónsdóttir, H. Jansson, L. Trap-Jensen, Þ. Úlfarsdóttir, 2008: ISLEX – an Icelandic-Scandinavian Multilingual Online Dictionary. I *Proceedings of the XIII Euralex International Congress*, red. av Elisenda Bernal og Janet DeCesaris. Barcelona, Universitat Pompeu Fabra: <http://islex.lexis.hi.is/islex/greinar/Euralex2008-ISLEX.pdf>
- Anna Helga Hannesdóttir, J.H. Jónsson, S. Tingsell, 2010: Mot en begrepps-baserad isländsk och svensk fraseologisk ordbok. Reflektioner kring pragmatiska idiom. I *Nordiska studier i lexicografi 10: Rapport från Konferensen om lexicografi i Norden*, red. av Harry Lönnroth og Kristina Nikula. Tammerfors: Nordiska föreningen för lexicografi.
- Anna Helga Hannesdóttir, 2012: Ekvivalensrelationer i tvåspråkig lexicografi. I *LexicoNordica 19*, red. av Henrik Lorentzen og Emma Sköldb-berg. Oslo: Nordisk forening for lexicografi.
- Anna Helga Hannesdóttir. Under utgjeving. Tvåspråkig lexicografi – ny teknik, nya funktioner. Ordboken som parallellkorpus. I *Framtidens svenska lexicografi* i serien Meijerbergs arkiv för svensk ordforskning.
- Anna Helga Hannesdóttir. Under utgjeving. Lemman och ekvivalenter i nya roller – en reviderad ordbokstypologi“. I *Nordiska studier i lexicografi 13: Rapport fra Konferansen om lexicografi i Norden*. Oslo: Nordisk forening for lexicografi
- Bergenholtz, Henning, I. Cantell, R. Vatvedt Fjeld, D. Gundersen, J.H. Jónsson, B. Svensén, 1997: *Nordisk lexicografisk ordbok*. Nr. 4 i Skrifter ut-gitt av Nordisk forening for lexicografi. Oslo: Universitetsforlaget.
- Fjeld, Ruth Vatvedt og Lars S. Vikør, 2008: *Ord og ordbøker: ei innføring i leksikologi og lexicografi*. Kristiansand, Høyskoleforlaget.
- Halldóra Jónsdóttir og Þórdís Úlfarsdóttir, 2008: ISLEX: Islandsk-skandi-navisk webordbog. I *Rapport från den 9:e konferensen om lexicografi i Norden*, red. av Ásta Svavarsdóttir o.a.. Akureyri 2007: Föreningen för Lexicografi i Norden.

- Halldóra Jónsdóttir og Þórdís Úlfarsdóttir, 2012: ISLEX - en flersproget nordisk ordbog. I *Nordiska studier i lexicografi 11: Rapport från Konferensen om lexicografi i Norden*, red. av Birgit Eaker, L. Larsson, A. Mattisson. Lund: Nordiska föreningen för lexicografi.
- Hellerud, Ylva, 2012: Ensam fras søker kontekst: Översättarens hantering av språkprov i en tvåspråkig ordbok. I *Nordiska studier i lexicografi 11: Rapport från Konferensen om lexicografi i Norden*, red. av Birgit Eaker, L. Larsson, A. Mattisson. Lund: Nordiska föreningen för lexicografi.
- Jansson, Håkan, 2012: Parallellkorpusen som resurs i lexicografiskt arbete. I *Nordiska studier i lexicografi 11: Rapport från Konferensen om lexicografi i Norden*, red. av Birgit Eaker, L. Larsson, A. Mattisson. Lund: Nordiska föreningen för lexicografi.
- Lilljegen, Joakim, 2011: Idiomekvivalens i tvåspråkiga ordböcker: En undersökning av svensk-isländsk och isländsk-svensk lexicografi. Magisteruppsats. Göteborg: Göteborgs universitet: https://gupea.ub.gu.se/bitstream/2077/26356/1/gupea_2077_26356_1.pdf
- Ljunggren, Anders, 2011: Från isländsk-svensk ordbok till isländsk-skan-dinavisk nätordbok. Några skärvar som kan fogas till en bild av vägen fram till ISLEX tillkomst: <http://islex.lexis.hi.is/islex/greinar/Ljunggren-nov2011.pdf>
- Rauset, Margunn, 2010a: Sorte får og svarte sauer. *Ekvivalensproblematikk og målformskilnader ved ordsamband i den islandsk-norske delen av ordboka ISLEX*. Masteroppgåve. Bergen: Universitetet i Bergen: <https://bora.uib.no/bitstream/handle/1956/4073/69434698.pdf?sequence=1>
- Rauset, Margunn, 2010b: Utfordringer med to norske målformer i ISLEX. I *Nordiska studier i lexicografi 10: Rapport från Konferensen om lexicografi i Norden*, red. av Harry Lönnroth og Kristina Nikula. Tammerfors: Nordiska föreningen för lexicografi.
- Rauset, Margunn, 2011: Islandsk møter norsk i eit nytt tusenår. Utfordringer og fordelar i den norske delen av ISLEX-prosjektet. I *Vestnordisk språkkontakt gjennom 1200 år*, red. av Gunnstein Akselberg og Edit Bugge. Tórshavn: Fróðskapur. Faroe University Press.
- Rauset, Margunn, A.H. Hannesdóttir, A. Sigurðardóttir, 2012: Ein-, to- eller fleirspråkleg ordbok? I *Nordiska studier i lexicografi 11: Rapport från Konferensen i lexicografi i Norden*, red. av Birgit Eaker, L. Larsson, A. Mattisson. Lund: Nordiska föreningen för lexicografi.

- Sigrún Helgadóttir og Eiríkur Rögnvaldsson, 2013: Language Resources for Icelandic. I *Proceedings of the workshop on Nordic language research infrastructure at NODALIDA 2013*. Oslo University, Norway: <http://www.ep.liu.se/ecp/089/006/ecp1389006.pdf>
- Svavar Gestsson, 2011: Ævintýri handa öllum heiminum: http://islex.lexis.hi.is/islex/greinar/avarp_SG-nov11.pdf
- Tognini-Bonelli, Elena, 2001: *Corpus linguistics at work*. Amsterdam: Benjamins.
- Þórdís Úlfarsdóttir, 2012: Upphaf ISLEX-verkefnisins: http://islex.lexis.hi.is/islex/greinar/Islex_upphaf.pdf
- Þórdís Úlfarsdóttir, 2013: ISLEX - norræn margmiðlunarorðabók. I *Orð og tunga 15*, red. av Ásta Svavarsdóttir. Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum: <http://islex.lexis.hi.is/islex/greinar/GreinISLEX.pdf>

Nettressursar

- Beygingarlýsing íslensks nútímamáls: <http://bin.arnastofnun.is/leit/>
- Dansk Islex-side: <http://www.islex.dk>
- Íslandsk Islex-side: <http://www.islex.is>
- Íslenskt textasafn: <http://corpus.arnastofnun.is/?%C3%8Ds-lenskt+textasafn=Leit+%C3%AD+textasafni>
- Íslenskur Orðasjóður: http://wortschatz.uni-leipzig.de/ws_isl/
- META-NET: <http://www.meta-net.eu>
- META-NORD: <http://www.meta-nord.eu>
- META-SHARE: <http://metashare.tilde.lv>
- Norsk Islex-side: <http://www.islex.no>
- Oxford English Dictionary (OED): <http://www.oed.com>
- Ritmálssafn Orðabókar Háskólans: <http://www.arnastofnun.is/page/ritmalssafn>
- Svensk Islex-side: <http://www.islex.se>
- Tímarit.is: <http://timarit.is>

Sprogværktøjet Frasar.net

Om fraser og fraseindlæring anskuet kontrastivt

Auður Hauksdóttir

Sprogværktøjet www.frasar.net er udviklet i samarbejde mellem islandske sprogforskere og danske og svenske datalingvister og sprogteknologer¹. Hensigten med sprogværktøjet er at fremme islændinges indlæring og brug af dansk, og samtidig er der blevet lagt vægt på, at det kan gavne danskere, der ønsker at kommunikere på islandsk. Ved udviklingen af fraseværktøjet er der blevet taget udgangspunkt i aktuelle teorier om fremmedsprogsundervisning, især om ordforråd og ordforrådsindlæring og ny viden inden for datalingvistik og sprogteknologi. Desuden tages der højde for den viden, som forskningen i dansk som fremmedsprog giver om islændingenes indlæring og brug af dansk (fx Auður Hauksdóttir 2001 og 2012a). Sprogværktøjet indeholder godt 7600 fraser på dansk af typerne billedelige og ikke billedlige idiomer, kollokationer, sammenligninger og kommunikative formler. På baggrund af komparative studier i denne del af ordforrådet på dansk og islandsk gives der oplysninger om en islandsk pendant til den danske frase, hvis den findes, eller alternativt en betydningsforklaring af den danske frase på islandsk. Endvidere indeholder sprogværktøjet komparative iagttagelser om lumske ligheder og forskelle mht. frasernes betydning, bøjning og brug. Sprogværktøjet muliggør hurtig kobling til danske korpora, hvor man kan iagttage frasernes forekomster i kontekst. I det følgende gøres der kort rede for det teoretiske udgangspunkt for fraseværktøjet og derefter for dets anvendelsesmuligheder og funktioner.

Kommunikativ kompetence som mål

Inden for fremmedsprogsundervisningen har teorier om kommunikativ

¹ Projektet ledes af Auður Hauksdóttir, docent i dansk på Islands Universitet. De vigtigste samarbejdspartner har været Guðrún Haraldsdóttir, Islands Universitet, Peter Juel Henriksen, docent på Handelshøjskolen i København, Ola Knutsson, lektor på Stockholms Universtitet og Robert Östling, doktorand på samme universitet.

kompetence været fremherskende i vores del af verden. Her er undervisningens overordnede mål, at eleverne opnår en alsidig kommunikativ kompetence på målsproget, som forudsætter, at de sættes i stand til at beherske flere kompetencer og færdigheder. Her står lingvistisk, herunder leksikalsk kompetence, sammen med pragmatisk kompetence og sociokulturel indsigt helt centralt, og mht. sprogfærdighed sættes der såvel på receptiv som produktiv beherskelse af målsproget. Mens de receptive færdigheder angår læse- og lyttefærdighed, dvs. forståelsen af skrevne tekster og talt sprog, drejer de produktive færdigheder sig om skrive- og talefærdighed, dvs. at udtrykke sig i tale og skrift. Ifølge det kommunikative paradigme opfattes sprogindlæringen som en individuel, kognitiv, social og kreativ proces, hvor brugen af målsproget anses for at være forudsætning for at sprogindlæring finder sted (Lund 1999, s. 13-39, Richards og Rodgers 2001, s. 151-162 og Auður Hauksdóttir 2001, 64-77). Forskningen i dansk som fremmedsprog i Island viser, at arbejdet med skriftlige tekster vægtes højt i danskundervisningen, og at de receptive færdigheder, især læsefærdigheden generelt opprioriteres, mens de produktive færdigheder, især talefærdigheden negligeres (Auður Hauksdóttir 2001, s. 250-284 og Auður Hauksdóttir 2012a, s. 144-157 og 176-188). Endvidere viser forskningen, at receptionen og især produktionen af talt dansk volder islændingene de største problemer, når de behøver at bruge dansk til kommunikation (Auður Hauksdóttir 2012a, s. 47-76).

Leksikalsk kompetence

I den seneste tid har der været tale om voksende interesse for ordforråd og ordforrådsindlæring, og hvad ordforrådet betyder for beherskelsen af et fremmedsprog (fx Nattinger & DeCarrico 1992, Nation 2001 og Schmitt 2004). Her skal det understreges, at leksikalsk kompetence omfatter både beherskelsen af enkeltord og forskellige typer af faste ordforbindelser eller fraser. Og det er ikke mindst frasernes og enkeltordenes brug i kontekst, som har været i fokus i den seneste tid. Dette kan bl.a. forklares ved de store muligheder som informationsteknologien og datalingvistikken giver for synkrone sproglige studier. Således er det vha. korpuslingvistikken lykkedes at opbygge omfattende autentiske synkrone korpora, som gør det muligt at undersøge ordforrådet og dets brug i kontekst. Korpusundersøgelser viser, at sprogbrugen ikke er så kreativ som Chomsky og visse andre lingvister har forventet, og at det ikke er tilfældigt, hvor og i hvilket omfang fraser og andre mere eller mindre faste leksikalise-

rede helheder forekommer i tekster, og hvordan de bruges. Således kan der konstateres visse restriktioner med hensyn til kombination af ord afhængigt af konteksten og sprogets pragmatiske funktion (Wray 2002, s. 11-13). Når det gælder opbygning af leksikalsk kompetence er beherskelsen af fraser af allerstørste vigtighed. Dels fordi fraserne er en vigtig del af ordforrådet, hvilket betyder at de må kunne afkodes, når de forekommer i tale og skrift, og indkodes, når det drejer sig om produktiv brug af målsproget. Det er også blevet påpeget, at beherskelsen af fraser er vigtigt for fluency på målsproget. Dette forklares ved, at i talesituationer er sprogbrugerens mentale energi yderst begrænset, og derfor fungerer de fraser og helheder, som sprogbruger behersker som en slags sproglige heller eller hvilepunkter, som giver sprogbruger muligheden for at koncentrere sig om andre sider af sprogbrugen. Således bidrager beherskelsen af fraser og andre helheder til fluency på målsproget (Pawley og Syder 1983).

Interessen for ordforrådet må også ses i lyset af en tungere vægtning af kulturaspektet i fremmedsprogsundervisningen og en større interesse for sprogets og sprogbrugens pragmatiske og sociale side. Birgit Henriksen (1999, 71) påpeger, at det netop er ordbetydningen, der udgør bindeledet mellem den kulturelle og sproglige side af sprogfagene. I forbindelse med sprogets kulturelle side er visse frasetyper af større interesse end andre. Det gælder ikke mindst idiommer, ordsprog og pragmatiske formler. Således vidner mange ikonografiske idiommer om forskellige livsmønstre, arbejdsforhold og dagligdagsfænomener, fx *bagerens dårlige øje*, *køre/være på skinner*, *toget er kørt*, *have/få vind i sejlene*. Idiomet opfattes som en selvstændig leksikalsk enhed, dvs. den samlede betydning er en anden end enkeltordenes betydning, og derfor har det karakter af et leksem.

Når en frase forekommer både som et idiom og som en ordinær kollokation, er der tale om dual kodering, dvs. både en overført og konkret betydning. Nogle idiommer afspejler forhold, som ikke mere er en del af moderne menneskers liv. I sådanne tilfælde bruges frasen primært overført i nutidssproget, mens den tilsvarende kollokation sjældent eller aldrig bruges. Eksempelvis kan nævnes idiommerne *have mange jern i ilden*, *spænde buen for højt*, *hverken have noget at bide eller brænde*. Af interesse er også mange metaforiske idiommer, som på forskellig vis relateres til generelle begrebsmetaforer, såsom at diskussion er krig, tid er penge, kærlighed er rejse og at forstå er at se (Lakoff og Johnson 2002, 14-18 og Ruus 2004, 131). Ifølge Brink (2006, 8-10) afviger både de metaforiske

og de non-metaforiske idiomers betydning fra den pålydende betydning og derfor er deres betydning ikke forudsigelig. Af disse grunde kan både idiommer i snæver forstand (ikonografiske/metaforiske) og non-metaforiske idiommer være svære at afkode for ikke modersmålstalende.

Det er dog især den produktive brug af idiomerne, som volder problemer. Selv om der er tale om nært beslægtede sprog såsom dansk og islandsk, hvor der kan konstateres et vist idiomfællesskab, er der ikke desto mindre tale om forskellige grader af forskelle og ligheder og i nogle tilfælde er der ikke tale om parallelle idiompar på de to sprog. På grund af idiomernes uforudsigelighed og da idiompar på de to sprog kan divergere på forskellig vis, er idiomerne problematiske. Det gælder i særdeleshed, når et dansk idiom ikke har en pendant på islandsk. I de tilfælde kan den receptive forståelse af idiomerne være besværlig, da man kan være i tvivl om, hvorvidt idiomet skal forstås overført eller konkret eller hvis ikke man er fortrolig med den implicite sammenligning. Når det gælder produktiv brug af idiomerne, er det ikke til at vide, hvordan man skal stykke ordene sammen. Det samme gælder faktisk også, når udfordringen er at finde frem til en kollokation på dansk, som ikke har en modspiller på islandsk. Noget andet gælder receptionen, hvor kollokationerne pga. deres forholdsvis konkrete betydning som regel er nemmere at afkode. En anden frasetype, som kan være svær at have med at gøre, er de såkaldte sammenligninger, fx *som sild i tønde*, *arbejde som en hest* og *lyve så hurtigt som en hest kan rende*. Når der er tale om nært beslægtede sprog som dansk og islandsk, har sprogene mange fraser til fælles, mens der i andre tilfælde er tale om større eller mindre grader af diversitet. Disse forhold gør, at fraserne kan være svære at mestre. Ved udviklingen af sprogværktøjet tages der højde for de skildrede forhold med den hensigt at bistå brugerne med deres indlæring og brug af fraserne.

I forbindelse med sprogets sociale eller pragmatiske side er frasetypen kommunikative eller pragmatiske formler (også kaldt rutineformler) af største vigtighed. Det er ikke tilfældigt, hvor fraser forekommer og hvordan de bruges i tale og skrift. I det øjeblik man har valgt et register, en social situation eller en sprogbrugssituation er det afgørende for sprogets brug, bl.a. mht. hvilket ordforråd, der trækkes på. De kommunikative formler er i høj grad karakteristiske for talesproget, og de adskiller sig fra andre typer af faste ordforbindelser ved deres pragmatiske og diskursmæssige funktion (Nattinger & DeCarrico 1992, 59-66). De er meget almindelige i hverdagssproget, og derfor findes der en vis form

for indforståethed om dem blandt modersmåltalende. For at kunne begå sig adækvat på et fremmedsprog er det afgørende at kunne trække på gængse kommunikative formler og andre sproglige virkemidler, der anses for at være passende i sammenhængen og som typisk ville bruges af modersmåltalende i de givne kontekster og situationer. Eksempelvis kan nævnes de fraser, som typisk ville blive brugt i forbindelse med lykønskninger, *Held og lykke*, *Til lykke med fødselsdagen*, anmodninger, *Må jeg godt bede om*, taksigelser, invitationer; *Tak for jeg/vi måtte komme*, *Tak for denne gang*, telefonsamtaler, *Træffer jeg ...*, *Hæng lige på*, *Vi høres/tales ved* og når det gælder småsnak eller samtaleregulering, *Hvordan går det*, *Hør en gang*, *Hvad var det jeg ville sige?*, *Det er jeg med på*. Et særligt problem for ikke-modersmåltalende i forbindelse med de gængse fraser er en tendens til at forkorte dem, fx *Værsgo* i stedet for *Vær så god* og *Hvadbehar?* i stedet for *Hvad behager?* Her er der tale om fraser, som sjældent findes i etsprogede ordbøger, da de er aldeles uproblematisk for modersmåltalende. De kommunikative formler adskiller sig syntaktisk fra idiomer og kollokationer ved, at de både kan udgøre en hel sætning eller en sætningsdel; i visse tilfælde kun et enkeltord, fx *Hej* eller *Halløj* (se nærmere i Auður Hauksdóttir 2012). Da fraserne er situationsbundne og mange af dem er højfrekvente i sproget, er de sjældent svære at afkode. Til gengæld kan det være svært at bruge dem produktivt, hvis ikke man kender til deres eksistens, og de ikke har en pendant i udgangssproget. Et andet problem med denne type af fraser er, at de ofte mangler et indholdsord, hvilket betyder, at de er svære at have med at gøre leksikografisk.

Den dansk-islandske frasebase www.frasar.net

I sprogværktøjet www.frasar.net findes der godt 7600 fraser på dansk og deres modspillere på islandsk eller alternativt forklaring på islandsk, hvis ikke der findes en islandsk pendant. Der findes lydeksempler på samtlige fraser, hvilket betyder, at brugerne kan lytte til hvordan fraserne udtales på dansk, når det passer dem. Idiomfrasebasen stammer for det meste fra *Den Danske Ordbog* (DDO) (Hjorth og Kristensen 2003-05), som Det Danske Sprog og Litteraturselskab velvilligt har stillet til rådighed. Desuden har Pia Jarvad bidraget med idiomer fra sin forskning, der angår nye ord i dansk. Betydningsforklaringer på dansk stammer fra samme DDO. For at sikre dækning af idiomerne blev DDO-frasebasen sammenholdt med de idiomer, som forekommer i ordbøgerne *Talemåder i dansk. Ord-*

bog over idiomer (Toftgaard Andersen 1998) og *Danske talemåder* (Røder 1998). I de tilfælde, hvor der fandtes idiomer i disse ordbøger, som ikke var med i DDO-frasebasen, blev de føjet til. Endvidere har de medvirkende i projektet været på jagt efter idiomer og i de tilfælde hvor der er fundet idiomer, som ikke findes i de nævnte kilder, og som ikke desto mindre ser ud til at være ret udbredt i nutidsdansk, er de blevet føjet til www.frasar.net. På nuværende tidspunkt indeholder www.frasar.net 3332 billedlige idiomer, 3281 ikke billedlige idiomer, 332 sammenligninger, 186 kollokationer og desuden 46 præpositionsforbindelser. Som før nævnt spiller de kommunikative formler en stor rolle i hverdagssproget, og beherskelsen af disse er afgørende for advækat og mundret sprogbrug. I arbejdet med denne frasetype viste det sig, at formlerne kun i begrænset omfang fandtes i danske ordbøger, og derfor var vi nødt til selv at samle denne frasetype. For at finde frem til de kommunikative formler gennemgik vi systematisk lærebøger i dansk som andetsprog og de parlører som vi kunne få fat i. Vi indsamlede desuden systematisk fraser, der angår højtider og mærkedage, fx jule- og nytårshilsner, samt fødselsdags-hilsner. Desuden udarbejdede vi et arbejdsark, hvor vi spurgte en mindre gruppe af danskere om, hvad man typisk ville sige i de forskellige situationer. Sidst men ikke mindst var vi selv på vagt som observatører, når det gjaldt danske spillefilm, debat i fjernsyn eller sprogets brug i autentiske situationer i dansk dagligdag, fx i banken, hos bageren eller i toget. Gennem dette arbejde er det lykkedes os at finde frem til 439 kommunikative formler i nutidssproget fra såvel tale som skrift.

Søgning og andre funktioner

Fordelen ved elektroniske fraseordbøger er bl.a., at man kan søge på enkeltord eller på frasen i sin helhed, og at det kan lade sig gøre at søge såvel på den danske som den islandske frase, dvs. såfremt der findes et mere eller mindre ækvivalent frasepar på de to sprog. Ved at klikke på det islandske eller det danske flag, skifter man fra islandsk til dansk metasprog og vice versa. Når det gælder afkodning af idiomer og andre fraser, hvis betydning ikke er transparent, er det en fordel at kunne finde frem til frasens betydning på dansk og få oplysninger om, hvorvidt der findes en islandsk modspiller til den danske frase, se figur 1:

The screenshot shows the Frasar.net website, which is a Danish-Icelandic language help tool. The main content area displays a search result for the Danish phrase "Købe katten i sækken". The Danish meaning is explained as "Blive snydt el. skuffet i en handel ved at købe noget ubeset og derefter opdage at det er ringere end forventet." The Icelandic equivalent is "Kaupa köttinn í sekki", with the meaning "Kaupa e-ð óséð; vera svikinn um það sem maður væntir (með ráðstöfun sinni) (MM)." Examples of usage are provided for both languages, illustrating the concept of buying something without seeing it. The page also includes a search bar, navigation links, and contact information.

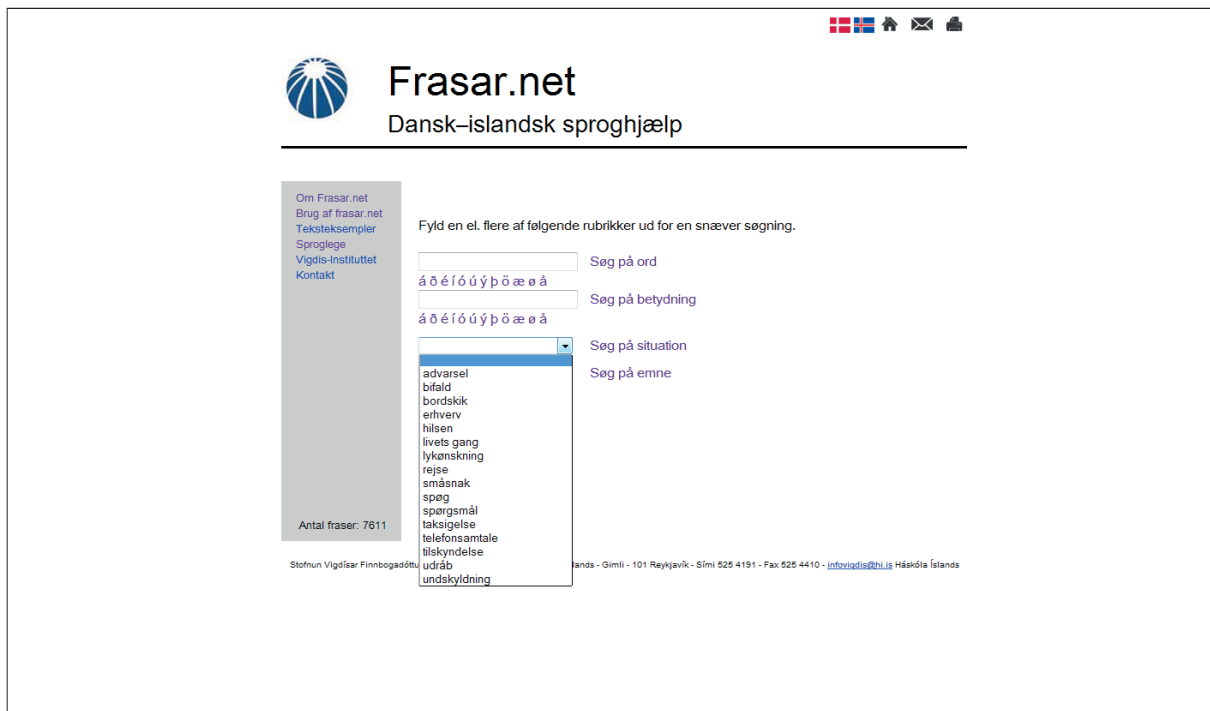
Figur 1. Et ækvivalent frasepar på dansk og islandsk.

Når det gælder produktiv brug af fraser på et fremmedsprog, som sprogbrugeren ikke helt er fortrolig med, er det vigtigt at kunne få oplysninger om frasens brug i kontekst for på den måde at kunne trække på det ordforråd, som frasen typisk forbindes med. Under søgefunktionen *Eksempler på brug*, se figur 1, kan man se forekomsten af fraserne i konstruerede eksempler, og ved at klikke på korpusDK eller Google kan man se dem i autentiske situationer. I de tilfælde, hvor der ikke findes en islandsk pendant til den danske frase, får man en betydningsforklaring på islandsk. Når der er tale om parallelle fraser på de to sprog og hvor der er anledning til at være på vagt over for lumske forskelle eller ligheder, dukker funktionen *Obs!* op, hvor der eksplicit gives komparative iagttagelser eller forklaringer, der angår fraserne betydning, bøjning eller brug. Hvor flere end én islandsk frase kan komme i betragtning som modspiller til den danske frase, dukker valgmuligheden *Andre varianter op*.

The screenshot shows the Frasar.net website interface. At the top, there is a logo and the text 'Frasar.net Dansk-islandsk sproghjælp'. Below this is a search bar with the text 'á ð é í ó ú ý þ ð æ ø á' and a 'Søg' button. To the right of the search bar is a 'Udvidet søgning' button. Below the search bar, there are several sections: 'Om Frasar.net', 'Brug af frasar.net', 'Tekstseksempler', 'Sproglege', 'Vigdís-instituttet', and 'Kontakt'. The main content area displays the search results for '(Nogen) glimrer ved sit fravær'. It includes the Danish phrase, its Icelandic equivalent '-oggg', and a note 'ironsk'. Below this, there is a section titled 'Láta ekki sjá sig' with an 'Obs!' icon, explaining that this is not a standard phrase and providing context. To the right, there is a box titled 'Eksempler på brug' with three examples of the phrase in use. At the bottom, there is a section titled 'Andre varianter' with an 'Obs!' icon, listing several variations of the phrase and their Icelandic equivalents. The footer contains contact information for Stofnun Vigdísar Finnboðadóttur.

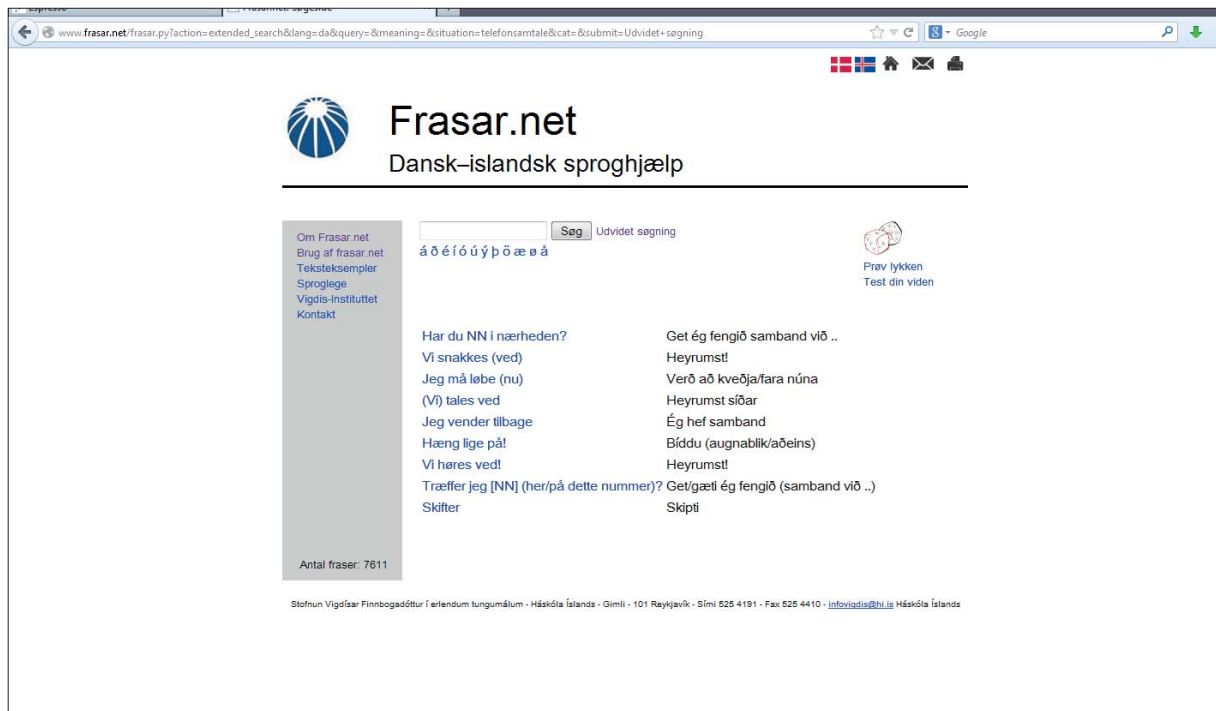
Figur 2. Ikke ækvivalente frasepar.

Mange af de kommunikative formler binder sig i høj grad til bestemte situationer eller til visse sproglige funktioner. Beherskelsen af formlerne spiller derfor en stor rolle for sprogbrugerens mulighed for at kunne begå sig adækvat, mundret og hensigtsmæssigt på fremmedsproget. Når det gælder beslægtede sprog såsom dansk og islandsk, finder man hurtigt ud af, at mange af de danske formler har deres modspiller på islandsk. Det er dog ikke altid tilfældet, og derfor kan det være svært at finde ud af, hvordan formlen ser ud og lyder på dansk. Det er ikke uproblematisk. For det første glimrer denne type af fraser ofte ved deres fravær i etsprogede ordbøger, og hvis de findes, er det ikke til at vide, hvad man skal lede efter. Et andet problem kan være fraværet af egentlige indholdsord og derfor er der sådan set ikke noget oplagt opslagsord. For at imødegå dette problem, er der i www.frasar.net udviklet særlige søgemuligheder for kommunikative formler. Som det fremgår af figur 3, kan man under søgemuligheden *Udvidet søgning*, søge på forskellige situationer:



Figur 3. Søgning på situationer.

Ved at klikke på ét af de opslagsord, der repræsenterer situationerne, dukker der en liste af formler op. Eksempelvis kan nævnes, at ved at klikke på *telefonsamtale*, dukker følgende situationsbundne fraser op:



Figur 4. Søgning på situationen telefonsamtale.

Alternativt kan man under *Udvidet søgning* lede efter fraser under søgemulighederne *Betydning* eller *Emne*.

Tekstseksempler

Ligesom mange hverdagssituationer kan forekommer ritualiserede, er visse teksttyper mere konventionelle end andre. Det gælder ikke mindst tekster, der bruges i forbindelse med sociale vaner i samfundet, samt sæd og skik, såsom at fejre jul, fødselsdage og barnedåb. Til trods for visse variationer er genrer af denne art forholdsvis fastbundne mht. form og ordvalg. Igennem årene har mange henvendt sig til mig, som skriver denne tekst, for at bede om hjælp med at formulere lejlighedstekster af denne art på dansk. Med det formål at lette og fremme produktiv brug af dansk i sådanne situationer, er der i sprogværktøjet gjort et forsøg med en særlig hjælpefunktion, som strækker sig ud over sætningsniveauet. Funktionen går ud på at hjælpe brugerne med at formulere forholdsvis fastkonstruerede almindelige nyttetekster, fx julekort, cv, anbefalinger og kondolence. Således indholder www.frasar.net en række nyttetekster. Som det fremgår af figur 5 er der tale om adskillige typer af tekster inden for to hovedkategorier, nemlig *Livets gang* og *Årets gang*.

The screenshot shows the Frasar.net website interface. At the top right, there are icons for the Danish flag, a home icon, an envelope icon, and a printer icon. The main header features the Frasar.net logo (a blue sunburst) and the text 'Frasar.net Dansk-islandsk sproghjælp'. Below the header is a search bar with a 'Søg' button and the text 'Udvidet søgning'. To the right of the search bar is a 'Prøv lykken Test din viden' button with a circular icon. On the left side, there is a vertical navigation menu with links: 'Om Frasar.net', 'Brug af frasar.net', 'Teksteksampler', 'Sproglege', 'Vigdis-Instituttet', and 'Kontakt'. Below the menu, it says 'Antal fraser: 7611'. The main content area is titled 'Livets gang' and contains a list of categories with sub-items:

- Barnedåb (Demo)
 - Indbydelse
 - Lykønskninger
- Bryllup (Demo)
 - Indbydelse
 - Lykønskninger
 - Takkekort
- Dødsfald (Demo)
 - Kondolence
 - Takkekort
- Eksamen (Demo)
 - Indbydelse
 - Lykønskninger
- Fødselsdag (Demo)
 - Indbydelse
 - Lykønskninger
 - Takkekort
- Konfirmation (Demo)
 - Indbydelse
 - Lykønskninger
 - Takkekort

Below the list is the heading 'Årets gang'.

Figur 5. *Teksteksampler og produktiv brug af dansk.*

Teksteksamplerne fremtræder som en slags ramme eller template filer for de forskellige teksttyper. For hver genre er der tale om visse variationsmuligheder, som brugerne kan manipulere med afhængigt af deres kommunikative behov. Således kan de vælge forskellige opstillede formuleringer og dermed afgøre tekstens formalitets- og intimitetsgrad og stil. Til sidst kan man ved at klikke på *Vis*, se ens tekst i sin helhed, se figur 6.

The screenshot shows the Frasar.net website, which is a Danish-Icelandic phrase generator. The page layout includes a header with the Frasar.net logo and the text 'Dansk-Islandsk sproghjælp'. Below the header, there is a search bar with a 'Søg' button and a 'Udvidet søgning' link. A sidebar on the left contains navigation links: 'Om Frasar.net', 'Brug af frasar.net', 'Teksteksempler', 'Sproglege', 'Vigdís-instituttet', and 'Kontakt'. The main content area displays a search result for the Icelandic phrase 'á þé í ó ú ý þ ö æ ø á'. Below this, there is a section titled 'Nedenstående teksteksempler kan bruges til at skrive et kort eller brev.' followed by instructions on how to use the examples. A 'Prøv lykken Test din viden' button is also present. The main content area shows a list of example phrases with dropdown menus for selection. The selected phrase is 'Jeg ønsker dig en god jul og et godt nytår.' The dropdown menu for 'en god' is open, showing options: 'en god', 'en rigtig god', 'glædelig', and 'rigtig glædelig'. The dropdown menu for 'snart er gået' is also open, showing options: 'snart er gået' and 'i det nye år.'. The page footer contains contact information for Stofnun Vigdísar Finnboğadóttur.

Figur 6. Eksempel på julekort med forskellige variationsmuligheder.

Sproglege

Udover de nævnte funktioner, indeholder www.frasar.net sproglege af forskellig karakter, som både har til formål at være til gavn og lyst. Sproglegene er taskbaserede og ved løsning af tasken sættes der fokus på forskellige sider af frasernes karakter, såsom deres ordmateriale, udformning, betydning og brug. Set i lyset af det kommunikative paradigme, hvor indlæringen betragtes om en individuel, kognitiv process, blev der lagt vægt på at udarbejde sproglege af forskellige sværhedsgrader og typer, se figur 7.



Figur 7. Sproglege af forskellig karakter og sværhedsgrad.

Mens nogle af legene går ud på at gætte sig frem til den rigtige løsning, forudsætter andre, at fraseværktøjet tages i brug for at fuldføre legen/ta-sken. Nogle af sproglegene er interaktive, således at brugerne straks får en reaktion på, om deres løsningsforsøg rammer plet eller ej.

Konklusion

Ved udviklingen af www.frasar.net er ny viden inden for sprogteknologi blevet anvendt for at fremme islændingenes indlæring og brug af dansk. Endvidere er der taget højde for aktuelle teorier om ordforråd og sprog-indlæring, samt den viden som forskningen i dansk som fremmedsprog har indhøstet. Erfaringen viser, at sprogværktøjet er et kærkomment sup-plement til eksisterende leksikografiske værktøjer. Tilbage står den store udfordring at finde frem til sprogteknologiske løsninger, som kan bi-stå med at fremme beherskelsen af udtale og spontan talefærdighed på dansk.

Auður Hauksdóttir er dosent i dansk ved Islands universitet og direk-tør for Vigdís Finnsbogaðóttir-instituttet på Island.

Litteratur

- Auður Hauksdóttir, 2001: Lærerens strategier – elevernes dansk. Dansk som fremmedsprog i Island. København: Nordisk Ministerråd (Tema Nord).
- Auður Hauksdóttir, 2007: Idiomerne glimrer ved deres fravær: Om dansk-islandsk idiomatik. I: Jørgensen, Henrik og Peter Widell (red.): Det bedre argument: Festskrift til Ole Togeby 7. marts 2007. Århus: Wessel og Huitfeldt, s. 197-215.
- Auður Hauksdóttir, 2012: At komme til orde på et mundret dansk: Om fraser, fraseindlæring og fraseværktøj anskuet kontrastivt. I: Henrichsen, Peter Juel et al (red.) Copenhagen Studies in Language, nr. 42. Speech in Action. Proceedings of the 1st SJUSK Conference on Contemporary Speech Habits, s. 123-142.
- Auður Hauksdóttir, 2012a: Dansk som fremmedsprog i en akademisk kontekst. Om islændingenenes behov for danskkundskaber under videreuddannelse i Danmark (Københavnstudier i tosprogethed nr. 68). København: Københavns Universitet, Humanistisk Fakultet.
- Brink, Lars, 2006: Den fraseologiske terminologi. I: Nordiske Studier i Leksikografi 8. København, s. 39-51.
- Henriksen, Birgit, 1999: Ordforråd og ordforrådsindlæring. I: Holmen, Anne og Karen Lund (red.): Studier i dansk som andetsprog. København: Akademisk Forlag, s. 71-106.
- Henriksen, Birgit, 2009: Det oversete ordforråd. I: Byram, M. S. m.fl. (red.): Sprogfag i forandring – pædagogik og praksis. København: Samfundslitteratur, s. 204-228.
- Lakoff, George og Mark Jonson, 2002. Hverdagens metaforer. København: Hans Reitzels Forlag.
- Lund, Karen, 1999: Sprog, tilegnelse og kommunikativ undervisning. I: Holmen, Anne og Karen Lund (red.): Studier i dansk som andetsprog. København: Akademisk Forlag, s. 11-69.
- Lund, Karen, 2009: Fokus på sproglæring og pædagogiske implikationer for undervisningen. I: Byram, M. S. m.fl. (red.): Sprogfag i forandring – pædagogik og praksis. København: Samfundslitteratur, s. 127-166.
- Nation, Paul I. S., 2001: Learning Vocabulary in Another Language. Cambridge: Cambridge University Press.
- Nattinger, James & Jeanette S. DeCarrico, 1992: Lexical Phrases and Language Teaching. Oxford: Oxford University Press.

- Pawley, Andrew og Frances Hodgetts Syder, 1983: Two puzzles for linguistic theory: nativelike selection and nativelike fluency. I: Richards, Jack C. og Richard W.
- Schmidt (red.). New York: Language and Communication: Longman, s. 191-226.
- Richard, Jack C. & Theodore S. Rodgers, 2001: Approaches and Methods in Language Teaching. Cambridge: Cambridge University Press.
- Ruus, Hanne, 2004: Det er da indlysende! Til det empiriske grundlag for den kognitive semantik. I: Rathje, Marianne og Linda Svenstrup (red.). Sprogpsykologi. Udvalgte kerneemner. København: Museum Tusulanums Forlag, s. 127-151.
- Wray, Alison, 2002: Formulaic Language and the Lexicon. Cambridge: Cambridge University Press.

Summary

The language tool [frasar.net](http://www.frasar.net) (<http://www.frasar.net>) was created as an aid for Icelanders learning and communicating in Danish. Frasar.net now contains more than 7600 phrases in Danish, organized in subcategories such as idioms, collocations, comparisons, and communicative formulas. Each Danish phrase thus has an Icelandic equivalent phrase or, alternatively, if no equivalent could be found, a paraphrase or explanation. Moreover, [frasar.net](http://www.frasar.net) provides contrastive information on inflectional patterns, false friends, practical usage etc. Frasar.net was developed as a computer assisted language learning (CALL) tool based on current theories of foreign language teaching and function of vocabulary and vocabulary learning, taking advantage of contemporary developments in computational linguistics and language technology. Current research on Danish as a foreign language in Iceland has also played an important role. Frasar.net is aimed at developing receptive as well as productive language skills and therefore includes a broad variety of options, including search options (e.g. for individual words, phrases, or situations), occasion-related text frames (e.g. for congratulations and greetings), and typical dialogue patterns. Frasar.net also provides audio files with native Danish pronunciations, linked with real-world examples.

Det islandske ordklasseopmærkede korpus (MÍM)

Sigrún Helgadóttir

I artiklen gives der en kort redegørelse for projektet "Det islandske ordklasseopmærkede korpus" som kaldes "MÍM" på islandsk. Der beskrives kort hvorfor projektet blev påbegyndt. Endvidere beskrives hvad et ordklasseopmærket korpus er for noget og hvordan det kan bruges. Desuden følger der en redegørelse for hvordan korpuset blev oprettet og dets tilgængelighed. Til sidst beskrives to andre korpuser som er tilgængelige på webpladsen <http://mim.arnastofnun.is/>.

Indledning

I 1998 nedsatte undervisnings-, forsknings-, og kulturminister Björn Bjarnason et udvalg for at undersøge statussen af sprogteknologi i Island og fremsætte forslag om hvordan sprogteknologien kunne styrkes (Rögnvaldur Ólafsson m.fl. 1999). Udvalget afleverede sin rapport i april 1999, og i 2000 blev undervisningsministeriets sprogteknologiske projekt lanceret. En af de sprogresurser som udvalget mente manglede var et ordklasseopmærket tekstkorpus. MÍM-projektet var et af de seneste projekter, som blev finansieret af det sprogteknologiske projekt, og det skulle aflevere et sådant korpus. Projektet blev startet i 2004 på *Orðabók Háskólans* (Leksikografisk institut ved Islands universitet) og færdiggjort på *Stofnun Árna Magnússonar í íslenskum fræðum* (Árni Magnússon-instituttet for islandske studier) i 2013. Artiklen giver en kort redegørelse for projektet. Ordklasseopmærket korpus beskrives, og derefter følger en redegørelse for MÍM-korpuset, hvordan tekster blev samlet, og hvordan man behandlede ophavsret. Endvidere beskrives hvordan teksten blev rensat og opmærket med oplysninger om ordklasse, bøjning og lemmaer. Tilgængelighed og brug beskrives og to andre korpuser som også er tilgængelige til søgning på samme måde som MÍM-korpuset. Til sidst gives et eksempel på søgning i MÍM.

Ordklasseopmærket korpus

Et ordklasseopmærket korpus (e. *tagged corpus*) er en samling af elektroniske eller digitale tekster. Hver enkelt tekst er forsynet med oplysninger om teksten som kaldes *metadata*, for eksempel titel, udgivelsesår, genre og forfatterens navn (eventuelt køn og fødselsår) for udgivne tekster. Hvert ord er opmærket med oplysninger om ordklasse, bøjning og lemma, og korpuset er gemt i standardiseret format, sædvanligvis xml-format. For at få et balanceret korpus er det en fordel, at teksterne kommer fra forskellige kilder, som skal give indtryk af, hvordan et sprog bliver brugt i en bestemt periode.

Fra korpuset kan man finde information om frekvens af ordklasser, ord og bøjningsformer, om ordforbindelser, syntaks og semantik osv. Korpusser er også nyttige, når man laver for eksempel ordbøger, programmer til stave- og grammatikkontrol, maskinoversættelse, talegenkendelse og talesyntese, til støtte for handicappede (blinde, døve og hørehæmmede, bevægelseshæmmede og ordblinde) og i sprogundervisning.

MÍM korpusprojektet

Formålet med korpusprojektet (Sigrún Helgadóttir m.fl. 2012) var at samle tekster, skrevet af mennesker som har islandsk som modersmål, fra en række forskellige kilder med i alt 25 millioner ord fra perioden 2000–2010. Kun tekster som var elektronisk tilgængelige, blev samlet, og man skulle sikre licens til brug af teksterne i korpuset fra indehavere af ophavsret. Det var tænkt som yderst vigtigt at sikre licens for ophavsretsbeskyttet materiale. Alle tekster skulle opmærkes automatisk med oplysninger om ordklasse og bøjning og med lemma.

I tabel 1 vises hvordan teksterne i korpuset fordeles mellem de forskellige kilder af tekster. Størstedelen af teksterne kommer fra trykte bøger og fra aviser, både trykte og webaviser. Offentlige tekster udgør en stor del af korpuset og dernæst kommer tekster fra tidsskrifter, trykte og elektroniske. Man fandt det også vigtigt at få tekster fra blogs. Vi fik tekster fra almene bloggere, teologer og politikere. Bloggene er anonyme. I gruppen ”Usorteret” findes der nogle tekster fra e-post lister, som også ser ud til at være uformel tekst ligesom bloggen. Det var meget vigtigt at få hjælp fra redaktøren til universitetets videnskabsweb¹. Han sikrede licens fra alle som skriver tekster for webben og afleverede teksten til kor-

1 <http://visindavefur.is/>

pusset. Videnskabswebben indeholder tekster om alt muligt fra socialvidenskab til astronomi.

Genre i MÍM	Antal tekster		
	(filer)	Antal ord	%
Trykte bøger	168	5.972.893	23,89
Aviser (trykte og elektroniske)	12.725	5.779.509	23,12
Offentlige tekster (rapporter, domme, forslag, love, drøftelser fra Alþingi)	1.246	3.513.990	14,06
Tidsskrifter (trykte og elektroniske)	311	2.501.222	10,00
Blogs	8.998	1.976.706	7,91
Artikler fra Universitetets videnskabsweb (Vísindavefurinn)	4.949	1.838.909	7,36
Tekster fra websites for virksomheder, organisationer og institutter	106	1.337.764	5,35
Tekster til oplæsning (blandt andet fra radio og tv)	1.196	694.506	2,78
Stile og skriftlige opgaver fra gymnasieelever og studenter	51	666.042	2,66
Talesprog	4	504.318	2,02
Usorteret	46	214.663	0,86
Totalt	29.800	25.000.522	100,00

Tabel 1. Tekster i MÍM efter genre.

I alt har man 58 % af tekstmaterialet i korpuset fra trykte kilder, 40 % fra webben og talesprog udgør 2 %. Ophavsretsbeskyttet tekst udgør omtrent 85,9 % af korpusteksterne.

Projektet havde ikke resurser til at samle talesprog. Talesprog blev skaffet fra fire forskellige talesprogsprojekter (Höskuldur Þráinsson m.fl. 2007), i alt 54 timer af transskriberet tale. Disse projekter omfatter monologer (taler fra Alþingi, delprojekt af ScanDiaSyn), interviews (MIN - Moderne importord i sprogene i Norden) og spontane samtaler mellem voksne mennesker, mænd og kvinder (ÍSTAL og Hvordan taler unge islændinge i begyndelsen af det 21. århundrede?). Alle transskriberede tekster er blevet opmærkede og inkluderet i korpuset, men er anonyme. Man planlægger at gøre talesproget tilgængeligt til søgning med lydfiles snart, men kun talerne fra Alþingi bliver tilgængelige for almenheden.

Ophavsret

I korpusprojektet arbejdede man med to juridiske dokumenter. Alle indehavere af ophavsretsbeskyttet udgiven tekst underskrev en deklARATION. Indehaverne fik information om projektet og en kopi af brugerlicensen. De fleste indehavere har givet licens til brug af alle deres tekster, men enkelte kun de tekster som blev inkluderet i korpusset. Forfattere af tekst som ikke var udgivet, fik en meget enklere deklARATION til underskrift. DeklARATIONen foreskriver blandt andet at alle tekster skulle være tilgængelige uden betaling, kun 80 % af udgivne tekster inkluderes i korpusset, og brugere må acceptere brugerlicensen.

Brugerlicensen foreskriver at brugeren kan bruge sine resultater (dvs. det som han lærer fra korpusset) frit. Man kan for eksempel forestille sig at man laver n-grams fra korpusset og bruger disse til at lave skrivnehjælp til ordblinde. Men teksterne må ikke kopieres eller videregives til andre, undtagen det, som er omfattet af citatretten.

I begyndelsen af projektet sikrede man samarbejde fra Den islandske forfatterforening² (Rithöfundasamband Íslands), Islandsk forfatterforening for faglitteratur og undervisning³ (Hagþenkir) og Den islandske forlæggerforening⁴ (Félag íslenskra bókaútgefenda). Alle tre foreninger anbefalede deres medlemmer at de skulle samarbejde med Árni Magnússon-instituttet om korpusprojektet. Det var yderst vigtigt at sikre at udgiverne var positive overfor projektet siden de bevarer de elektroniske kopier af udgivne bøger. Projektet fik et meget fint samarbejde med nogle af de største udgivere i Island.

Forberedelse af tekster for korpusset

Tekst blev skaffet i forskellige formater, som pdf, xml eller Word filer, som tekst fra databaser, webtekst og så videre. Teksten måtte uddrages af formatet. Man fjernede også fremmedsprogede og oldnordiske citater, fodnoter, indholdsfortegnelser, indekser, digte, tabeller, billeder og så videre samt bindestreg. Alle tekster er tilgængelige i UTF-8 tegnkodningstabel.

Den sædvanlige arbejdsgang ved korpustekst, når teksten er blevet rensat, er at segmentere i sætninger og tokenisere, dvs. fordele teksten i tokens eller ord. Det blev gjort med *IceNLP* softwaresystemet (Hrafn

2 <http://rsi.is/>

3 <http://hagthenkir.is/>

4 <http://www.bokatidindi.is/index.php/forsiea>

Loftsson og Eiríkur Rögnvaldsson 2007). Derefter blev teksten ordklasseopmærket og lemmatiseret.

Opmærkningen giver hvert ord en bogstavsstreng, et tag, som viser ordklasse og bøjning og lemma. Det første bogstav i taggen angiver ordklassen, og de andre (op til 5) angiver for eksempel køn, tal og kasus for substantiver og person, tal og tempus for verber. Lemma er ordets grundform eller opslagsform og er for eksempel nominativ, ental for substantiver og infinitiv for verber. Som eksempel kan man analysere sætningen *ég sagði* (jeg sagde). Lemma til ordet *ég* er *ég* og taggen bliver **fp1en**, hvor **f** står for pronomener, **p** står for personligt pronomener, **1** står for første person, **e** står for ental og **n** står for nominativ. Lemma til ordet *sagði* er *segja* og taggen bliver **sfg1eþ** hvor **s** står for verb, **f** står for indikativ, **g** står for aktiv, **1** står for første person, **e** står for ental og **þ** står for datid.

Analysen bygger på det tidligere arbejde med *Íslensk orðtíðnibók* [Den islandske frekvensordbog] (Jörgen Pind m.fl. 1991) og tagsættet indeholder omtrent 700 tags.

Ordklasseopmærkning sker med fire ordklassetagere og et system til at vælge imellem de fire tags. Til sidst bruges en lemmatiserer for at finde lemmaet for et ord. Hele processen er pakket ind i et system som vi kalder *CorpusTagger* (Hrafn Loftsson m.fl. 2010).

Alle tekster i korpusset er forsynet med metadata. For udgivne tekster består metadata af bibliografiske oplysninger som titel, udgivelsesår, genre og forfatterens navn (eventuelt køn og fødselsår). For andre tekster registreres metadata, som identificerer teksten. For talesprog er oplysninger om sessionen og talerne registreret. De fleste metadata er blevet registreret manuelt, for enkelte filer fra avisen *Morgunblaðið*, for eksempel, og taler fra Alþingi er metadata oprettet automatisk. Når man søger i teksterne via søgegrænsefladen, bliver metadata vist og alle xml-filer, som kan downloades, indeholder metadata.

Tilgængelighed og brug

Korpusset er tilgængeligt på to forskellige måder. Man kan søge i teksterne og bruge taggene for at lave en mere nøjagtig søgning. Man kan finde eksempler på hvordan et ord eller en vending bliver brugt i islandsk. Resultatet vises i en konkordans, som er en række linjer, hvor søgeudtrykket optræder i den umiddelbare kontekst. For at få et overblik over konkordansen kan man evt. sortere den. Kilden, hvor teksten i hver enkelt linje i

konkordansen stammer fra, bliver også vist. Søgesiden for *MÍM*-korpuset er <http://mim.arnastofnun.is/>.

Søgegrænsefladen bygger på Glossa⁵ fra Universitetet i Oslo og Glossa bruger “corpus search engine”, IMS Corpus Workbench (CWB)⁶ fra Universitetet i Stuttgart.

Brugere som vil undersøge og se eksempler på sproglige fænomener, sådan som de optræder i naturligt forekommende islandske tekster, bruger søgesiden. Det kan være nyttigt både for dem, der beskæftiger sig professionelt med sprog (fx journalister, lærere og sprogforskere), og for dem, der bare synes sprog er interessant og sjovt. *MÍM* er allerede blevet brugt i undervisningen i Háskóli Íslands. Vi har planer om at præsentere, hvordan korpuset kan bruges i undervisning. Der er skrevet en artikel om *MÍM* og dets brug i undervisning i et tidsskrift udgivet af sammenslutningen af islandske modersmålslærere (Sigrún Helgadóttir 2013).

I forbindelse med META-NORD projektet har den islandske META-NORD gruppe etableret webpladsen www.málföng.is, hvor forskellige sprogresurser er tilgængelige. *MÍM*-korpuset er tilgængeligt for download fra denne webplads i en pakke som består af 29.800 filer i TEI-konform⁷ xml-format. Brugere, som vil hente korpuset, må acceptere en særlig brugerlicens. Denne pakke er hovedsagelig tænkt for dem, som laver sprogteknologiske værktøjer. Korpuset er allerede blevet brugt til at komplettere BÍN (database for moderne islandske fleksioner) (Bjarnadóttir 2012), for at lave en database med semantiske relationer for at udarbejde et program til stavekontrol og for at udforme ”word prediction” til ordblinde.

Andre korpuser

På *MÍM* søgesiden, <http://mim.arnastofnun.is/>, findes der to andre korpuser. Det første er Korpus for den islandske frekvensordbog (IFD Korpusset). Frekvensordbogen blev udgivet 1991 (Jörgen Pind m.fl. 1991) og er lavet af en samling af tekster udgivet 1980–1989, hovedsagelig litterære tekster. Teksterne giver et billede af sproget, som det blev skrevet i udgivne tekster i perioden 1980–1989. Tekstsamlingen indeholder omtrent 500.000 ord fra 100 tekster, opmærkede med oplysninger om ordklasse,

5 <http://www.hf.uio.no/tekstlab/glossa.html>

6 <http://cwb.sourceforge.net/>

7 <http://www.tei-c.org/index.xml>

bøjning og lemmaer plus metadata om hver tekst. Korpusset er velegnet til undervisning i morfologi, siden taggene blev korrigeret manuelt. Korpusset er tilgængeligt på webpladsen <http://www.málföng.is/> til søgning og download med særlig brugerlicens (omtrent identisk brugerlicensen for *MÍM*). IFD korpusset er også blevet brugt til træning af "data-driven" taggere og til udvikling af regelbaserede taggere. Korpusset er tilgængeligt til dette formål på webpladsen <http://málföng.is/>.

Det andet korpus indeholder tekster fra de islandske sagaer og kaldes *Sagakorpusset* (Eiríkur Rögnvaldsson og Sigrún Helgadóttir 2011). Korpusset indeholder 44 digitale tekster fra sagaer, dvs. fra 41 islandske sagaer samt værkerne *Sturlunga*, *Heimskringla* og *Landnámabók*. Korpusset indeholder 1.659.385 ord, opmærkede med oplysninger om ordklasse, bøjning og lemma plus metadata om hver tekst. Teksterne er blevet normaliseret til moderne retskrivning og nogle bøjningsendelser blev ændret. Korpusset er tilgængeligt på webpladsen <http://málföng.is/> til søgning og download med brugerlicens CC BY 3.0⁸ (Creative Commons Attribution 3.0 Unported). *Sagakorpusset* kan bruges blandt andet til at undersøge brug af ord og konstruktioner i sagaerne.

Eksempel på søgning

Man kan søge efter enkelte ord eller ordforbindelser i *MÍM* og finde eksempler på sprogbrug i formelle eller uformelle tekster, som blev skrevet i perioden 2000–2010. Man kan bruge taggene for en mere præcis søgning. Det kan for eksempel være nyttigt at finde, hvilket adjektiv bruges med forskellige substantiver.

Vi kan for eksempel søge efter hvilke adjektiver bruges med ordet "kjóll" (kjole). I så fald kan man definere søgning efter alle bøjningsformer af ordet "kjóll" med et adjektiv, som står foran substantivet.

Figur 1 viser hvordan søgningen er defineret. Det første søgeord bør være et adjektiv (lýsingarorð) og det andet søgeord en bøjningsform af ordet "kjóll". Det realiseres ved at specificere "nefnimynd" (lemma)⁹.

Søgningen giver 198 konkordanslinjer, som alle stammer fra udgivne bøger. Figurer 2 og 3 viser nogle konkordanslinjer, som er resultat af søgningen specificeret i Figur 1. Med en mouse-overfunktion kan man se or-

8 <http://creativecommons.org/licenses/by/3.0/>

9 Søgegrænsefladen er på islandsk men man har planer om en engelsk grænseflade i den nærmeste fremtid.

denes analyse. På Figur 2 kan man se en analyse for ordet "hvítum" (maskulin, dativ af ordet "hvítur", dansk hvid).

På venstre side af hver konkordanslinje er en tegnstreng for kilden af tekstudsnittet. Med en mouse-overfunktion på tegnstrengen kan man se oplysninger om kilden. På Figur 3 kan man se at teksten „...klæða sig . Fyrst fóru þær í **ljósbláan kjól** eða kyrtil sem náði þeim niður á...” stammer fra bogen *Valkyrjan*, som er skrevet af forfatteren Elías Snæland Jónsson, udgivet af Vaka-Helgafell i 2003.

Konklusion

Det er allerede klart at *MÍM*-korpuset er meget vigtigt for dem, som beskæftiger sig med lingvistiske undersøgelser. Det er også vigtigt for dem, som laver sprogteknologiske værktøjer, da det er det eneste store islandske ordklasseopmærkede korpus. Det var uhyre vigtigt at sikre licens til brug af ophavsretsbeskyttet tekst således at forskere, eventuelt med nogle restriktioner, kan bruge korpuset til forskellige formål.

Mörkuð íslensk málheild
 [Mörkuð íslensk málheild | Orðtíðnibók | Fornrit]

Leitarlína 1

orð á milli
 frá
 til

Leita

Hreinsa
ALLT

Fjöldi niðurstaðna á síðu:
 Fjöldi orða umhverfis leitarorð:
 vinstri hægr
 Hámarksfjöldi niðurstaðna :

Mörkuð íslensk málheild = 25.000.522 lesmálsorð.
Fornrit = texti úr 44 sögum úr útgáfu Svarts á hvítu, samtals 1.659.385 orð
Orðtíðnibók = textar Orðtíðnibókar

Allar athugasemdir, ábendingar og tillögur eru vel þegnar. Þær má senda á malfong[hja]malfong.is

© Stofnun Árna Magnússonar í íslenskum fræðum - Orðfræðisvið
 Neshaga 16 - 107 Reykjavík - Sími 525 4430, malfong[hja]malfong.is

Figur 1. Søgning for adjectiver med ordet kjóll.

BAEKUR-B0M BAEKUR-B0O BAEKUR-B0O BAEKUR-B0O BAEKUR-B0Q BAEKUR-B0R BAEKUR-B0R BAEKUR-B0R BAEKUR-B0S	hvíta perlufesti um hálsinn , í flegnum liggur hún á gólfinu heima , í einu sinni að hafa séð hana í í hús . Þar lá mamma í Ókei ... Mamma þín er sorgmædd (klæða sig . Fyrst fóru þær í Ruth og Leslie . Ég hentist í og hjóðfærin . Ég kom tiplandi í Hattamaðurinn í Íðnó . Ég fór í sinnar . Þær voru allar mættar í	hvítum kjól hvítur lýsingarorð karlkyn eintala þágufall sterk-beyging frumstig bláa kjólinn fínasta kjólnum	, og Jack í hvítri skyrtu með og allir eru að stumra yfir og það var þegar hún var fjallkonan og fullt af fólki að stumra yfir en hún er í hvíld núna þá kyrtil sem náði þeim niður á og fann skip við höfnina með matsölustað eftir strætum New York borgar . Af með rauðu rósunum frá Sýrlandi og hafði sinum úr gullfalegum Afrikuefnum með borða yfir
--	--	--	--

Figur 2. Nogle linjer fra resultat af søgningen specificeret i Figur 1.

BAEKUR-B00	Okei ... Mamma þín er sorgmædd (svarti kjóllinn) en hún er í hvíld núna
BAEKUR-B00	klæða sig . Fyrst fóru þær í	ljósbláan kjól	eða kyrtil sem náði þeim niður á
BAEKUR-B0R	Textasafn: Úr bók	í rauða kjólinn	og fann skip við höfnina með matsölustað
BAEKUR-B0R	Titill: Valkyrjan	í rauða kjólnum	eftir strætum New York borgar . Af
BAEKUR-B0R	Höfundur: Elías Snæland Jónsson	í bláa kjólinn	með rauðu rósunum frá Sýrlandi og hafði
BAEKUR-B0S	Útgefandi: Vaka-Helgafell	í fínasta kjólnum	sinum úr gullfallegum Afríkuefnum með borða yfir
BAEKUR-B0T	Ár: 2003	« Flottur kjóll	, » sagði Arnþrúður sem komið hafði
BAEKUR-B0T	God , serðu utganginn á mér !	Hvernig kjóllinn	er farinn ! Hvernig ég lit út
BAEKUR-B1A	Og barna er mynd af mömmu í	fallegum kjól	að syngja . Barna er líka innrömmuð

Figur 3. Nogle linjer fra resultat af søgningen specificeret i Figur 1

Finansiering og tak

Korpusprojektet var hovedsagelig finansieret af undervisningsministeriets sprogteknologiske projekt, som lanceredes i 2000. Men det er også blevet finansieret fra andre kilder. Man kan nævne forskningsprojektet ”Tilbrigði í setningagerð”¹⁰, som har leveret talesprogs materiale til *MÍM*, Nordisk Ministerråd (Nordisk Netordbog), Rannís¹¹ – Det islandske forskningsfond (Viable Language Technology beyond English – Icelandic as a test case¹²), Islandske studenter innovationsfond¹³ (nogle stipendier), Universitetets forskningsfond (nogle stipendier) og META-NORD¹⁴. Forfatteren vil også gerne takke Eiríkur Rögnvaldsson og Halldóra Jónsdóttir for at læse manuskriptet med henblik på indhold og sprog.

Sigrún Helgadóttir er statistiker og har arbejdet med sprogteknologi på Árni Magnússon-instituttet for islandske studier i mange år.

Emneord: korpus, opmærkning, taggers, sprogteknologi, islandsk

Litteratur

Eiríkur Rögnvaldsson og Sigrún Helgadóttir, 2011: Morphosyntactic Tagging of Old Icelandic Texts and Its Use in Studying Syntactic Variation and Change. I: Caroline Sporleder, Antal P.J. van den Bosch og Kalliopi A. Zervanou (red.): *Language Technology for Cultural Heritage: Selec-*

10 http://malvis.hi.is/tilbrigdi_i_setningagerd

11 <http://rannis.is/>

12 <http://iceblark.wordpress.com/>

13 <http://rannis.is/funding/icelandic-student-innovation-fund/>

14 <http://www.meta-nord.eu/>

- ted Papers from the LaTeCH Workshop Series*. s. 63–76. Berlín: Springer.
- Hrafn Loftsson og Eiríkur Rögnvaldsson, 2007: IceNLP: A Natural Language Processing Toolkit for Icelandic. I: *Proceedings of Interspeech 2007, Special Session: "Speech and language technology for less-resourced languages"*. Antwerpen.
- Hrafn Loftsson, J.H. Yngvason, S. Helgadóttir, E. Rögnvaldsson, 2010: Developing a PoS-tagged corpus using existing tools. I: *Proceedings of "Creation and use of basic lexical resources for less-resourced languages", workshop at the 7th International Conference on Language Resources and Evaluation (LREC 2010)*. Valetta, Malta.
- Hrafn Loftsson, S. Helgadóttir, E. Rögnvaldsson, 2011: Using a morphological database to increase the accuracy in PoS tagging. I: *Proceedings of Recent Advances in Natural Language Processing, RANLP*. Hissar, Bulgaria.
- Höskuldur Thráinsson, Á. Angantýsson, Á. Svavarsdóttir, T. Eythórsson, J. G. Jónsson, 2007. The Icelandic (Pilot) Project in ScanDiaSyn. I: *Nordlyd*, 34(1), s. 87–124.
- Jörgen Pind, F.Magnússon, S. Briem, 1991: *Íslensk orðtíðnibók* [Den islandske frekvensordbog]. Reykjavík: Leksikografisk institut ved Islands Universitet.
- Kristín Bjarnadóttir, 2012. The Database of Modern Icelandic Inflection. I: *Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages – SaLTMiL 8 – AfLaT2012*. Istanbul, Tyrkiet.
- Rögnvaldur Ólafsson, E. Rögnvaldsson, Þ. Sigurðsson, 1999: *Tungutækni* [Sprogteknologi]. Skýrsla starfshóps [Komité rapport], Reykjavík: Menntamálaráðuneytið [Undervisnings-, forsknings- og kulturministeriet].
- Sigrún Helgadóttir, Á. Svavarsdóttir, E. Rögnvaldsson, K. Bjarnadóttir, H. Loftsson, 2012: The Tagged Icelandic Corpus (MÍM). I: *Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages -SaLTMiL 8 – AfLaT2012*. Istanbul, Tyrkiet.
- Sigrún Helgadóttir, 2013: Mörkuð íslensk málheild. *Skíma*, 36(1), s. 26–27. Reykjavík.

Summary

The article describes the development of a morphosyntactically tagged corpus of Icelandic, the MÍM corpus. The corpus consists of about 25 million tokens of contemporary Icelandic texts from the years 2000–2010, collected from varied sources during the years 2006–2010. The corpus is intended for use in Language Technology projects and for linguistic research. Text collection and permission clearance is described. Text cleaning and annotation phases are also described. The corpus has been available since spring 2013 for search through a web interface and for download in TEI-conformant XML format. Two other corpora available from the same web pages are also described.

Status for samisk språkteknologi

Sjur Nørstebø Moshagen

Heilt frå utviklinga av den aller fyrste språkteknologien – det skrivne språket – og fram til i dag har språkteknologien vore både ei hjelp og ei grense. Hjelp for dei som har hatt han, og eit hinder for deltaking på like vilkår for dei som ikkje har hatt han, og ofte som eit middel for språkundertrykking overfor dei som ikkje har hatt han. Artikkelen gjev ei kort innføring i historia til samiske språk i eit språkteknologisk ljøs, og korleis utviklinga av språkteknologi har vore med på å forma relasjonen til andre språk. Deretter gjev artikkelen ei oversikt over statusen for samisk språkteknologi i 2014, kva som er dei viktigaste ideane bakom arbeidet som er gjort, og kva som bør gjerast framover.

Bakgrunn

Hovudtanken bak arbeidet med samisk språkteknologi er at språka skal overleva og ha ei framtid som bruksspråk for heile det samiske samfunnet. Eit språk som ikkje blir brukt, er i praksis eit dødt språk. I og med at samfunnet er så gjennomdigitalisert som det er, betyr det at for å kunna brukast må dei samiske språka vera tilgjengelege og brukande i alle digitale samanhengar.

Språkteknologi og samfunn

Føresetnadene for å nytta språk, og samanhengane dei blir nytta i, har vorte dramatisk endra over dei siste fem hundre åra, og aller mest dei siste hundre. Dette heng i hop med den språkteknologiske utviklinga i vid forstand, og dei endringane som ho har ført med seg for samfunnet.

Kort skissert kan utviklinga summerast opp i desse punkta:

- utviklinga av skriftspråk og alfabet, med ulike teknologi (kiler og leire, kniv og tre, penn og papir)
- utviklinga av boktrykkjarkunsten

- utviklinga av massemedium og innføringa av massekommunikasjon
- utviklinga av datateknologi

Før noko av dette vart utvikla, var bruken av språk berre avhengig av menneska sjølve. Talen var den einaste modusen (eg ber teiknspråklege om orsaking for at eg forenkler litt), og overføringa av eit språk frå ein generasjon til neste gjekk naturleg og sjølvsgt.

Men med innføringa av språkteknologi endra dette seg. Kvart steg i den teknologiske utviklinga har ført til nye barrierar i bruken av språka, som oftast kopla til dei samfunnsmessige endringane som fylgde med den teknologiske utviklinga.

Med innføringa av skriftspråk kom det fyrste skiljet, mellom dei som har eit skriftspråk og dei som ikkje har. Dei fyrste samiske tekstene er frå 1600-talet, og med ordlister og bibelomsetjingar utover 1700- og 1800-talet. Ein eigentleg skriftspråkstradisjon kan ein ikkje prata om før kring 1900, og sjølv då er han svært avgrensa. Enno i dag er det samiske språk som ikkje har ei etablert skriftnorm, til liks med nesten fem tusen andre språk¹.

Innføringa av skriftspråk i ulike samfunn førte til store samfunnsmessige endringar og framsteg, samtidig som skriftspråka vart ein delar: deltaking i samfunnet føreset bruken av skriftspråket, og bruken av andre språk blir dermed indirekte (eller direkte) undertrykt, med mindre samfunnet aktivt tillét at ein nyttar fleire skriftspråk parallelt. Sjølv om det finst mange døme på slik parallell bruk, er dei unnatak heller enn regel i den store samanhengen.

Utviklinga av boktrykkjarkunsten var fyrste steg på vegen mot massekommunikasjon, med Internett og sosiale media som dei nyaste stega på den vegen. Dei kommunikasjonsbaserte samfunna vi ser i dag, er svært skriftspråksbundne, og bruken av språkteknologi er ein grunnleggjande føresetnad for å kunna delta i samfunnet. Ein kan ikkje delta på eigne premiss og eige språk om språket ikkje er skriftfest, eller om det tilhøyrande skriftspråket ikkje er kodifisert i Unicode². Og sjølv om skriftsymbola i språket er koda i Unicode, er det ikkje sikkert at det finst i tilgjenge-

1 Jf. [Ethnologue-statistikk](#) – språk utan skriftspråk er rekna som summen av EGIDS-kategoriane 6a-9, dvs. 4889 av i alt 7105 språk (pr. 4.3.2012). Dette rimar bra med andre kjelder, t.d. føreordet til [Thousand Languages](#), som viser til at kring 2/3 av språka i verda er utan skriftspråk.

2 <http://www.unicode.org>

lege skrifter for datamaskiner, eller at det finst tastaturopsett som gjer at ein kan skriva det. Alt dette er språkteknologi som t.d. norsk- og svensk-talande tek som sjølvst, men som ikkje finst for store delar av språka i verda (Moshagen & Trosterud, 2008).

Ei viktig drivkraft bak arbeidet med samisk språkteknologi er altså å gje det samiske samfunnet dei same verktøya og den same teknologien som finst for majoritetsspråka kring dei, slik at samisk i framtida òg har dei naudsynte språkteknologiske føresetnadene for å bli brukte. Utan språkteknologi blir språkbruken privatisert og utestengd frå det meste av samfunnet, og vil til slutt ikkje eksistera, og då er språka daude.

Domenetap og domenevinning

Ei anna side av denne utviklinga og andre endringar i samfunnet er knytt til dei samfunnsområda, domena, som språket kan bli brukt i og på.

Alle språk har i utgangspunktet vore samfunnsberande, dvs. vore det einaste språket ein har trengt for å klara seg i det samfunnet ein lever i. Men i samband med utviklinga av dei moderne statane, og særleg knytt til utviklinga av nasjonalstaten, har bruken av andre enn det statsberande språket vorte marginalisert, og i dei nordiske landa – og i mange andre land – vorte aktivt undertrykt.

Dette har ført til store domenetap for samisk. Sjølv om den offisielle undertrykkingspolitikken no er avslutta, er samisk i praksis framleis ute-stengt frå svært mange domene, av ulike grunnar. Ei anna viktig drivkraft for arbeidet med samisk språkteknologi er difor å gje verktøy for å ta tilbake og vinna nye domene, og å ta tilbake tapte talarar.

Sosial status og demografi/stigma

Ein konsekvens av samfunnsendringane som er nemnde over, er at statusen til språka har endra seg kraftig, frå å vera heilt sjølvst og naturleg til å bli kraftig stigmatisert som ein konsekvens av undertrykkingspolitikken. I dag er stigmaet borte mange stader, men ikkje over alt, og ikkje for alle.

Sjølv om språkteknologi truleg spelar ei avgrensa rolle når det gjeld sosial status og stigma, så vil han vera ein av mange faktorar som dreg i rett retning: det at samisk har dei same verktøya som t.d. norsk, er eitt teikn på at dei er likeverdige.

Tilhøve mellom majoritet og minoritet

Det er likevel langt att til fullt likeverde mellom samisk og majoritetsspråka. All kommunikasjon mellom språkgruppene skjer på premissa til majoritetsspråka, *på* majoritetsspråka. Det er heile tida minoriteten som må tilpassa seg. Dette avgrensar talet på arenaer der samisk kan bli brukt kraftig, og er eit stort hinder for ein meir utbreidd bruk av samisk.

For å få ein betre balanse mellom språkgruppene burde kommunikasjonen gjerast meir på premissa til minoriteten. Det var slik det ofte fungerte før dei moderne statskonstruksjonane vart bygde, då kommunikasjonen mellom sentralmakt og dei ulike folkeslaga typisk gjekk gjennom tolk. Tolking vil alltid gje informasjonstap, så enno betre er det sjølvsagt om sentralmakta lærer seg språka til dei samfunna ein skal administrera. Og når det gjeld tolking, omsetjing og språklæring, kan språkteknologi vera til hjelp.

Institusjonsbruk

Med slike verktøy vil det på sikt bli mogleg for samiske institusjonar å fungera heilt på samisk sjølv om dei kommuniserer mykje med majoritetssamfunnet. I dag fungerer dei aller fleste heilt på norsk, svensk eller finsk. Om ikkje sametinga kan fungera på samisk, kven kan då gjera det? Kva slags signal gjev det til både det samiske samfunnet og majoritetssamfunnet at sjølve symbola på det samiske folket ikkje opererer på samisk?

eSápmi

eSápmi var eit prosjekt starta i 2002 med sikte på å planleggja og førebu ulike satsingar for å gjera samisk tilgjengeleg og brukande på digitale plattformer, nettopp med utgangspunkt i at dei samiske språka må finnast på slike plattformer for at dei skal bli brukte i framtida. Mykje av det som har vorte gjort seinare vart alt skildra i planane som vart utforma den gongen. eSápmi har såleis vore den språkpolitiske grunnsteinen som alt arbeid med språkteknologi hjå det norske Sametinget har bygt på, og framleis er det prosjekt og teknologi som vart planlagt den gongen som enno ikkje er ferdige. Det gjeld til dømes talesyntese for samisk, der eit prosjekt for nordsamisk er i gang og vil vera ferdig i 2014.

Oppsummering av bakgrunnen for arbeidet med samisk språkteknologi

Som teksten over viser, trengst språkteknologi i vid forstand for at samiske språk skal ha ei framtid og haldast i aktiv bruk. Dei må kunna brukast på datamaskiner og andre digitale plattformer. Det vil vera med på å heva statusen og hjelpe til med å ta att tapte domene og vinna nye. Motsett vil mangel på tilgang til slik teknologi forsterka tendensane til språkdaude.

Status i byrjinga av 2014

Elementær databruk

Teiknsett: Innføringa av Unicode var eit enormt framsteg for alle språk i verda. Ved at det er ein altomfattande standard som ligg til grunn for all moderne tekstkoding på alle digitale plattformer, betyr det at det ikkje lenger er eit problem å koda tekst for databruk, uansett språk, inklusive utdøyddde språk. Alle samiske språk er representerte i Unicode, og for språk som enno ikkje er det, finst det standardiserte rutinar for korleis ein skal leggja til nye data og oppdatera standarden. Unicode blir brukt over alt, frå minimale mobiltelefonar til store serversystem, og ein treng ikkje lenger gå ut frå at om ein sender ein tekst frå maskin A til maskin B, så er teksten forvrengd eller øydelagd når han kjem fram. Sjølv om det framleis hender, har det gått frå å vera regel til å vera unnatak.

Skrifter: Unicode er ikkje nok, Unicode er berre ein standard for korleis tekst skal kodast inne i datamaskina. For å visa teksten trengst det skrifter, og at dei tilgjengelege skriftene har alle teikn som trengst. Dessverre er det ein hovudregel at skriftene *ikkje* har dei samiske teikna. Det finst ingen standard for skrifter tilsvarande Unicode for teiknkoding. Det finst i dag i praksis to vegar fram for å skapa fleire skrifter med full samisk dekning: å støtta utviklinga av opne skrifter, og å krevja full samisk dekning i offentlege innkjøpsavtaler av datasystem.

Tastatur: dette vart løyst for Linux, MacOSX og Windows kring 2003, og er ikkje lenger ei sak for desse plattformene. Dessverre har ikkje dei løysingane vorte overførte til dei nye mobile systema Android, iOS, og Windows Phone/RT, og dermed er vi tilbake på rute 1. Det er enno ikkje mogleg å skriva dei samiske språka på mobile system, men Unicode-organisasjonen leier eit initiativ – CLDR³ – for å samla og halda ved like både tasta-

3 <http://cldr.unicode.org/index>

turdata (frå versjon 22 av CLDR) og andre data (sjå neste avsnitt). Sjølv om ikkje alle OS er dekte av tastaturopsett skrivne i dette dataformatet (både iOS og Windows Mobile/RT manglar i lista over støtta OS), vil det vera eit viktig steg framover å leggja til støtte for samiske språk der det manglar.

Sortering og andre grunnleggjande data om språk: Dei aller fleste data-system føreset visse grunnleggjande data om eit språk for å kunna handtera det skikkeleg, og mangel på slike data fører ofte til at eit språk heller blir fjerna frå eit system enn at det blir inkludert med mangelfull støtte. Data det gjeld er slike ting som sorteringsrekkefølge, format for dato og tid, namn på land og språk, osb. CLDR (sjå førre avsnitt) inneheld alle slike data, men berre nordsamisk er dekt i rimeleg grad. Det vil vera eit viktig arbeid i åra som kjem å leggja til data for dei andre samiske språka.

Annan databruk

Sjølv om det er mogleg å skilja mellom det ein kunne kalla *vanleg* bruk og meir *avansert* bruk, så er overgangen mellom dei flytande, og av og til føreset såkalla vanleg bruk at ein i utviklingsarbeidet tek vegen om meir avansert bruk. Heile spekteret er difor samla i dette avsnittet. Inndelinga under fylgjer i staden ei inndeling etter kva språk tekstprodusent og motakar har.

For fyrstespråksbrukarar

Den overordna målsetjinga er å tilby dei same verktøya og tenestene som for majoritetsspråka, og samtidig tilby verkøty og tenester som gjer at samisktalande vel å skriva på samisk heller enn på eit majoritetsspråk. Til dette trengst det i alle fall stavekontrollar⁴, grammatikkontrollar, ordbøker⁵, terminologi⁶, omsetjingsverktøy⁷ og talesyntese.

Talesyntese er viktig av fleire grunnar, både i situasjonar der ein sjølv vanskeleg kan lesa, og i skrivesituasjonar. Ein stor del av den vaksne samiske befolkninga har ikkje fått skriveopplæring, og har difor vanskeleg for å uttrykkja seg i skrift sjølv om dei har samisk som morsmål. Talesyntese vil i kombinasjon med ein stavekontroll og ein grammatikkontroll gjera det mogleg for mange av desse å byrja å uttrykkja seg skriftleg, ved

4 divvun.no

5 satni.org og sanit.oahpa.no

6 <http://gtsvn.uit.no/termwiki/index.php/Váldosiidu>

7 <http://divvun.no/doc/tools/autshumato.html>

at dei fyrst nyttar korrekturverktøya til å retta så mange feil som mogleg, og deretter nyttar talesyntesen til å sjekka at teksten høyrer rett ut.

For *nordsamisk* finst det meste, og dei manglande verktøya grammatikkontroll og talesyntese er under utvikling. For *lule-* og *sørsamisk* finst stavekontrollar, ordbøker og noko terminologi, medan resten av verktøya ikkje er påbyrja. Dei manglande verktøya vil kunna byggja på det arbeidet som blir gjort med nordsamisk, men det er opplagt at det må leggjast ned mykje arbeid for å gjera alle verktøy tilgjengelege for desse språka.

For alle andre samiske språk har det vorte starta arbeid på morfologiske analysatorar og elektroniske ordbøker, og i nokon mon terminologi. Men det meste er ugjort. Dei vil likevel dra stor nytte av arbeidet som blir gjort for nord-, lule- og sørsamisk, i form av at dei kan nytta ein ferdig infrastruktur som vil gje dei ferdige rammeverk og malar for å utvikla alle desse verktøya. Dei treng ikkje finna opp hjulet på nytt, det er nok å fylla det med språkleg innhald (men det er sjølvsagt ein stor jobb, og like stor uansett om det finst 50 eller 50 000 talarar), t.d. kan dei morfologiske analysatorane direkte byggjast som stavekontroll for LibreOffice⁸.

For andrespråksbrukarar

Målsetjinga for slike brukarar er dels dei same som for fyrstespråksbrukarar, det vil seia tekstproduksjon på samisk, kanskje med verktøy tilpassa dei behova andrespråkstalarar har. I tillegg er det eit mål å få fleire samisktalande.

Andrespråkstalarar har delvis andre behov enn fyrstespråkstalarar, fordi dei typisk gjer andre feil og har ein annan språkleg bakgrunn. Samtidig er grensa mellom fyrste- og andrespråkstalarar mykje meir flytande og breiare for samisk enn for t.d. norsk. Det finst eit breitt spekter av ulik kompetanse i syntaks, bøyning, ordforråd og idiom, og det er ikkje mogleg å setja opp klare kategoriar basert på språkleg kompetanse. I tillegg til dette kjem det konstante presset frå majoritetsspråka som gjer at sjølv morsmålstalarar ofte har mykje språkleg materiale frå majoritetsspråket med i det samiske språket sitt.

Det er opplagt at ordbøker, terminologisamlingar og omsetjingsverktøy er viktige for andrespråksbrukarar, til liks med ulike skrivestøtteprogram. Dei eksisterande stavekontrollane er sjølvsagt til hjelp for desse brukarane òg, men det har vist seg at andrespråksbrukarar og personar

8 <http://divvun.no/libreofficeoxt.html> (beta)

som lærer seg samisk innimellom, har problem med å bruka verktøya fordi stavekontrollen tillèt ordformer som er marginale men korrekte, og der brukaren gjer skrivefeil som stavekontrollen godtek fordi skrivefeilen tilfeldigvis er identisk med slike marginale former. Løysinga på dette vil vera å laga stavekontrollar tilpassa ulike grupper av andrespråksbrukarar. Dette har enno ikkje vorte gjort, men rammeverket for å laga slike er stort sett på plass.

Tilsvarande vil det sikkert vera behov for å laga ein grammatikkontroll tilpassa andrespråksbrukarar. Eit slikt arbeid ligg lenger fram, og må venta til den fyrste grammatikkontrollen for morsmåstalarar er ferdig.

I tillegg til verktøy for tekstproduksjon er verktøy for språkopplæring viktige for denne brukargruppa, både for dei som kan litt og dei som kan ingen ting. Det finst ei stor gruppe ungdomar og vaksne som ikkje har fått opplæring i det samiske språket av ulike grunnar, og som er svært motiverte for å læra seg det. I byrjinga av 2014 fanst det språkopplæringsprogram⁹ for nordsamisk og sørsamisk, og enklare versjonar for skolt samisk, enaresamisk og kildinsamisk. Det blir òg utvikla språkopplæringsprogram for ikkje-samiske språk med utgangspunkt i den same plattformen.

For kommunikasjon frå minoritet til majoritet

Den overgripande målsetjinga her er å gjera det mogleg for minoriteten å nytta sitt eige språk mest mogleg i staden for at samisktalande alltid skal byta til majoritetsspråket.

Kommunikasjonen er anten skriftleg eller munnleg. Skriftleg kommunikasjon kan ein dela i to: tekst der detaljane og tekstkvaliteten er viktig, og tekst der det viktigaste er å få med seg hovuddraga i innhaldet. Desse to kommunikasjonstypene krev ulike verktøy.

For at munnleg kommunikasjon skal fungera mellom to språk, og slik at begge partar kan nytta sitt eige språk, krevst det tolk. Det finst enno ikkje språkteknologi som kan nyttast i staden for tolkar, men det finst fleire verktøy som kan vera til hjelp for tolkar. Dette vil typisk vera elektroniske ordbøker og termsamlingar. Som nemnt tidlegare, så finst det slike, men omfanget av ordbøkene og termsamlingane er ikkje stort nok for mange tolkeformål, og det meste av innhaldet er skriva for språkparet nordsamisk-norsk. Det er altså mykje arbeid som står att for alle samiske språk.

9 <http://oahpa.no>

For skriftleg innhaldsformidling vil det ofte vera nok med maskinomsetjing. Sjølv om resultatet som oftast er langt frå perfekt, er det likevel som regel bra nok til å formidla innhaldet i ein tekst i grove drag. Det finst i dag slik maskinomsetjing frå nordsamisk til norsk bokmål¹⁰. Det finst ingen ting for dei andre samiske språka.

For å produsera presisjonstekst krevst det andre verktøy. Målsetjinga er at teksten skal kunna produserast på samisk, og at ein så effektivt som mogleg kan omsetja teksten til eitt av majoritetsspråka etterpå. Relevante verktøy er maskinomsetjing, omsetjingsminne, ordbøker, termsamlingar og korrekturverktøy på målspråket.

Det er mogleg å nytta *maskinomsetjing* som eit fyrste steg i ein omsetjingsprosess og retta opp teksten etterpå, men kvaliteten er i dag for dårleg til at ein sparar tid på det, og går som nemnt berre frå nordsamisk til norsk. Over tid vil maskinomsetjinga bli betre og gje meir hjelp i omsetjingsprosessen, og truleg vil det koma fleire språkpar, både frå andre samiske språk og til dei andre majoritetsspråka.

Dei viktigaste verktøya vil vera *omsetjingsminne* og *parallellkorpus* – som langt på veg er to sider av same sak. Med eit omsetjingsminne vil ein få forslag til omsetjing av setningar basert på tidlegare omsetjingar, og for ein god del byråkratisk tekst med mange repetitive element og liknande tekststrukturar frå år til år burde det vera mogleg å spara mykje tid i omsetjingsprosessen utan at dette går ut over kvaliteten. Eit parallellkorpus inneheld i praksis dei same parallelle setningane, men gjer det mogleg å søkja manuelt i staden for å lita på dei automatiske algoritmane i omsetjingsminnet. Sidan svært mykje av tekstproduksjonen til no har vorte gjort på norsk, finst det nesten ikkje parallelltekstar frå nordsamisk til norsk, og dermed heller ikkje noko omsetjingsminne. Derimot finst det eit norsk-nordsamisk parallellkorpus som det er mogleg å søkja i¹¹. Det finst enno ikkje tilsvarande for dei andre samiske språka, men eit nytt prosjekt vil byggja opp parallellkorpus for norsk-lulesamisk og norsk-sørsamisk.

Andre viktige verktøy for ein omsetjar vil vera ordbøker og termsamlingar. Det finst slike for dei fleste samiske språka (jf. fotnote 5 og 6), men omfanget av dei er svært avgrensa for alle andre språk enn nordsamisk. Til sist vil omsetjarane sjølvsgt ha nytte av korrekturverktøy på målspråka, og slike verktøy finst for både finsk, svensk og dei to norske skriftspråka.

10 <http://gtweb.uit.no/mt/index.php?lang=nno>

11 http://gtweb.uit.no/korp/?mode=parallel#parallel_corpora=nob&lang=nb

For kommunikasjon frå majoritet til minoritet

Dei språkpolitiske målsetjingane for denne kommunikasjonsretninga bør vera at kommunikasjonen skjer på minoritetsspråket, av same grunn som diskutert tidlegare: det er minoriteten som er pressa språkleg, og som har alt å vinna på å nytta sitt eige språk. Det er altså viktig for å oppnå ein betre balanse mellom språka.

I denne samanhengen vil det i utgangspunktet alltid vera feil å nytta maskinomsetjing. Maskinomsetjing gjev i praksis alltid dårlegare resultat enn manuell omsetjing, og sidan minoriteten alltid kan majoritetsspråket, vil det å nytta maskinomsetjing fort verka mot sin eigen hensikt – det vil få språkbrukarane til å venda seg bort frå den samiske teksten. Ein annan grunn til å unngå maskinomsetjing sjølv om han skulle bli brukbar, er at omsetjinga alltid vil ha ganske mange drag av originalspråket i seg. Så sjølv om t.d. ein nordsamisk maskinomsett tekst vil vera korrekt nordsamisk, vil det framleis vera ein veldig norskfarga nordsamisk tekst. Dette gjeld òg sjølv om ein språkvaskar teksten etterpå.

Dette betyr at ein bør satsa på manuell omsetjing, med støtte av verkøy som omsetjingsminne, ordbøker, terminologisamlingar og parallellkorpus. Som det vart nemnt i førre avsnitt så finst det eit parallellkorpus frå norsk bokmål til nordsamisk. Det inneheld ca. 156 000 setningspar, frå i all hovudsak administrative tekstar. Tilsvarende finst det eit omsetjingsminne basert på dette korpuset, til hjelp ved omsetjing av slike tekstar. Og det finst ordbøker og termsamlingar, som òg har vorte nemnt før. Men som det alt har vorte klårt, så finst det meste berre for nordsamisk og språkparet norsk bokmål-nordsamisk. Det finst lite eller ingen ting for dei andre samiske språka, og dei potensielle parallellkorpusa er òg så små at dei ikkje vil vera til stor hjelp.

For mange av dei samiske språka vil det difor kunna vera ei brukbar løysing å basera seg på omsetjingane til nordsamisk:

Kommunikasjon mellom minoritetsspråka

Språkpolitisk vil det vera mykje betre om dei samiske språkbrukarane kommuniserer på samisk seg i mellom enn at dei tyr til eitt av majoritetsspråka når dei skal kommunisera på tvers av dei samiske språkgrensene.

For munnleg kommunikasjon trengst det opplæring i nabospråksforståing og ordbøker mellom dei samiske språka. Ingen av delane finst

enno, men eksisterande termsamlingar inneheld noko terminologi for fleire samiske språk, og kan difor vera til hjelp.

For skriftleg kommunikasjon vil settet av verktøy langt på veg vera det same som for kommunikasjon frå minoritetsspråk til majoritetsspråk (dvs. maskinomsetjing, omsetjingsminne, terminologisamlingar og parallellkorpus). Men det er ein viktig skilnad: sidan språka er så like, er det truleg mykje som kan overførast direkte frå eitt språk til eit anna heller enn å omsetjast. Og det er ikkje umogleg å laga maskinomsetjing som vil gjera det raskare å vaska teksten etterpå enn å omsetja han manuelt. Eit fyrste eksperiment med maskinomsetjing frå nordsamisk til sørsamisk har vorte gjennomført, med lovande resultat (sjå Antonsen, Tyers og Trosterud (i kjømda)).

Problema med å nytta maskinomsetjing som vart skildra over, for kommunikasjon frå majoritetsspråk til minoritetsspråk, er heller ikkje til stades i same grad nettopp fordi språka er så like: ein ikkje-neglisjerbar del av leksikonet vil vera felles, mykje av syntaksen er lik, og morfologi og ordstruktur er stort sett lik. Dessutan er det rimeleg å tru at det vil vera betre med eit preg av eit anna samisk språk enn at majoritetsspråket skin gjennom på ulike vis. Dette betyr sjølvsagt ikkje at det enkelt å få eit fungerande system, men resultatata frå eksperimentet med nordsamisk-sørsamisk er så lovande at det no startar eit stort prosjekt ved UiT Norges arktiske universitet som skal gå over fleire år og omfattar både studentar, ein doktorgrads-student og ei postdok.-stilling.

Oppsummeringsvis finst det altså enno svært lite språkteknologi og ressursar for kommunikasjon mellom dei samiske språka, men det er arbeid på gang, og det vil truleg vera synlege resultat om nokre år.

For forskarar

Bakom dei fleste av dei omtalde ressursane og verktøya ligg det ulike verktøy og ressursar meir retta mot forskarar, og utvikla av språkforskarar. Det gjeld til dømes slike ting som *morfologisk analyse*. Eit analyseprogram er mest nyttig for forskarar, men er samtidig grunnmuren i all annan språkprosessering, inklusive sluttbrukarverktøya.

Både syntaktisk analyse og analyserte korpus er svært viktige verktøy for både forskarar og til dømes terminologar. I praksis betyr dei tilgjengelege korpusressursane for nordsamisk nærmast eit paradigmeskifte for

forsking på samisk syntaks ved at ein stor del av den samiske tekstproduksjonen er tilgjengeleg på nettet, og slik at ein kan søkja både på ordformer, grunnformer og ulike syntaktiske kontekstar. Etter at prosjektet med talesyntese er avslutta, vil det òg gje verktøy som opnar nye høve til forsking, t.d. kring intonasjon og uttale.

Oppsummering av ressursar og verktøy som finst

Nedanfor er statusen for dei ulike språka oppsummert i tabellform.

	Nordsamisk	Lulesamisk	Sørsamisk	Enaresamisk	Skoltesamisk	Kildinsamisk	Pitesamisk	Umesamisk
I Unicode	3	3	3	3	3	3	3	3
Har alle bokstavane i minst ein systemfont	3	3	3	3	3	3	3	3
Har tastatur for datamaskin	3	2	2	2	2	2	1	1
Kan skrivast på mobilsystem	1	1	1	1			1	
Ordliste på mobilsystem								
morfologisk analysator	3	3	3	1	1	1	1	1
retteprogram	3	3	3	1	1	1	1	1
automatisk orddeling	3	3	3					
grammatikkontroll	1							
elektroniske ordbøker mellom samisk og majoritetsspråk	3	2	3		2	2		
elektroniske ordbøker mellom dei samiske språka	1	1						
terminologisamlingar	3	2	2	2	2			
maskinomsetjing frå samisk til majoritetsspråk	3							
maskinomsetjing frå samisk til samisk	1	1	1					
omsetjingsminne	3							
korpus	3	2	2			1		
parallellkorpus	3	1	1			1		
syntaktisk analyse	3	2	2					
talesyntese	2							
språklæringsprogram	3		3	2	2	2		

Tabell over verktøy og ressursar for dei ulike samiske språka.

Finst ikkje	
Påbyrja / finst for nokre system	1
Under utvikling / finst for dei fleste systema	2
Ferdig versjon / finst for alle system	3

Nykel til fargekodane i tabellen.

Infrastruktur for å byggja ressursar og verktøy

I tillegg til dei verktøya og ressursane som er skildra over, har det vorte bygt opp ein infrastruktur som gjer det mogleg å produsera mange av ressursane og verktøya over, ut frå rådata eller frå eit sett med grunnressursar. Til dømes blir morfologisk analysator (og generator), stavekontrollar og orddelingsprogram i all hovudsak bygde frå den same kjeldekoden, og morfologisk analyse er ein komponent i fleire av dei andre verktøya.

Med ein felles infrastruktur vert det mogleg å produsera langt fleire verktøy for mange fleire språk, sjølv om det grunnleggjande språkarbeidet alltid må gjerast for kvart enkelt språk. I mars 2014 var det 42 språk som det blir jobba aktivt med i denne infrastrukturen, og alle desse språka vil automatisk få fleire av dei nemnde verktøya automatisk når ein utviklar den morfologiske analysatoren.

Slik automatisk gjenbruk er viktig for språksamfunn med svært avgrensa middel og menneskelege ressursar ein kan leggja ned på dette arbeidet. I tillegg blir alle språkteknologiprojekt oppmoda til å nytta ein så open lisens som mogleg, slik at ein ikkje stengjer ute andre som vil nytta kjeldekoden til andre formål. På den måten vil ein få mest mogleg att for det arbeidet ein legg ned på å utvikla dei språkteknologiske ressursane.

Oppsummering og framtidstankar

Situasjonen for nordsamisk er på mange måtar bra, mange av verktøya og ressursane som har vorte nemnde for at samisk skal kunna fungera heilt samfunnsberande igjen, er på plass, om enn i mindre målestokk enn for andre språk, og mindre omfattande enn det ein treng. Ein kan seia at for nordsamisk er veldig mange av verktøya og ressursane i versjon 1.0, og det trengst mykje arbeid enno for å utvikla dei vidare. På den andre sida har arbeidet med nordsamisk lagt grunnen for tilsvarende arbeid for andre språk, og gjort det mykje lettare for dei å koma i gang med tilsvarende

arbeid. Det same kan seiast om arbeidet med å byggja ein infrastruktur som støttar opp om gjenbruk, felles bruk og eit felles sett med verktøy.

Situasjonen for lule- og sørsamisk er noko dårlegare enn for nordsamisk. Dei viktigaste ressursane og verktøya er på plass, men mykje står att. Dette gjeld særleg korpusressursar, der hovudproblemet er at det finst så få tekstar å leggja inn i eit korpus. På sikt kan ein vona at tilgangen til dei nye verktøya vil stimulera til meir tekstproduksjon som vil gje meir tekst til forskings- og utviklingsarbeid, som igjen vil gje betre verktøy. På den måten kunne ein få ein god sirkel.

For dei andre samiske språka er mykje av arbeidet enno i startgropa. Samtidig blir det arbeidd med dei aller fleste av dei, i større eller mindre grad. Og gjennom den felles infrastrukturen vil alt grunnarbeid som blir gjort for eitt språk, automatisk koma andre språk til gode. Slik sett ser det positivt ut.

Dei to største utfordringane når det gjeld språkteknologi for samiske språk, er å gjera dei praktisk brukbare på mobile system, og mangelen på tekstar. Dei mobile systema er mykje meir lukka enn dei etablerte systema for borddatamaskiner, og har ikkje overført grunnleggjande ressursar for skriving frå bordsystem til mobilsystem. Mobilsystema har difor i praksis vorte ei ny digital sperre for minoritetsspråk, nett som den førre digitale sperra var på veg til å bli rive ned. Sjølv om det no er mogleg å skriva t.d. nordsamisk på iOS-einingar, finst det ikkje støtte for ordlister og ordfullføring. Ein må nytta det finske tastaturet som betyr at ein for alle samiske teikn, må trykkja og halda på ein bokstav for å få fram det samiske teiknet. Berre dei mest ihuga språkbrukarane gjer det.

Mangelen på tekst på samiske språk er ikkje noko som enkelt kan løysast. Men synet på samisk har dramatisk endra seg til det betre over dei siste tjue-tretti åra, og i kombinasjon med dei skrivestøtteverktøya som er utvikla eller er på veg til å bli utvikla, kan situasjonen på sikt bli mykje betre.

Den teknologiske plattformen som har vorte bygt opp og heile tida blir utvikla, gjev eit godt grunnlag for språkteknologiske verktøy og ressursar for dei samiske språka. Samtidig finst det ei positiv og aukande interesse for dei samiske språka både i samfunnet generelt og i ulike finansieringsorgan. Det er difor grunn til å vera optimistisk når det gjeld utviklinga av samisk språkteknologi og verktøy og ressursar for det samiske samfunnet. Til sjuande og sist er det likevel andre faktorar som er avgjerande: språka må haldast i bruk og overførast frå ein generasjon til den neste. På

det punktet kan språkteknologien berre vera eitt av mange tiltak som aukar statusen til samisk og på det viset får fleire til å nytta språka.

Dei fleste samiske språksamfunna er svært små og sårbare, og har vorte hardt råka av språkpolitikken på 1900-talet. Haldningane og kunnskapen har heldigvis endra seg kraftig, men situasjonen er framleis svært kritisk for mange av dei samiske språka. Språkteknologiske verktøy og ressursar er berre ein del av arbeidet med å føra språka vidare. Men med det grunnlaget som no er bygt opp og den aktiviteten som har starta, så finst det håp om at språkteknologi kan gje ei viktig hjelp til dei samiske språka.

Sjur Nørstebø Moshagen er prosjektleder for Divvun-prosjektet ved Universitetet i Tromsø.

Summary

Language technology in all aspects has been a divider among languages since the very invention of writing. Today access to language technology tools that are taken for granted among the majority language speakers can be the factor that ensures a language will continue to be used in the future, and thus survive as a living language. The article goes briefly through the use of language technology among the Sámi languages, and then describes the present state. At the end there's a short look at the future work and possibilities for the Sámi languages and their use of language technology. Three aspects are important for future use: use of open standards, all Sámi resources should be available as open source, and the use of a shared, co-developed infrastructure to ensure that all languages get access to the same support and tool-set.

Litteratur

Antonsen, Lene, Tyers, Francis og Trosterud, Trond (i kjømda): A North Saami to South Saami machine translation prototype, i *Proceedings of the SALTMIL Workshop at LREC2014*

Moshagen, Sjur og Trosterud, Trond, 2008: Datorstöd för samiska och andra minoritetsspråk, i *Tekniken bakom språket*, red. Domeij, Språkrådet, Sverige.

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 samfunds bærende 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 opridsrer 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 samfunds bærende 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, sprogsprogs målet 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 drænedes 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 skrivehjælp 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, *Kukkuníiaat*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 *Kukkuníiaat* 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 *Kukkuníiaat* er en neologisme, der direkte oversat betyder 'fejlfjerner'. 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 interesseret 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 entydigt 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 oversættelsestjenester 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-grønlændere, 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-grønlændere. 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. Komplicerede 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 tanker 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 ressourcedelingen. 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.

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.

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.

FIN-CLARIN – en humanistisk forskningsinfrastruktur med betoning på språk

Krister Lindén

Miljardvis med ord och tusentals timmar med audio och video behövs som material för humanistisk forskning och i synnerhet språkforskning. Dessutom behöver forskarna redskap för att förädla och jämföra sina egna datasamlingar med allmänna datasamlingar. När ett forskningsprojekt är slut behövs det lagrings- och spridningsplatser för att göra rådata, redskap och forskningsresultat tillgängliga och användbara. Data, redskap och gemensamma användningsmöjligheter bildar tillsammans en forskningsinfrastruktur, som gör det möjligt att verifiera tidigare resultat och effektivare göra nya rön, när alla inte behöver starta från noll med att samla data och bygga analysredskap.

FIN-CLARIN – en forskningsinfrastruktur

FIN-CLARIN är en forskningsinfrastruktur som tillhandahåller språkdata och språkredskap för humanistiska forskare. FIN-CLARIN är en nationell nod i Finland för CLARIN ERIC¹ (European Research Infrastructure Consortium) – den europeiska infrastrukturen för humanistisk forskning, som grundades den 29 februari 2012. Genom att samarbeta med andra inom CLARIN ERIC, kan FIN-CLARIN fokusera på de två officiella språken i Finland, dvs. finska och svenska, även om FIN-CLARIN också tillgängliggör andra språks resurser producerade av forskare i Finland.

FIN-CLARIN² är ett nationellt konsortium av universitet och forskningsinstitut i Finland. FIN-CLARIN är en distribuerad forskningsinfrastruktur, där de olika konsortiemedlemmarna bidrar med språkdatasamlingar och språkredskap från sina olika styrkeområden. FIN-CLARIN koordineras av Helsingfors universitet, som samlar in allmänt tillgängliga språkdataresurser och som i Finland har specialiserat sig på att bygga

1 www.clarin.eu

2 www.helsinki.fi/finclarin

språkteknologiska analysredskap. En viktig partner i samarbetet är CSC – Tieteen tietotekniikan keskus Oy³, som är ett IT-center för vetenskap, ägt av staten och administrerat av undervisnings- och kulturministeriet. CSC tillhandahåller en gemensam plats för lagring och sökning i dataresurser och användning av språkredskap. En annan viktig partner är Institutet för de inhemska språken⁴, som upprätthåller omfattande språkliga arkiv och samlingar.

FIN-CLARIN är en virtuell forskningsinfrastruktur som erbjuder två centraliserade tjänster: Språkbanken i Finland⁵ (Kielipankki) och Vetenskapstermbanken i Finland⁶ (Tieteen termipankki). För att rationalisera utvecklingen av språkresurser för svenska har Språkbanken i Finland speciellt värnat om samarbetet med Språkbanken i Göteborg, medan Vetenskapstermbanken i Finland samarbetar med Rikstermbanken i Sverige. Forskarna i samiska vid Universitetet i Tromsø använder redskap utvecklade av FIN-CLARIN och har börjat utveckla sin egen terminologiplattform med utgångspunkt i Vetenskapstermbanken i Finland. Förbundet för hörselskadade i Finland utvecklar en videokorpus och en videobaserad ordbok för teckenspråk där de kombinerar videosökredskap från Språkbanken i Finland och lexikografiska redskap från Vetenskapstermbanken.

FIN-CLARINs förberedelser inför CLARIN ERIC

FIN-CLARIN har förberett sig för Finlands inträde i CLARIN ERIC genom följande åtgärder:

1. Det autentiserings- och auktoriseringssystem som krävs för att få tillgång till språkdata inom CLARIN ERIC har implementerats av CSC med delfinansiering från Helsingfors universitet och det har testats av Finland, Tyskland och Holland i form av en tjänstefederation som kallas CLARIN SPF för att alla CLARIN-center ska få tillgång till varandras tjänster.
2. FIN-CLARIN erbjuder forskningsmaterial till hundratals forskare och tusentals studerande i Finland och många fler i hela Europa. CLARIN

3 www.csc.fi

4 www.sprakinstitutet.fi

5 www.kielipankki.fi

6 www.tieteentermipankki.fi

och FIN-CLARIN erbjuder öppna sökportaler⁷ för språkdata i Europa och för att informera användare om vilka språkresurser som finns, var de finns och på vilka villkor de kan användas. En del material kan användas genom sökgränssnitt som Korp⁸ för textmaterial och LAT⁹ för audio- and videomaterial.

3. Licenstemplat för deponering och användning av språkresurser har skapats av Helsingfors universitet och de kan användas inom hela CLARIN (Oksanen & al., 2010).
4. CLARIN skapade en pilot för gemensam användning av språkredskap och språkdata som befinner sig på olika CLARIN-center genom en molntjänst kallad Weblicht. Språkredskap placerade i Finland deltog också pilotprojektet¹⁰.
5. FIN-CLARIN har skapat nya multifunktionella språkresurser så som den syntaktiskt analyserade finska textsamlingen FinnTreeBank (Voutilainen & al., 2012) och den finska synonymordboken FinnWordNet (Lindén & Carlson, 2009; Lindén & Niemi 2013) samt skapat redskap med HFST - Helsinki Finite-State Technology (Lindén & al., 2009, 2011, 2013) för att bearbeta språkdata, vidareutveckla existerande och integrera nya språkmaterial i Språkbanken i Finland.
6. För att utvidga språkmaterialiet i Språkbanken i Finland har FIN-CLARIN ingått ett licensavtal med Kopiosto beträffande digitaliserade material i Finlands nationalbibliotek. Materialet omfattar nu cirka 5 miljarder ord finskspråkiga tidskrifter från åren 1820-1940 och cirka 3 miljarder ord svenskspråkiga tidskrifter från åren 1750-1940.
7. FIN-CLARINs personal vid Språkbanken i Finland inom Helsingfors universitet och CSC har rest runt till språkresurskonsortiet FIN-CLARINs medlemsorganisationer och gett råd om hur man kan använda språkredskap och standarder i forskningsprojekt för att skapa resurser som kan återanvändas inom CLARIN.
8. Språkbanken i Finland har integrerats med Scientist's User Interface vid CSC¹¹ för att skapa en plattform för forskare att bearbeta och distribuera språkdata.
9. FIN-CLARIN ordnar möten med sin styrgrupp ungefär var tredje må-

7 www.clarin.eu/vlo/ och metashare.csc.fi

8 korp.csc.fi

9 <http://lat.csc.fi>

10 <https://weblicht.sfs.uni-tuebingen.de/>

11 <https://sui.csc.fi>

nad för att informera om den senaste utvecklingen och samla feedback och utvecklingsförslag. Dessutom har FIN-CLARIN årligen organiserat kurser vid CSC med avsikt att träna nya forskare i att använda forskningsinfrastrukturen och få konkreta exempel på hur infrastrukturen används i praktiken.

Målsättningen för CLARIN och FIN-CLARIN

CLARIN är en virtuell infrastruktur som erbjuder en ny plattform för humanistisk forskning, där en forskare kan arbeta vid sin egen arbetsstation, hitta språkmaterial i enorma datasamlingar, friktionsfritt få de nödvändiga tillstånden och påbörja sin forskning utan fördröjning.

CLARIN grundar sig på nationella initiativ som producerar språkmaterial för lagring i pålitliga CLARIN-center. Språkresurserna består av språkmaterial såsom maskinellt läsbara texter, lexikon, terminologier, digitaliserade inspelningar av talspråk och redskap för att bearbeta, söka och förädla sådant språkmaterial. Målet med CLARIN är att lösa tre problem som för tillfället hindrar forskare att till fullo utnyttja språkmaterial:

1. Digitalt språkmaterial existerar vanligen, men användarna kan inte hitta det. Detta kan lösas genom att man erbjuder metadata för språkresurserna som kan samlas in i en gemensam databas. Federerad sökning, dvs. en innehållssökning i samlingarna på alla datacenter inom CLARIN, förbättrar ytterligare möjligheterna att hitta lämpliga språkresurser.
2. När relevant material hittas, är det inte alltid lätt att veta hur man ska få tillstånd att använda det. Detta kan lösas genom att kategorisera och standardisera användarlicenserna och erbjuda potentiella användare en standardiserad autentiseringsmetod och rättighetsinnehavarna en gemensam auktoriseringsmetod.
3. Med behörigt användartillstånd är alla delar av materialet ändå inte alltid kompatibla med varandra eller med de redskap som finns. Detta kan lösas med standardiserade dataformat, gemensamma applikationsgränssnitt och harmoniserad terminologi.

Med nya redskap kan många problem, som tidigare tog flera veckor att lösa, klaras upp på mindre än en timme eller rentav på några minuter. Många påståenden som tidigare baserade sig på intuition kan baseras

på mer objektiva och välunderbyggda fakta. Nya regelbundenheter och undantag är lättare att upptäcka än med traditionella redskap.

Forskningen inom de humanistiska vetenskaperna kommer också att bli lättare att bekräfta när sådana språkmaterial som argumenten baserar sig på är tillgängliga för andra forskare och påståendena kan verifieras eller falsifieras.

FIN-CLARIN kommer i praktiken genom Språkbanken i Finland och Vetenskapstermbanken i Finland att betjäna hundratals forskare och tusentals studerande i Finland och många fler i resten av EU och världen. Användningen av standardiserad terminologi via Vetenskapstermbanken i Finland kommer att förbättra den vetenskapliga och tvärvetenskapliga diskussionen och rapporteringen om forskningsresultat både på nationell och på internationell nivå. Vetenskapliga upptäckter innebär också terminologiska innovationer som lätt kan förmedlas via termbanken.

Nya typer av forskning kommer att bli möjliga, då många av de nuvarande metoderna för språkbehandling kräver flera miljarder ord för att sammanställa pålitlig statistik för många av sina analysmetoder.

Tillgängligheten på tillräckliga språkresurser är nödvändig för att bygga adekvat språkteknologi för ett språk. Som påvisades av en internationell förfrågan inom META-NET¹², finns det alarmerande små språkresurser¹³ för finska jämfört med många grannländer för att inte tala om större länder som Tyskland, Frankrike och Storbritannien. Målet med FIN-CLARIN är att underlätta situationen. (För ytterligare information jämför t.ex. Finnish Language White Paper, Koskeniemi & al., 2012).

Implementering av den internationella forskningsinfrastrukturen

FIN-CLARIN är en del av den europeiska CLARIN-infrastrukturen som bygger ett europeiskt nätverk av center som erbjuder språkmaterial för humanisterna och språkforskarna i Europa. CLARIN var med på den första ESFRI vägkartan med 34 infrastrukturer som skulle byggas. Ett pilotprojekt under 2008–2011 förberedde en mera permanent CLARIN ERIC för att konstruera och driva nätverket av CLARIN center. CLARIN ERIC grundades den 29 februari 2012 med den Europeiska kommissionen beslut 2012/136/EU. För tillfället är Bulgarien, Danmark, Estland, Holland,

¹² <http://www.meta-net.eu/>

¹³ <http://www.meta-net.eu/whitepapers/key-results-and-cross-language-comparison>

Polen, Tjeckien, Tyskland och Österrike samt den holländska språkunionen medlemmar och Norge är observatör. Finland avser att gå med under 2014. Detta kräver en initialsatsning på 5 år för att utveckla och upprätthålla CLARIN ERIC och FIN-CLARIN som dess nationella nod.

Dr Krister Lindén är forskningsdirektör vid FIN-CLARIN och nationell koordinator för CLARIN i Finland.

Litteratur

- Koskenniemi, Kimmo, Lindén, K., Carlson, L., Vainio, M., Arppe, A., Lennes, M., Westerlund, H., Hyvärinen, M., Nuolijärvi, P., Piehl, A., 2012: Suomen kieli digitaalisella aikakaudella [Finnish in the Digital Age], Berlin. 85 p. <http://www.meta-net.eu/whitepapers/volumes/finnish>
- Lindén, Krister, Carlson, L., 2010: FinnWordNet - WordNet på finska via översättning. [FinnWordNet – WordNet in Finnish by Translation] / In: *LexicoNordica*, Vol. 17, 11.
- Lindén, Krister and Jyrki Niemi, 2013: Is It Possible to Create a Very Large WordNet in 100 days? – an Evaluation / In: *Language Resources and Evaluation*, Springer Verlag.
- Lindén, Krister, Silfverberg, M., Axelson, E., Hardwick, S., Pirinen, T., 2011: HFST–Framework for Compiling and Applying Morphologies. / In *Systems and Frameworks for Computational Morphology*. Edited by Cerstin Mahlow and Michael Pietrowski. Vol. 100 Springer. p. 67-85 (Communications in Computer and Information Science).
- Lindén, Krister, Silfverberg, M., Pirinen, T., 2009: HFST Tools for Morphology – An Efficient Open-Source Package for Construction of Morphological Analyzers. / In *State of the Art in Computational Morphology* edited by Cerstin Mahlow, Michael Pietrowski. Berlin, Heidelberg, Springer Berlin Heidelberg. p. 28-47 (Communications in computer and information science).
- Oksanen, Ville, Lindén, K., Westerlund, H., 2010: Laundry Symbols and License Management – Practical Considerations for the Distribution of LRs based on experiences from CLARIN. In: *Proceedings of LREC 2010 : Workshop on Language Resources: From Storyboard to Sustainability and LR Lifecycle Management*.
- Voutilainen, Atro, Muhonen, K., Purtonen, T.K., Lindén, K., 2012: Specifying Treebanks, Outsourcing Parsebanks: FinnTreeBank 3. In: *The*

eighth international conference on Language Resources and Evaluation (LREC).

Lindén, Krister, Axelson, E., Drobac, S., Hardwick, S., Kuokkala, J., Niemi, J., Pirinen, T., Silfverberg, M., 2013: HFST—a System for Creating NLP Tools. In: *Systems and Frameworks for Computational Morphology Systems and Frameworks for Computational Morphology: Communications in Computer and Information Science*, edited by Cerstin Mahlow and Michael Piotrowski, Springer-Verlag, 2013. (Communications in Computer and Information Science).

Summary

The virtual distributed research infrastructure FIN-CLARIN is the Finnish national node of CLARIN ERIC. FIN-CLARIN provides two centralized services: the Language Bank of Finland and the Bank of Finnish Terminology in Arts and Sciences. The Language Bank of Finland makes available collections of digital language resources and tools for analyzing them via its on-line service center. It serves a wide research community of humanists as well as social scientists and computer scientists. Its relevance lies in the amount and diversity of materials as well as in the seamless access provided to researchers. The goal is to make available collections from different periods, genres and regions as well as different modalities such as text, audio, pictures and video containing language data. The Bank of Finnish Terminology in Arts and Sciences is a multidisciplinary and multilingual project aiming at developing a permanent and easily updated terminological database for all fields of research in Finland.

Sprogteknologi og sproginstitutioner. Hvilken rolle kan sprognævnene spille i forhold til sprogteknologi?

Sabine Kirchmeier-Andersen

De nordiske sprognævn har igennem flere år arbejdet ihærdigt for at fremme kendskabet til sprogteknologi og for at bringe relevante aktører på området sammen for at styrke det nordiske samarbejde. Sprognævnenes arbejdsgruppe for sprogteknologi i Norden (ASTIN) har bl.a. afholdt flere succesfulde konferencer og workshops med sprogteknologi på programmet. I det følgende vil det blive belyst hvilke resultater dette arbejde har skabt, og hvordan samarbejdet om sprogteknologi kan videreudvikles for at styrke brugen af de nordiske sprog i fremtidens informationsteknologi.

Sprogteknologi, den nordiske sprogdeklaration og nordisk sprogpolitik

Sprogteknologi nævnes ikke direkte i den nordiske sprogdeklaration (Deklaration om nordisk sprogpolitik 2006), men der fremhæves under arbejdsspørgsmål 1 tre applikationer som forudsætter sprogteknologi:

- Internordiske ordbøger i papirform og elektronisk form udarbejdes
- Maskinoversættelsesprogrammer for Nordens samfunds bærende sprog og programmer til søgning i nordiske databaser udvikles (Deklaration om nordisk sprogpolitik 2006, s. 13)

Sprogteknologi er imidlertid ikke begrænset til dette, men er et middel der gør det muligt hurtigere og bedre at opnå flere af de andre målsætninger der opstilles i sprogdeklarationen, og flere de arbejdsspørgsmål som prioriteres, især dem som vedrører parallelsproglighed og mangesprogethed/flersproglighed. Udviklingen af sprogteknologi er også en forudsætning for at man kan nå målet om at gøre Norden til en sproglig foregangsregion.

Sprogteknologiske programmer indgår i undervisningsprogrammer på alle niveauer, i interaktive netsteder, ordbøger, talegrænseflader, stavetkontroller osv., og sprogteknologi har længe været blandt de faktorer der kan medvirke til at bevare truede sprog (Crystal 2000, s. 143).

Der er gået 8 år siden sprogdeklarationen blev vedtaget, og netop på det sprogteknologiske område er der sket meget i denne tid. Hvis sprogdeklarationen skulle revideres i de kommende år, ville det være nærliggende at inkludere mere sprogteknologi, heriblandt fx taleteknologi, især talegenkendelse, blandt de applikationer som skulle fremmes for at styrke de nordiske sprog. De mål for sprogteknologi som er opstillet i deklARATIONEN, er kun delvist nået. Således er der kommet enkelte elektroniske ordbøger, fx Islex (<http://islex.lexis.hi.is>), og der er også udviklet programmer til tværspørglig søgning, som dog kun kører som prototype på Nordisk Ministerråds hjemmeside (http://cst.ku.dk/projekter/projekter_slut/netordbog/). Inden for området maskinoversættelse er der stadig lang vej endnu.

Sprogteknologi spiller også en rolle i de nordiske landes sprogpolitikker. Alle sprogpoltiske redegørelser, hvad enten de er blevet efterfulgt af egentlig sproglovgivning eller ej, indeholder overvejelser om sprogteknologiens betydning. Der er endvidere en generel anerkendelse af at de nordiske sprog som marked betragtet ikke er store nok til at internationale virksomheder som indarbejder sprogteknologi i deres produkter, fx Microsoft (stavetkontrol) eller Apple (talegenkendelse), kan motiveres til at investere nævneværdige beløb i produkter af høj kvalitet for disse sprog.

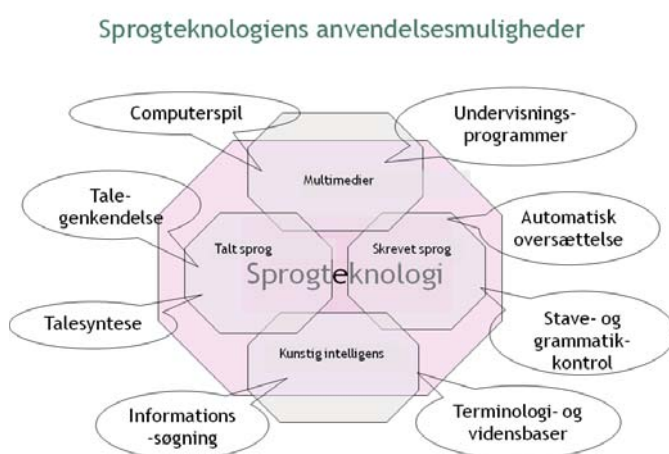
Anbefalingerne i de sprogpoltiske redegørelser går typisk på at opbygge sprogteknologiske infrastrukturer og satse på forskning i sprogteknologiske produkter der tager højde for de specielle karakteristika man finder i de nordiske sprog. Der er med andre ord en klar erkendelse af at sprogteknologi for de enkelte sprog bør styrkes, men der er stor forskel på hvor meget de enkelte lande vælger at investere, fx har Norge og Sverige for flere år siden etableret nationale termbanker, og der arbejdes nu på at oprette permanente sprogbanker til organisering og tilgængeliggørelse af sprogresurser. Språkrådet i Norge har presset på over for de politiske beslutningstagere for at få gennemført planerne om sprogbanker. Etableringen af en permanent sprogbank blev besluttet i 2009 og påbegyndt i 2010. I Sverige er arbejdet ikke så langt fremskredet, men der arbejdes nu på et pilotprojekt om en sprogbank med talesprogsdata. Til

trods for at termbanker og sprogbanker har været en del af de sprogpoltiske anbefalinger i Danmark i de sidste 10 år og forslag om termbanker har været på forskningsministeriets kortlægning af forskningsinfrastruktur og flere gange har været behandlet i Folketinget, er der endnu ikke er blevet givet permanente bevillinger til nogen af delene.

Hvad er sprogteknologi, og hvordan virker det?

Sprogteknologi er en fællesbetegnelse for teknikker der anvendes til at få computere til at bearbejde tekst eller talt sprog. Fra simple programmer til optælling af ord i løbende tekst i 1960'erne har sprogteknologien i takt med udviklingen i informationsteknologien udviklet sig til at omfatte mange forskellige aspekter af sproget og dækker i dag mange forskellige anvendelsesfelter. Sprogteknologi overlapper delvist med andre it-områder som multimedieteknologi og kunstig intelligens. Der er meget forskellige teknikker knyttet til at håndtere input på henholdsvis talt og skrevet sprog, mens de teknikker som bruges til at analysere indholdet er fælles.

Sprogteknologi indgår efterhånden i alle former for informations- og kommunikationsteknologi lige fra undervisningsprogrammer og computerspil til informationssøgning og automatisk oversættelse. Også moderne ordbøger og terminologi- og vidensbaser indeholder sprogteknologiske komponenter. Automatisk oversættelse og tolkning er blandt de største udfordringer for sprogteknologien idet oversættelse kræver en meget avanceret forståelse af både kildesproget og målsproget og af relationen imellem dem.



Figur 1

Mange computerprogrammer, databaser og webapplikationer som involverer sprog, bliver udviklet uden en grundlæggende viden om sprogteknologi, og dette medfører ofte uhensigtsmæssigheder som kunne være undgået hvis producenterne havde involveret mennesker med sprogteknologisk kompetence.

Stort set alle applikationer der involverer sprogteknologi, er skabt ved hjælp af to grundlæggende byggesten: En teknikresurse, typisk et sprogteknologisk program eller softwaremodul, og en sprogresurse, fx store samlinger af talt eller skrevet sprog – det sproglige råstof.



Figur 2

Disse byggesten bygges sammen til applikationer med to grundlæggende metoder, regelbaserede og statistiske. Når der bruges regelbaserede metoder, udarbejder sprogteknologer datamatiske ordbøger, grammatikker og oversættelsesregler eller regler for hvordan tale omsættes til skrift, ud fra deres viden om sprog og ud fra de observationer de kan gøre med udgangspunkt i sprogresurserne. Disse regler bygges så ind i programmer der fx kan analysere og omforme et spørgsmål fra en kunde til en forespørgsel i en database eller generere et svar på et spørgsmål man har stillet til en chatrobot, en service på nettet hvor man kan få besvaret sine spørgsmål af computeren, fx IKEA's chatrobot Anna.

Når der bruges statistiske metoder, udleder et program reglerne automatisk fra de digitale sprogresurser, fx kan sandsynligheden for at et givet ord på et sprog skal oversættes med et givet ord på et andet sprog, beregnes ud fra de statistiske informationer der kan udledes af en sprogresurse bestående af en samling tekster på et sprog og deres oversættelse til et andet sprog. De beregnede sandsynligheder og regler indgår derefter i applikationer som fx Google Translate.

Begge metoder har fordele og ulemper. Regelbaserede metoder er langsomme at udvikle fordi de er mere manuelle og kræver mange persons arbejde. Til gengæld er de ofte mere generelt anvendelige. Statistiske metoder er markant hurtigere at udvikle, men kun hvis der allerede foreligger digitale sprogresurser i rigelige mængder, og det er særdeles vanskeligt for sprog som færøsk, grønlandsk og samisk (Langgård 2014, Moshagen 2014). Endvidere er der den ulempe ved statistiske metoder at de kun kan analysere og producere den type sprog som anvendes i sprogresurserne. Træner man et maskinoversættelsessystem på fx debatter fra EU-parlamentet, vil de virke fint for politiske debattekster, men komme til kort når man skal oversætte en brugsanvisning til en husholdningsmaskine.

I de sidste 20 år har systemer som er baseret på statistiske metoder, domineret markedet. Mange store applikationer er blevet udviklet, fx systemer som Google Translate, automatisk talegenkendelse og diktering i bl.a. sygehussektoren, fx Nuance/MacManus, og avanceret informationsøgning, fx IBM's søgemaskine Watson (www-03.ibm.com/innovation/us/watson) der har slået alle rekorder i den amerikanske paratvidenquiz Jeopardy og nu bl.a. bliver anvendt i medicinsk forskning og i finanssektoren.

Man er dog ved at nå til et punkt hvor det er vanskeligt at opnå nævneværdige forbedringer af kvaliteten af de statistiske metoder, og et mere avanceret samspil mellem regelbaserede og statistiske metoder synes at være vejen frem. Tendensen går i retning af at kombinere de to metoder således at de regler som udvikles statistisk, finjusteres ud fra en regelbaseret tilgang, eller at man træner systemerne på sprogresurser hvor der allerede er foretaget en regelbaseret analyse, fx tekster som er manuelt opmærket med syntaktiske funktioner eller dependensoplysninger som en slags guldstandard.

Sprognævnenes engagement i sprogteknologi

Sprognævnene er både brugere og producenter af sprogteknologi og sprogteknologiske resurser. Det giver god mening at sprognævnene følger med i den sprogteknologiske udvikling. For det første fordi det ikke kan udelukkes at teknologien kan påvirke sprogbrugen, og for det andet fordi der bliver udviklet viden og værktøjer der støtter sprognævnene i nogle af deres kerneopgaver. De ordbøger, tekstsamlinger og svarsamlinger som sprognævnene selv producerer eller er med til at producere,

udgør en vigtig resurse for andre producenter af sprogteknologi, og det vil være fremmende for udviklingen hvis sprognævnene kan stille disse resurser til rådighed for forskere og udviklere. Sidst men ikke mindst ser sprognævnene det som en vigtig opgave at styrke udviklingen af nordisk sprogteknologi som en del af opfølgningen på den nordiske sprogdeklaration.

Netværket for sprognævnene i Norden (NSN) har nedsat en række arbejdsgrupper som har til opgave at følge op på centrale mål i sprogdeklarationen, heriblandt i 2005 Arbejdsgruppen for sprogrøgt og sprogteknologi i Norden (ASTIN).

ASTIN har følgende opgaver:

- at arbejde for at sprogteknologisk forskning og udvikling får adgang til den nødvendige infrastruktur i form af grundlæggende sprog- og teknikressurser for de nordiske sprog
- at arbejde for at vigtige sprogteknologiske applikationer udvikles for de nordiske sprog, og at disse tilpasses grupper med særlige behov også når den kommercielle interesse savnes
- at gøre opmærksom på teknologiens konsekvenser for skriftsprogsnormen og sprogansværelsen i Norden og at arbejde for at nordisk sprogteknologi har en god sproglig kvalitet
- at styrke det nordiske sprogrøgtarbejde ved hjælp af sprogteknologi
- at fremme samarbejdet mellem sprognævn, forskning og industri i Norden på det sprogteknologiske område

ASTIN består p.t. af

- Torbjørg Breivik, Språkrådet i Norge
- Rickard Domeij, Språkrådet i Sverige
- Per Langgård, Grønlandsk Sprognævn

- Sjur Nørstebø Moshagen, Sametinget
- Sabine Kirchmeier-Andersen, Dansk Sprognævn

ASTIN har siden 2005 først og fremmest arrangeret konferencer og seminarer for sprognævn, forskere og virksomheder:

2005	Språkverktøy for språkene i Norden, 21.-22. april, Pargas, Finland, ca. 50 deltagere
2006	Språkteknologisk infrastruktur for språkene i Norden, 26. oktober, Göteborg, ca. 60 deltagere (samarbejde med SLTC (Swedish Language Technology Conference))
2007	Sprogrøgt, sprogpolitik og sprogteknologi, 29.-30. oktober, København, ca. 60 deltagere
2008	Oversettingsteknologi i Norden, 23.-24. oktober, Oslo, ca. 45 deltagere
2009	Nordiske perspektiver på CLARIN's infrastruktur for sprogresurser, 14. mai, Odense, ca. 30 deltagere (arbejdsseminar på de Nordiske Datalingvistikdage (NoDaLiDA 2009))
2009	Sprogteknologisk terminologi, 12. juni, København, ca. 30 deltagere (arbejdsseminar i tilknytning til Nordterm 2009)
2010	Språkteknologi for økt tilgjengelighet, Linköping, 27.-28. oktober, 45 deltakere (i samarbejde med SLTC 2010)
2011	Visability and availability of LT resources, 10. mai, Riga, ca. 35 deltakere (arbejdsseminar på de Nordiske Datalingvistikdage (NoDaLiDa 2011))
2012	Taleteknologi - hva har vi, og hva vil vi? Lund, 24.-25. oktober, 55 deltakere (konference i tilknytning til SLTC 2012)
2013	Bidraget til programmet for det nordiske sprog møde som havde sprogteknologi som tema.

Tabel 1

Derudover har ASTIN fungeret som styre- eller rådgivningsgruppe for forskellige projekter, fx sprognævnens svarbaseprojekt (www.nord-svar.org) og projekter om kortlægning af nordiske sprogresurser, fx den nordiske vismandsrapport om sprogteknologi (Koskenniemi, Lindén & Nordgård 2007). ASTIN's medlemmer er typisk også involveret i natio-

nale sprogteknologiske projekter, typisk infrastrukturprojekter som fx sprogbanker og i de nordiske afdelinger af EU-projekterne CLARIN og META-NET.

Sprognævnenes sprog- og teknikressurser

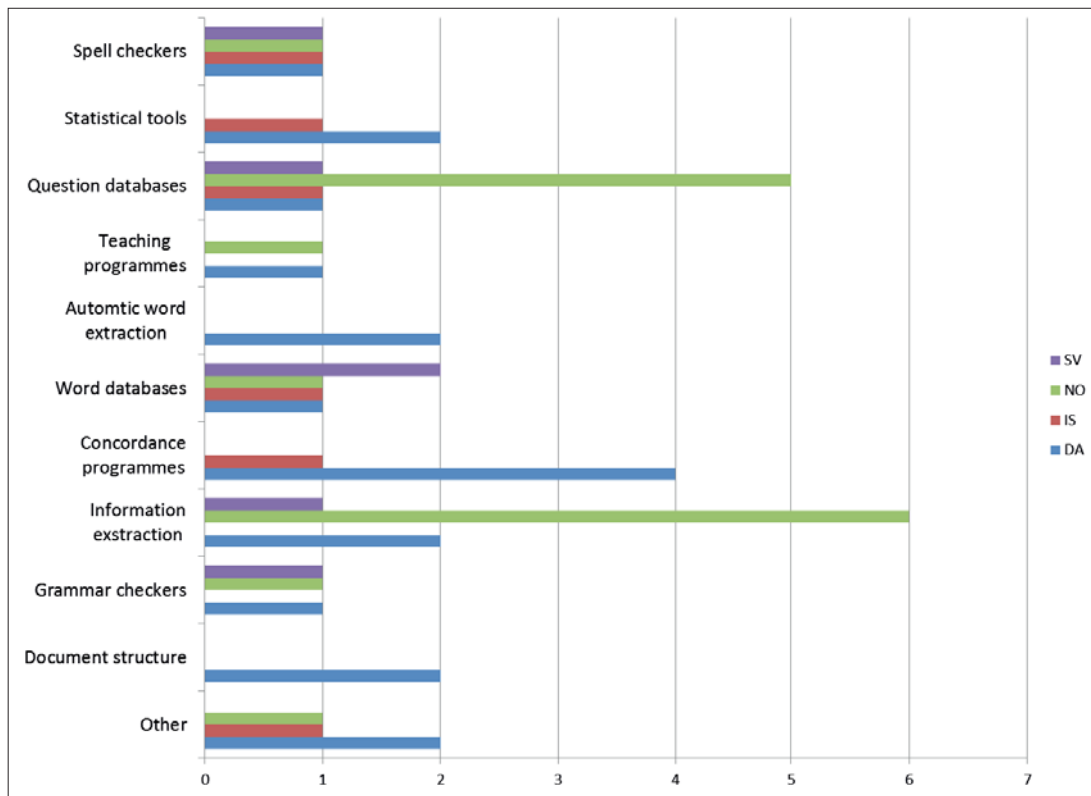
I 2007 foretog ASTIN en undersøgelse af de resurser som de nordiske sprognævn bruger eller selv opbygger, for at få et bedre billede af på hvilke områder der kunne være mulighed for et tættere samspil med forskningsmiljøer og sprogteknologiske virksomheder (Kirchmeier-Andersen 2011).

Undersøgelsen tog udgangspunkt i såkaldte BLARK-koncept (Basic Language Resource Kit) (Krauwer 1998 og www.blark.org) som opregner de sprogteknologiske resurser der er nødvendige for at sprogteknologi kan udvikles for et givet sprog.

Blandt de teknikressurser som blev kortlagt, var fx parsere (dvs. automatisk morfologisk analyse og sætningsanalyse), programmer til informationsekstraktion, værktøjer til automatisk opmærkning, konkordansprogrammer eller transskription af talt sprog, statistiske værktøjer, stavetjekkontroller, grammatikkontroller, programmer til automatisk ekstraktion af termer og nye ord, orddatabaseprogrammer, programmer til udvikling af sprogøvelser og -tests mv.

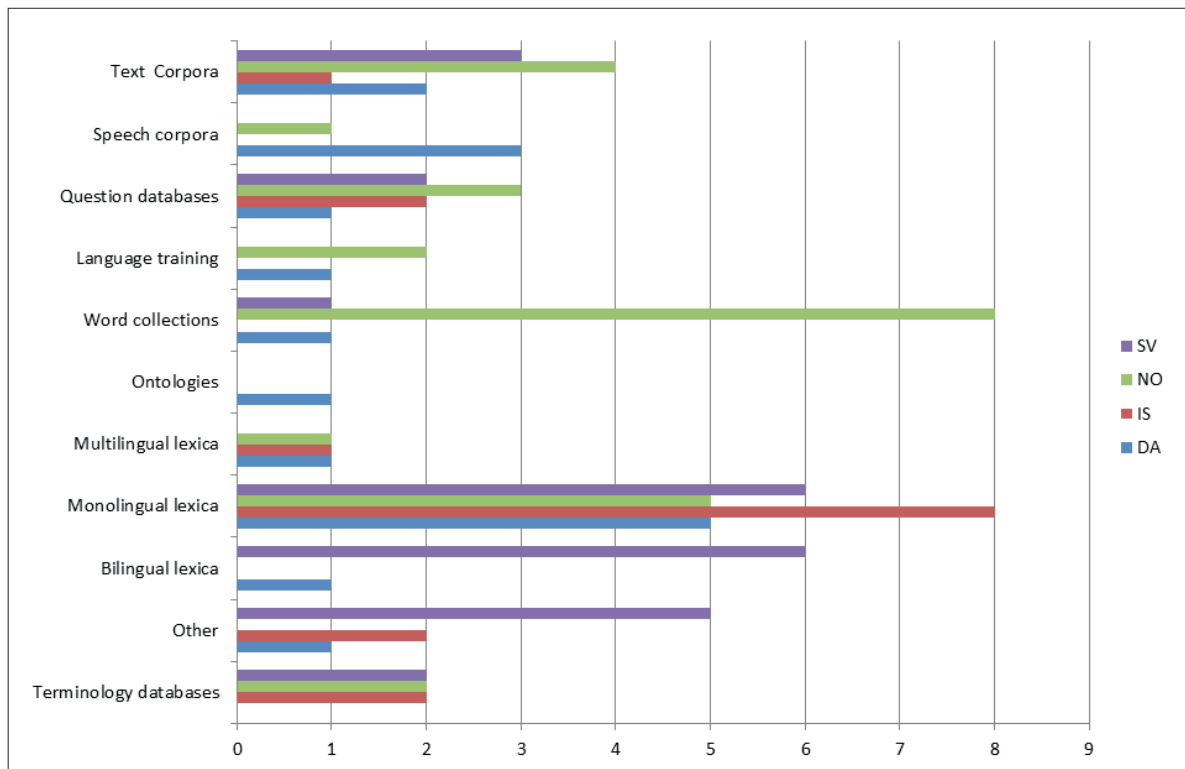
Blandt de sprogresurser der blev kortlagt var fx digitale ordbøger (mono-, bi- og multilingviale), korpuser (tale, tekst, flersproglige, og parallelle), ordnet og tesaurusser, ontologier, svarsamlinger, ordsamlinger, sproglige øvelser, terminologisamlinger. I undersøgelsen deltog Sverige, Norge, Island og Danmark.

Hvad angår teknikressurser, altså sprogteknologisk software, viste det sig at kun stavetjekkontroller, svardatabaser og orddatabaser var i brug i alle fire sprognævn, og at der typisk var forskellige typer af svardatabaser, konkordansprogrammer og programmer til informationsekstraktion til rådighed. Programmer til analyse og generering af dokumentstruktur og automatisk ekstraktion af nye ord var på dette tidspunkt kun i brug i Danmark.



Oversigt over teknikressurser i de nordiske lande

Hvad angår sprogresurser, viste det sig at tekstkorpuser, ordsamlinger og ensprogede elektroniske ordbøger fandtes i alle fire sprognævn, mens fx sprogundervisningsmateriale kun fandtes i Danmark og Norge, mens databaser med terminologiske data kun fandtes i Sverige, Norge og Island.



Overview of language resources in the Nordic countries

Dataene blev præsenteret på et seminar i København i 2007 og dannede efterfølgende grundlag for dialog med virksomheder og forskningsmiljøer. Det blev bl.a. tydeligt at sprognævnene ikke selv producerer sprogteknologi i nævneværdigt omfang, men at de gerne indgår et tæt samarbejde med udviklervirksomheder og forskningsinstitutioner om udvikling og afprøvning af nye programmer. Endvidere blev sprognævnene opmærksomme på at de resurser de selv udvikler, kan have stor værdi for forskere og udviklere. En øget resursedeling, fx efter et open source-princip, stiller imidlertid både krav til klarlægning af de ophavsmæssige rettigheder til resurserne og til sprognævnernes strukturering af de sproglige data.

Sprogteknologi i de nordiske lande i dag

Selv om sprognævnene i Norden ikke dækker alle former for sprogteknologi i Norden, udgør de i kraft af deres opgaveportefølje og sproglige opgaver en meget avanceret brugergruppe, og det er tydeligt at ønsket om at kunne bruge sprogteknologi generelt i de nordiske samfund og i sprognævnene i særdeleshed er steget betydeligt siden 2007. Endvidere

er nye kommunikationsformer kommet til, bl.a. var udviklingen i de sociale medier som Facebook og Twitter og i app-markedet til telefon og tabletcomputere kun lige begyndt da sprognævnens kortlægning blev iværksat.

I takt med vedtagelser af sproglove i Sverige, Island og Grønland, og nye sprogpoltiske tiltag i de andre lande er der i varierende grad blevet investeret offentlige midler i udvikling af sprogteknologi og sproglige ressourcer. For tiden ser vi de største offentlige investeringer i termbanker og sprogbanker for norsk, svensk og finsk. Endvidere er det lykkedes at udvikle avanceret sprogteknologi for grønlandsk og samisk omfattende bl.a. stavekontroller, talesyntese og maskinoversættelse, og også i Island er der på det seneste opbygget et sprogteknologisk basisberedskab i form af sprogteknologiske korpuser, ordbøger og værktøjer til håndtering af islandsk. Danmark har udviklet værktøjer og indsamlet i sproglige ressourcer i DK-Clarín fra 2008-2011 (Halskov & Asmussen 2009), og der bliver løbende udviklet sprogteknologi i forbindelse med nye forskningsprojekter, fx inden for taleteknologi og semantik, fx et dansk Wordnet (Nimb 2006 og www.wordnet.dk), men der er ikke foretaget større nationale satsninger siden, og der foregår kun sporadisk koordinering af de eksisterende projekter og ressourcer.

Norden som sprogteknologisk underudviklet region?

Samtidig med at udviklingen for de større sprog accelererer stadig kraftigere, ser vi at mange basale problemer for de nordiske sprog stadig ikke er løst. Hvor man i de engelsktalende lande allerede er langt fremme med stemmestyrede applikationer, avanceret maskinoversættelse og semantisk fortolkning af tekst, er helt basale problemer, stadig ikke løst.

Der findes fortsat programmer hvor η , δ , \ae , \o , \ddot{o} , \aa , fremstår som andre tegn eller omskrives til fx ae, oe, aa, og der er løbende problemer med at genfinde information fordi systemernes interne søgeprogrammer ikke er ordentligt gearet til de nordiske sprog. Selv om man har fundet en teknisk løsning på brugen af de særlige nordiske bogstaver i URL'er, er denne løsning ikke særlig udbredt, og da den ikke er blevet fulgt op af en tilsvarende løsning for e-mailadresser, ser man løbende især virksomheder og offentlige institutioner ændre stavningen af deres officielle navne til noget der er kompatibelt med det engelske alfabet. I Danmark skiftede byen Århus i 2012 således officielt til stavemåden Aarhus.

Selv simple applikationer som stavekontroller bliver ikke tilpasset de

nordiske sprog. Microsofts stavekontrol i Office 2010 foreslår fx som default at ord der ikke kan genkendes, deles i to, fx *garderobenicherne* foreslås skrevet *garderobe nicherne*, og for et simpelt ord som *dørhul* foreslås *dør hul*. Der findes endnu ingen undersøgelser af hvilke konsekvenser dette har for stavefærdigheden hos nordiske borgere, især børn og unge, men de er formentlig i det lange løb bedre tjent med ikke at få et forslag end at få et der er så misvisende.

Ser vi på kvaliteten af mere avanceret sprogteknologi, som fx maskinoversættelse, lader kvaliteten i fx Google Translate af de sprogpar som involverer nordiske sprog, fortsat meget tilbage at ønske, hvilken kan skyldes at der ikke er tilstrækkeligt mange parallelle tekster tilgængelige som involverer de nordiske sprog, sammenlignet med hvad der er tilgængeligt fx for de officielle UN-sprog som har dannet udgangspunktet for udviklingen af programmet (Och 2003).

Eksemplerne viser at der fra kommerciel side ikke satses på nordiske sprog, formentlig fordi markedet er for lille. Mulighederne for at sprogteknologien kan bidrage substantielt til at gøre Norden til en sproglig foregangsregion, er derfor kun til stede hvis forskning og udvikling inden for sprogteknologi får højere prioritet, og det tværnordiske samarbejde på området intensiveres.

Bedre sprogteknologi gennem samarbejde

Når det for de store softwareproducenter tydeligvis ikke er en god forretning at prioritere de nordiske sprog, er det helt afgørende at regeringerne i de enkelte lande er opmærksomme på disse problemer, og at de træder til med midler til at holde den sprogteknologiske udvikling i Norden på niveau med andre lande.

Samarbejdet kan med fordel foregå på flere niveauer, dels mellem sprognævn, forskere og virksomheder i de enkelte lande, dels mellem disse forskellige aktører på tværs af de nordiske lande. Endvidere bør repræsentanter for både nationale og nordiske offentlige institutioner deltage, da det i stigende grad også er den offentlige sektor som har brug for sprogteknologi. Da udviklingen på dette område går meget hurtigt, kan alle drage nytte af videndeling på tværs af sektorer og lande.

Et eksempel på dette er de danske kommuner som satser intensivt på at indføre taleteknologi som led i effektiviseringen af den offentlige sektor. Odense Kommune alene investerede i 2012 mere end 50 millioner kr. på udvikling af statistisk baseret taleteknologi. Mere end 2000 medar-

bejdere skulle begynde at tale til deres computer i hverdagen i stedet for at skrive breve, rapporter og indberetninger. Projektet var tæt på at kuldsejle da det viste sig at kvaliteten var alt for ringe, bl.a. fordi leverandørerne havde alt for lidt kendskab til dansk, og systemerne ikke var trænet på kommunale tekster. Kommunen skulle derfor hurtigt fremskaffe store tekstmængder for at virksomheden kunne gentræne systemerne og forbedre kvaliteten.

Fremskaffelsen af det sproglige råmateriale til træning af systemerne er dyrt og tidkrævende, men i modsætning til andre råstoffer kan materialet bruges igen og igen uden at blive slidt eller opbrugt. Der er derfor god økonomi i at sørge for at så meget af materialet kan genbruges, og det forudsætter afklaring af ophavsrettigheder og ejerforhold. Stort set alle nordiske kommuner vil i løbet af de kommende år foretage lignende investeringer som Odense Kommune. Kommunernes Landsforening har gennem sit eget selskab KOMBIT i samarbejde med DANCAST-centret ved CBS i København iværksat initiativer til at finde metoder til at sikre udbud af sprogteknologiske opgaver til en rimelig pris hvor de særlige krav til det sproglige råmateriale tilgodeses, og hvor kommunerne sikres mulighed for og rettigheder til at genbruge og dele det sproglige råmateriale (se også Juel Henrichsen i dette bind).

Brugen af nye input-metoder som talegenkendelse åbner for mange nye muligheder især i den offentlige sektor. I takt med at vi i stigende grad går over til borgerbetjening på nettet, vil borgere med læse-/skrivehandicap få problemer. Muligheden for talt input kunne være en løsning for mange. En del kommunale medarbejdere i socialektoren har ligeledes svært ved at udtrykke sig korrekt på skrift i rapporter og indberetninger, og også her kan en talegrænseflade hjælpe. Opgaven med at sikre et korrekt og klart sprog lægges dermed i højere grad over på softwareprogrammerne – et område som de nordiske sprognævne derfor bør overvåge, således at de kan sikre at programmerne overholder de officielle standarder for skriftsproget.

Den digitale udvikling og den stigende mængde af skriftligt materiale på nettet og i diverse applikationer stiller imidlertid også store krav til sprognævnenes egen brug af sprogteknologi. Der er brug for software der kan hjælpe sprognævnene med deres løbende opgaver, fx

- Tekstanalyseværktøjer til analyse af aktuel sprogbrug (fx konkordansprogrammer, ordtrawlere m.m.)

- Orddatabaser til avanceret strukturering af ordbøger
- Programmer til udvikling af sproglige øvelser og tests
- Oversættelsesværktøjer mellem de nordiske sprog, især minoritetsprog, fx svensk-finsk, dansk-grønlandsk, norsk-samisk mv., og mellem de nordiske sprog og engelsk
- Værktøjer til kvalitetsanalyse af sprogteknologi, fx af stavekontrolprogrammer og af output fra talegenkendelse

Denne type software kan udvikles i samarbejde med virksomheder og forskningsinstitutioner, og også her vil der være meget at spare gennem videndeling og fælles nordisk programudvikling.

Sprognævnene kan som sagt også selv bidrage til at forbedre sprogteknologi ved at dele deres resurser med forskere og udviklere, fx

- Tekstsamlinger
- Talesprogsdata
- Ordbøger
- Grammatikker
- Andre typer af sproglig information, fx samlinger af nye ord og udtryk

Det ville være en stor fordel hvis dette skete systematisk og regelmæssigt, da man derved i højere grad kan sikre at de standarder som nævnene anbefaler, kommer ind i den software der udvikles.

Et hidtil overset forskningsområde er hvordan sprogteknologisk software kan holdes løbende ajour med den sproglige udvikling. Når et program trænes på en bestemt tekstsamling, vil det kun være i stand til at genkende og producere inden for de konstruktioner og det ordforråd som det er blevet trænet på, og hvis det ikke bliver opdateret, vil de fleste systemer blive forældet i løbet af 5-10 år. Sprognævnene har en særlig erfaring med at dokumentere forandringer i skriftsproget og bør også her

sørge for at dele deres viden og deres sprogresurser, fx nyordslister, mest muligt.

Til sidst kan sprognævnene bidrage til at øge kendskabet og udbredelsen af sprogteknologi og ikke mindst kendskabet til teknologiens begrænsninger og korrekte anvendelse.

Nævnene bør derfor fortsat

- opfordre beslutningstagere til at støtte udviklingen af sprogteknologi
- opfordre beslutningstagere til at gøre sprogresurser tilgængelige for sprogteknologi
- overbevise beslutningstagere om at det er vigtigt at der findes ekspertise inden for sprogteknologi for deres sprog

Det er helt afgørende at der på nationalt og på nordisk plan kontinuerligt iværksættes forskningsprogrammer og infrastrukturinitiativer der støtter udviklingen af sprogteknologi hvis de nordiske lande ikke skal sakke yderligere agterud på dette område.

Hvad angår sproglige rådata, bør man overveje at tænke en sprogteknologisk infrastruktur sammen med det nyligt reviderede PSI-direktiv fra EU (DIRECTIVE 2013/37/EU)¹ som fastlægger regler for genbrug af data i den offentlige sektor. Direktivet fastlægger at alle offentlige dokumenter skal kunne genbruges inden for direktivets rammer, at de skal gøres tilgængelige digitalt og forsynes med relevante metadata der følger åbne standarder, og at omkostningerne til dette skal holdes på minimumsniveau. Ved den seneste revidering i 2013 blev biblioteker, museer og arkiver ligeledes omfattet af direktivet, og det åbner yderligere mulighed for at inddrage relevante sproglige rådata.

Udvikling af sprogteknologisk software er ikke en opgave der kan overlades til almindelige it-ingeniører eller udenlandske softwarehuse. Det kræver både datalogisk indsigt og sproglig ekspertise – ikke blot en generel kompetence inden for kognition og lingvistik, men også en om-

1 <http://ec.europa.eu/digital-agenda/en/european-legislation-reuse-public-sector-information>

fattende indsigt i de sprog som softwaren skal håndtere, helst som modersmålstalende.

Nordiske sprogteknologer og datalingvister udspringer af højt specialiserede niceuddannelser som er meget sårbare over for de stigende effektiviseringskrav der stilles til universiteterne da der typisk kun uddannes en håndfuld i hvert land om året. I takt med at sprogteknologisk software bliver mere og mere udbredt, vil behovet for veluddannede sprogteknologer begynde at stige. Det er vigtigt i alle nordiske lande at man sørger for at der er stærke faglige miljøer på universiteterne der kan tiltrække nye sprogteknologer og uddanne dem på alle niveauer til alle nordiske sprog.

Konklusion

Sprogteknologi er vigtig for at sikre de nordiske sprogs status og brug i alle sammenhænge. Det påpeges allerede i den nordiske sprogdeklaration i 2006, men i lyset af den rivende udvikling på området burde anbefalingerne i sprogdeklarationen vedr. sprogteknologi revideres. De nordiske sprognævn har en forpligtelse til at sikre at sprogteknologi af høj kvalitet bliver udviklet for deres respektive sprog. Samarbejdet mellem forskere og udviklere inden for sprogteknologi og de nordiske sprognævn bør yderligere styrkes for at sikre løbende vidensudveksling. Beslutningstagere bør især have fokus på at etablere en permanent sprogteknologisk infrastruktur der sikrer indsamling og genbrug af sproglige rådata i form af sprogbanker, fx i tilknytning til implementeringen af EU's PSI-direktiv.

Sabine Kirchmeier-Andersen er sprogteknolog og direktør i Dansk Sprognævn.

Litteratur

Kirchmeier-Andersen, Sabine, 2011: Language Technology for Language Institutions. I: Stickel, Gerhard & Tamas Varadi (eds.): *Language, Languages and New Technologies. ICT in the Service of Languages*. Duisburg Papers on Research in Language and Culture. Vol. 87. Peter Lang, s. 21-32.

- Crystal, David, 2000: *Language Death*. Cambridge. Cambridge University Press.
- Koskenniemi, Kimmo, Lindén, K., Nordgård, T., 2007: *Språkvis – Språkteknologisk vismansrapport. En utvidgad sammanfattning av The Nordic Countries – A leading region in Language Technology*. Helsingfors. EU (DIRECTIVE 2013/37/EU)
- Deklaration om nordisk sprogpolitik 2006. Nordisk Ministerråd.
- Krauwer, Steven, 1998: *ELNET and ELRA: A common past and a common future*. In: ELRA Newsletter Vol. 3 N. 2.
- Nimb, Sanni & Hartvig Sørensen, Nicolai, 2006: DANNET – et leksikalsk semantisk wordnet for dansk. I: Peter Widell og Ulf Dalvad Berthelsen (udg.): *11. Møde om Udforskningen af Dansk Sprog*. Århus.
- Juel Henriksen, Peter, 2014: Tale er Guld. Om talegenkendelse i de danske kommuners tjeneste. I: *Språk i Norden 2014*.
- Langgård, Per, 2014: Selvfølgelig snakker vaskemaskiner grønlandsk i fremtiden. I: *Språk i Norden 2014*.
- Moshagen, Sjur, 2014: Status for samisk språkteknologi. I: *Språk i Norden 2014*.
- Och, Franz Josef, 2005: Statistical Machine Translation: Foundations and Recent Advances. In: *Proceedings of the Tenth Machine Translation Summit*, Phuket, Thailand 2006.
- Halskov, Jakob & Jørg Asmussen, 2009: Compiling, annotating and publishing corpora in DK-CLARIN, the Danish incarnation of the pan-European initiative for a common research infrastructure. In: *Work-in-progress report. Corpus Linguistics*. Liverpool University.

Summary

Language technology is important for the status and the use of Nordic languages in all domains. Language institutions have an obligation to ensure that language technology is developed for their language. The Nordic language councils have worked eagerly to promote the knowledge of language technology and to bring together relevant players in the field in order to strengthen Nordic cooperation. The language council working group for language technology in the Nordic countries (ASTIN) has held a number of successful conferences, seminars and workshops on various aspects of language technology. The article describes the results of

ASTIN's work, and how the cooperation on language technology can be strengthened in order to meet the challenges of the information society of the future. The recommendations include the sharing of language data in a strong language technology infrastructure and a constant focus on the education of language technology experts with competence in Nordic languages.

Sprog i fokus: Islandsk

Jóhannes B. Sigtryggsson

Islandsk er et vestnordisk sprog, og færøsk er dets nærmeste slægtning. Artiklen er en kort præsentation af det islandske sprogs historie, udtale, grammatik, orddannelse osv. Der er også en kort redegørelse for purisme og sprogrøgt, som har været stærk i Island gennem tiderne. Sprogets officielle status i Island har ændret sig ret meget i de sidste år, mest pga. en ny lov om islandsk og islandsk tegnsprog fra 2011.

Udbredelse

I begyndelsen af 2013 boede der i Island omtrent 322.000 mennesker. Deraf var cirka 21.500 udenlandske statsborgere (6.7%) (Hagstofa Íslands, www.hagstofan.is). Der savnes en officiel sprogstatistik, men man antager at langt de fleste islandske statsborgere i Island har islandsk sprog som modersmål. Det eneste indenlandske minoritetssprog er islandsk tegnsprog som har cirka 250 modernmålsbrugere (Hilmarsson-Dunn & Ari Páll Kristinsson 2010, s. 102). I slutningen af 1800-tallet emigrerede titusinder af islændinge, de fleste til Canada. Kun et fåtal af deres efterkommere taler nu islandsk i noget omfang.

Der er undervist i islandsk ved omtrent 100 universiteter i 25 lande, deraf moderne islandsk ved cirka 40 universiteter.

Sproghistorie

Islandsk er et vestnordisk sprog. De nærmeste slægtninge er færøsk og fjernere (sydvest)norsk. De ældste bevarede håndskriftsfragmenter skrevne på islandsk er fra midten af 1100-tallet. På 1100-tallet var der kun små forskelle mellem islandsk og norsk, men de blev flere under 1200-tallet og især under 1300-tallet.

Islandsk udtale har ændret sig ret meget fra 1100-tallet, men ordforrådet og strukturen af sproget har for største delen holdt sig intakt. Udtalen har ændret sig mere end man skulle tro hvis man ser udelukkende på retskrivningen. Den moderne islandske retskrivning skjuler hvor meget udtalen i det islandske sprog har ændret sig fra oldsprogets tid. Her

er et eksempel på hvordan en berømt sætning i Njáls saga fra 1200-tallet er skrevet:

Ek var ung gefin Njáli ok hefi ek því heitit honum at eitt skyldi ganga yfir okkr bæði. (oldislandsk, normaliseret)

Ég var ung gefin Njáli og hef ég því heitið honum að eitt skyldi ganga yfir okkur bæði. (moderne islandsk)

Det normaliserede skriftsprog fra den ældste periode er endnu forståeligt for islændinge. Derimod har lydssystemet ændret sig ret meget, især vokalerne. Den største ændring i lydssystemet fra oldislandsk er at vokallængde nu betinges af antallet af efterfølgende konsonanter, men er ikke indbygget i hver vokal (og diftong). F.eks. oldislandsk: *áll* [au:l:] '(nom.sg.) ål, strømfure' - *ál* [au:l] '(akk.sg.) ål, strømfure' men nyislandsk *áll* [autl] - *ál* [au:l].

Status

I de senere år har islandsk lovgivning ændret sig meget hvad angår sproget. Den islandske sprogpolitik (*Íslenska til alls* 2009, <http://www.menntamalaraduneyti.is/nyrit/nr/5201>) blev vedtaget i Altinget den 12. marts 2009. Islandsk sprognævn skrev teksten til deklARATIONEN og lagde hovedvægt på sprogets status og dens vigtighed for sprogets overlevelse.

Deklarationens hovedformål er at sikre at islandsk bliver brugt på alle områder i det islandske samfund. Man fremhæver i deklARATIONEN diverse områder i samfundet, f.eks. forretningslivet og universiteterne.

Det vigtigste skridt i udviklingen af lovgivningen blev taget den 7. juni 2011 da lov nr. 61/2011 om det islandske sprogs og det islandske tegnsprogs status (*Lög um stöðu íslenskrar tungu og íslensks tákn máls*, <http://www.althingi.is/altext/stjt/2011.061.html>) blev vedtaget. I lovenes 1. paragraf står: "Íslenska er þjóðtunga Íslendinga og opinbert mál á Íslandi." ['Islandsk er Islændingenes nationalsprog og det officielle sprog i Island.'] I paragraf 6 står der også: "Íslenska er mál Alþingis, dómstóla, stjórnvalda, jafnt ríkis sem sveitarfélaga, skóla á öllum skólastigum og annarra stofnana sem hafa með höndum framkvæmdir og veita almannþjónustu." ['Islandsk er Altingets, domstolernes, myndighedernes, både de statlige og kommunale, skolernes i alle skolestadier og andre tjenstlige institutioners sprog.']

Indtil denne lovgivning har det ikke stået klart i landets lov at islandsk skulle være det officielle sprog i Island, men dets status har været støttet af diverse love, f.eks. om at sproget i domstolene skal være islandsk (Hilmarsson-Dunn & Ari Páll Kristinsson 2010, s. 120, 136–141). Man syntes ikke nødvendigt at præcisere sprogets status og rettigheder skriftligt i lovgivningen, sprogets status var for de fleste en selvfølge. Men nu når samfundet bliver stadig mere multikulturelt, er det blevet nødvendigt at vedtage love om det islandske sprogs og sprogbrugerens rettigheder. Det islandske samfund har ændret sig over nogle årtier fra at være ensartet hvor næsten alle talte det samme sprog (islandsk) til at have mange forskellige nationaliteter og sprog som polsk og litauisk.

Loven fik stor og bred støtte i Altinget og fra almenheden.

Tegnsprog

I loven om det islandske sprogs og det islandske tegnsprogs status af 2011 står i paragraf 3: “Íslenskt táknmál er fyrsta mál þeirra sem þurfa að reiða sig á það til tjáningar og samskipta og barna þeirra.” [‘Islandsk tegnsprog er det første sprog for dem som er nødt til at bruge det i kommunikationen også med deres børn.’] I tilknytning til loven blev der også oprettet et specielt sprognævn som kaldes *Málnefnd um íslenskt táknmál* [‘Sprognævn for det islandske tegnsprog’]. Med loven har tegnsprog og tegnsprogsbrugere for første gang fået vigtige rettigheder.

Alfabet og udtale

aA áÁ bB dD ðÐ eE éÉ fF gG hH iI íÍ jJ kK lL mM nN oO óÓ pP rR sS tT uU úÚ vV xX yY ýÝ þÞ æÆ öÖ.

Bogstavet *þ* (thorn) var allerede i brug i de ældste håndskrifter fra midten af 1100-tallet. Det har været brugt uafbrudt i islandsk skrift i gennem tiderne. Bogstavet *ð* (eth) forsvandt derimod ud af skriftsproget omkring 1400. Det kom igen ind i alfabetet i 1800-tallet, blandt andet under indflydelse fra den danske sprogforsker Rasmus Kristian Rask, og har siden været brugt i skriftsproget.

Bogstaverne *i* og *y* har den samme udtale i nyislandsk ([i(:)]). Det samme gælder også for *í* og *ý* ([i(:)]) og diftongerne *ei* og *ey* ([ei(:)]).

Accenttegn over vokaler betegner anden lyd kvalitet end de tilsvarende bogstaver uden accenttegn, f.eks. *a* [a(:)] men *á* [au(:)], *i* [i(:)] men *í* [i(:)], *o* [ɔ(:)] men *ó* [ou(:)], *u* [(:)] men *ú* [u(:)], *æ* [ai(:)], *ö* [ø(:)]. Bogstavet *é* står for udtalen [jɛ(:)], f.eks. *fé* [fjɛ:] ‘penge, får’. Bogstavet *e* har udtalen

[ɛ(:)]. De oprindelige norrøne diftonger er i islandsk blevet til *ei/ey* [ei(:)], *au* [øi(:)].

Nogle særtræk i islandsk udtale kan nævnes her. Vokaler i trykstærk stavelse kan enten være korte eller lange, betinget af hvor mange konsonanter der følger efter vokalen, f.eks. *hús* [hu:s] ‘hus (nominativ sg.)’ men *húss* [hus:] ‘hus (genitiv sg.)’. Diftonger kan også enten være korte (f.eks. *nótt* [nouht] ‘nat’) eller lange (f.eks. *nót* [nou:t] ‘not’). Præaspiration findes i konsonantgrupperne *pp*, *tt*, *kk* og i *p(p)*, *t(t)*, *k(k)* foran *l*, *m*, *n*, f.eks. *stappa* [stahpa] ‘stampe’, *Atli* [ahtlɪ], *drukknir* [tryhknɪr] ‘fuld (nominativ pl.)’. Ustemte *l*, *n*, *r*, *j* forekommer i forlyd (skreve *hl-*, *hn-*, *hr-*, *hj-*), f.eks. *hlaða* [l̥a:ða] ‘hølade’, *hnöta* [n̥ota] ‘nøddetræ’, *hrífa* [ri:va] ‘rive’, *hjól* [çou:l] ‘cykel’. Ustemte varianter af nasaler forekommer foran klusilerne *p*, *t*, *k* (*kj*) og en ustemt *l* foran *p*, *k* (*kj*), f.eks. *hampur* [hampɸr] ‘hamp’, *vanta* [vaŋta] ‘mangle’, *banki* [paŋci] ‘bank’, *úlpa* [ulpa] ‘vinterjakke’, *hylki* [hɪci] ‘hylster’. Et dialekttræk med stemte nasaler foran disse klusiler forekommer i Nordlandet, f.eks. *hampur* [hampʰɸr].

Ordenes hovedtryk er altid på første stavelse i islandsk som i finsk. Et ekstra tryk er på tredje stavelse og siden hveranden stavelse.

Dialekter

Et unikt særpræg ved islandsk er at der ikke findes dialekter, kun mindre dialekttræk (Hilmarsson-Dunn & Ari Páll Kristinsson 2010, s. 101). Et udbredt dialekttræk i Nordlandet er at klusiler bag en lang vokal er aspirerede, f.eks. *tapa* [ta:pʰa] ‘tabe’, *vaka* [va:kʰa] ‘være vågen’ (andre steder er udtalen uden aspiration: [ta:pa], [va:ka]).

Grammatik

Det oldislandske bøjningssystem har for største delen holdt sig intakt i nyislandsk. Substantiver, adjektiver, pronomener, bestemt og ubestemt artikel og nogle talord bøjes i fire kasus, i singularis og pluralis og i tre grammatiske køn (hankøn, hunkøn, intetkøn). F.eks. ordet *hestur* ‘hest’:

Tabel 1

	sg.	pl.
nominativ	<i>hestur</i>	<i>hestar</i>
akkusativ	<i>hest</i>	<i>hesta</i>
dativ	<i>hesti</i>	<i>hestum</i>
genitiv	<i>hests</i>	<i>hesta</i>

Bøjningsmønster for ordet *hestur* ‘hest’.

Verber bøjes i tid, person, tal og modus og følger enten stærk eller svag bøjning. De bøjes i flere forskellige bøjningsmønstre, specielt inden for de stærke verber hvor aflyd spiller en stor rolle. F.eks. verbet *skjóta* ‘skyde’:

Tabel 2

infinitiv	<i>skjóta</i>
3. p. sg. nutid	<i>skýtur</i>
3. p. sg. datid	<i>skaut</i>
1. p. pl. datid	<i>skutum</i>
m. sg. nom. participium	<i>skotinn</i>

Bøjningsmønster for verbet *skjóta* ‘skyde’:

Orddannelse

De fleste af de almindeligste ord som bruges er fremdeles de samme som i oldislandsk, f.eks. *faðir* ‘far’, *kýr* ‘ko’, *stökkva* (oldislandsk *stökkva*) ‘springe, hoppe’. Nye ord har selvfølgelig været tilføjet gennem tiderne, og de laves enten af eksisterende ordmateriale eller med tilpasning af låneord. Siden 1700-tallet har der været sat pris på dette pga. gennemsigtighed og bevaring af sproget (Ari Páll Kristinsson 2002, s. 140).

Islandsk orddannelse bygger i højere grad end i de andre skandinaviske sprog på ældre hjemlige orddele som bruges til at lave nye ord (Ari Páll Kristinsson 2002, s. 140–141). Nye ord af hjemlige rødder er enten lavet med præfikser/suffikser/aflyd eller med sammensætning. Som eksempler på det førstnævnte kan nævnes ord som *lyfta* ‘elevatør’ og *þota* ‘jetfly’. Af stammen *kvik-*, *kveik-*, som betegner hurtighed eller livlighed, og findes i ord som *kvikindi* ‘kryb’, *kveikja* ‘tænde’, *kvikna* ‘opstå, blive tændt’, har man lavet nye ord som *kvikmynd* ‘film, spillefilm’ og *kvikari* (forholdsvist nyt term for animator). Det er også populært at lave nye ord med sammensætning. Eksempler på det er *stjörnufræði* af *stjarna* ‘stjerne’ og *fræði* ‘videnskab’. Her er eksempler om nogle forholdsvis nye neologismer hvor sammensætning har været brugt: *áfallahjálp* ‘krisehjælp’, *endurvinnsla* ‘genbrug’, *fellihýsi* ‘(af engelsk) folding trailer (sammenklappelig anhænger)’, *fjölmennning* ‘multikulturalisme’, *víðóma* ‘stereo’, *þolfimi* ‘aerobic’. Der er også mange låneord i islandsk. Størsteparten er tilpasset retskrivningen og lydsystemet. Eksempler om tilpassede låneord: *biskup* ‘biskop’, *kaffi* ‘kaffe’, *sjeik* ‘sheik’, *jeppi* ‘firehjulstrækker (SUV /Jeep)’, *blogg* ‘blog’.

Syntaks

Rækkefølgen i islandske sætninger er oftest subjekt – verbum – objekt. Pga. af det rige bøjningssystem er ordenes rækkefølge ret fri i islandske sætninger. Her er nogle enkelte sætninger i islandsk:

Systir mín fór í ferðalag. (Min søster gik på en rejse.)

Fór systir mín í ferðalag? (Gik min søster på en rejse?)

Bílararnir eru hvítir. (Bilerne er hvide.)

Navne

I islandsk lever endnu det gamle germanske navnesystem hvor efternavne normalt laves af farens (eller morens) efternavn. Farens (eller morens) navn får en genitivendelse og til dem tilføjes enten *-son* eller *-dóttir*. Hvis en mand som hedder *Þormóður Sigtryggsson* har børn vil de komme til at hedde *Þormóðsson* (søn) eller *Þormóðsdóttir* (datter). Over 80% af de børn som nu bliver navngivet får dobbelte fornavne og den skik er dominerende at det første navn gerne er bisyllabisk eller polysyllabisk, mens det sidste er monosyllabisk, f.eks. *Ari Freyr*, *Linda Björk*.

Den nugældende navnelov i Island er lov nr. 45 fra 1996. Loven gør krav på at navne af udenlandsk oprindelse tilpasses fonologisk og morfologisk, og at de følger islandsk retskrivning. Loven fortsætter også et forbud, som har været en del af den islandske navnelovgivning fra 1925, mod optagelse af nye familienavne.

Sprogrøgt

Islandsk har en stærk litteraturtradition og en konservativ og beskyttende sprogpolitik (Hilmarsson-Dunn & Ari Páll Kristinsson 2010, s. 100). Sprogpurisme har stået stærk i Island fra 1800-tallet, men allerede i 1500-tallet kan man finde sprogpuristiske skrifter. Purismen har altid været stærkt tilknyttet til den gamle islandske kulturarv, specielt sagaerne. Den vigtigste ide i purismen har været at bevare islandsk med henblik på at bevare den kontinuiteten i skriftsproget som har været til stede fra 1100-tallet (Ari Páll Kristinsson 2002:140). I 1800- og 1900-tallet prægedes purismen også meget af kampen mod danismer. Eksempler på danismer som nu har forsvundet ud af sproget er *altan* 'altan', som nu altid kaldes *svalir* i islandsk, og *fortó(f)* 'fortov' som nu kaldes *gangstétt*.

I året 1964 blev *Íslensk málnefnd* (Islandsk sprognævn) etableret. 1980 fik sprognævnet et sekretariat som nævntes *Íslensk málstöð*. I året 2006 ændredes *Íslensk málstöð* til sprogrøgtsafdeling (*málræktarsvið*) indenfor *Stofnun Árna Magnússonar í íslenskum fræðum* [‘Árni Magnússon-instituttet for islandske studier’] mens Islandsk sprognævn fik i loven andre opgaver, specielt sprogpoltiske. I den nye lov om det islandske sprogs og det islandske tegnsprogs status af 2011 er der også en paragraf om Islandsk sprognævn (paragraf 16) som har de samme opgaver og i loven af 2006. I Islandsk sprognævn skal der være 16 medlemmer fra forskellige organisationer i samfundet mens formanden og næstformanden er valgt af kultur- og uddannelsesministeriet. Nævnet har som opgave at rådgive og vejlede myndighederne om alt angående islandsk sprog. Det udgiver også hvert år en rapport om sprogets status. Hvert år har der været et specielt tema i rapporten. Sidste år, 2013, handlede den om islandsk som fremmedsprog.

En udførlig oversigt over sprogsituationen i Island findes hos Hilmarsson-Dunn & Ari Páll Kristinsson 2010.

Jóhannes B. Sigtryggsson er projektleder i sprogrøgtsafdelingen ved Árni Magnússon-instituttet i islandske studier, Island.

Litteratur

Ari Páll Kristinsson, 2002: Språk i fokus: Islandsk. *Språk i Norden* 2002, s. 137-141.

Ari Páll Kristinsson & Amanda Hilmarsson-Dunn, 2013: Recent developments in the language situation in Iceland. I: Kaplan, Robert B., R.B. Baldauf, Jr., N.M. Kamwangamalu (red.): *Language Planning in Europe. Cyprus, Iceland and Luxembourg*, s. 26-28. London and New York: Routledge.

Hagstofa Íslands, www.hagstofan.is

Hilmarsson-Dunn, Amanda & Ari Páll Kristinsson, 2010: The Language situation in Iceland. I: *Current Issues in Language Planning*, volume 11, issue 3, s. 207-276. London and New York: Routledge.

Íslenska til alls. Tillögur íslenskrar málnefndar að íslenskri málstefnu samþykktar á Alþingi 12. mars 2009. Mennta- og menningarmálaráðuneyti. <http://www.menntamalaraduneyti.is/nyrit/nr/5201>

Lög um stöðu íslenskrar tungu og íslensks táknmáls, nr. 61/2011.

<http://www.althingi.is/altext/stjt/2011.061.html>

Summary

Icelandic is a West-Nordic language most closely related to Faroese. The article gives a short description of its history, pronunciation, grammar, word-formation etc. In recent years the legal status of Icelandic has changed, esp. with new language laws from 2011 that concern both Icelandic and the Icelandic sign-language. There is also an account of the history of language purism in Iceland, which has been strong for a long time.

Det nordiske sprogsamarbejde i 2013

Anne Kjærgaard og Sabine Kircheimer-Andersen

Nordisk sprogkoordination koordinerer indsatsen for et velfungerende sprogfællesskab i Norden. Artiklen giver en oversigt over fremdriften for projekter som Sprogkoordinationen har hovedansvar for eller indgår som samarbejdspartner i. Netværket er ansvarligt for projekterne under Netværket for Sprognævnene i Norden, mens Sprogkoordinationen har et samordnende ansvar for projekterne og modtager rapporter om fremdriften.

Ved at oprette Nordisk Sprogkoordination (NOSK) og Program for Sprogkurser og ved at tildele øremærkede projektmidler til aktiviteter i regi af Netværket for Sprognævnene i Norden ønskede Nordisk ministerråd at:

- koordinere aktiviteter på sprogområdet som ikke før blev koordineret
- give aktører på sprogområdet bedre kendskab til hinanden og motivere dem til at samarbejde
- få synergieffekt ud af samspillet mellem sprognævnene/sprogrådene, samt koordineringen af Program for sprogkurser, dvs. Nordspråk, Nordkurs og de Nordiske Perler
- skabe større effektivitet i arbejdet, idet der er en tidshorisont for planlægning, og der foretages løbende opfølgning på resultater
- skabe kontinuitet i driften
- gøre det nordiske sprogsamarbejde mere synligt for omverdenen

NOSK skal nå målsætningerne ved at koordinere aktiviteterne som foregår i regi af Netværket for Sprognævnene i Norden og Program for Sprogkurser (dvs. Nordkurs, Nordspråk og De Nordiske Perler).

Bodil Aurstad tog orlov fra stillingen som nordisk sprogkoordinator

i begyndelsen af 2013, og i foråret 2013 blev FNF Foreningerne Norden udpeget til ny værtsinstitution for Nordisk Sprogkoordination fra 2014. Hulda Zober Holm blev ansat som ny nordisk sprogkoordinator pr. 1.1. 2014. Derfor har den hidtidige værtsinstitution Dansk Sprognævn i løbet af 2013 haft fokus på at fortsætte driften af NOSK samt at afslutte og afvikle igangværende projekter.

Projekter og aktiviteter under Netværket for Sprognævnene i Norden i 2013

En af Nordisk Sprogkoordinations hovedopgaver i perioden 2009-2013 var at administrere projekter under Netværket for Sprognævnene i Norden (NSN).

Projektene som netværket sætter i gang, skal så vidt muligt bidrage til at styrke tiltag som på nationalt niveau følger op på intentionerne bag den nordiske sprogdeklaration (fra 2006) og de fire arbejdsspørgsmål i deklARATIONEN.

Netværket disponerer over DKK 800 000. 13 % af midlerne dækker administrationsomkostninger i Dansk Sprognævn, mens resten af midlerne er afsat til netværksprojekter. NOSK administrerer netværkets projekter, mens netværkets ledergruppe prioriterer hvilke projekter der skal støttes. Projektmidlerne fordeles efter diskussioner på det årlige netværksmøde hvor hele netværket samles.

Sprog- og netværksmøde 2013

Det nordiske sprogmøde skal samle medlemmer fra NSN og sproginteresserede nordboer til en årlig konference om sprog. Konferencerne arrangeres hvert år et nyt sted i Norden for at sørge for at så mange nordboer som muligt kan få mulighed for at deltage i et sprogmøde. Det årlige netværksmøde for NSN afholdes umiddelbart før eller efter sprogmødet og er forbeholdt medlemmer af netværket.

Sprogmødet 2013 blev afholdt i Akureyri d. **28.-29. august 2013** i samarbejde mellem Íslensk málnefnd og NOSK. Temaet for konferencen var de nordiske sprogs fremtid inden for informationsteknologien, og programmet dækkede foredrag fra næsten hele Norden fra foredraget *Selvfølgerlig snakker vaskemaskiner både grønlandsk og samisk i fremtiden* af Per Langgård, seniorrådgiver hos Oqasileriffik og Sjur Nørstebø Moshagen hos UiT, Norges Arktiske Universitet til Krister Linden, forskningsdi-

rektør ved Helsingfors universitet, der fortalte om *FIN-CLARIN – integrering av språkresurser i Finland*.

Ekstraordinært netværksmøde

I forbindelse med reorganiseringen af sprogområdet blev der den 2. december 2014 afholdt et ekstraordinært netværksmøde for NSN. Formålet med mødet var at diskutere hvordan netværkets aktiviteter kan fortsætte efter 2013 hvor netværkets bevilling sættes ned fra 800.000 DKK til 250.000 DKK årligt. På mødet blev det aftalt hvilke projekter netværket skal fokusere på i 2013. På mødet blev også planerne for det kommende sprogkonference, der afholdes i Sverige 2014, diskuteret. Mødet fungerede desuden som samlet forberedelse til konferencen om nordisk sprogforståelse d. 3. december 2013 i Stockholm.

Holdninger til modersmålsundervisning – for eller imod på dagbladenes læserbrevssider?

Projektet bygger på en sammenligning mellem læserbreve om *modersmålsundervisning* (dvs. den betegnelse man bruger i Sverige om undervisning i og på andre modersmål end svensk) på dagbladenes læserbrevssider. Der er blevet lavet en gennemgang af svenske læserbreve i årene 2006-2011. Formålet med dette projekt er at foretage en tilsvarende gennemgang af danske og norske læserbreve (dog ikke nødvendigvis i samme periode). Projektet falder inden for det svenske Språkråds opgave med at overvåge situationen for minoritetssprogene i Sverige samt den svenske sproglövs (2009:600) bestemmelse om at myndigheder i Sverige skal give alle mennesker adgang til og mulighed for at anvende og udvikle deres sprog. Ved at undersøge offentlighedens holdninger til undervisning i fremmede sprog og udpege evt. forskelle de nordiske lande imellem, er ambitionen at kunne give indblik i hvordan denne type undervisning, og i forlængelse af dette også det mangesprogede samfund generelt, diskuteres i Sverige, Norge og Danmark.

I løbet af efteråret 2013 blev der påbegyndt en undersøgelse af hvordan data, altså læserbreve i norske og danske aviser, kunne indsamles. Projektlederen undersøgte de tilgængelige medietjenester, samlede information fra relevante bibliotekstjenester og er på det grundlag begyndt at formulere en plan for dataindsamlingen der forventes at blive påbegyndt inden sommeren 2014. Muligheden for at få hjælp til indsamlingen fra en assistent eller praktikant er også blevet undersøgt. Det er dog fort-

sat uklart om det kommer til at fungere tidsmæssigt. Alt i alt har udredningen af en passende indsamlingsmetode taget betydeligt længere tid end beregnet. Svenska bibliotek har desværre meget begrænset materiale fra nordiske dagblade hvilket indebærer at dataindsamlingen bliver betydeligt mere kompliceret og tidskrævende end antaget i begyndelsen af projektet. Projektet afsluttes i løbet af 2014.

Projektledelsen ligger hos Språkrådet i Sverige.

ASTIN (Arbejdsgruppen for sprogteknologi i Norden)

ASTIN er netværkets arbejdsgruppe for sprogteknologi i Norden. Gruppen arbejder med sprogteknologi fra et sprogpolitisk perspektiv og et sprogrøgtsperspektiv. Sprogteknologi er vigtigt for sprogene i Norden, og særlig vigtigt for sprog som grønlandsk og samisk. Begge disse sprog er repræsenteret i ASTIN.

ASTIN har i 2013 bidraget med indspil til det nordiske sprog møde som havde de nordiske sprogs fremtid i informationsteknologien som tema. ASTIN har også været repræsenteret i styregruppen for projektet Nordiske Sprog & Sprogforståelse i de nye Kommunikationsmedier (NoSSink) og deltog i den forbindelse i en paneldebat på slutkonferencen for projektet i 2013. ASTIN har endvidere været styregruppe for NORDSVAR-projektet.

Projektledelsen ligger i Språkrådet i Norge.

Det nordiske sprogsamarbejdes historie

Formålet med projektet er at skrive en rapport som skal ligge tilgængelig på nettet i et printbart format. Historikken vil være nyttig i forbindelse med kommende omorganiseringer af det nordiske samarbejde på sprogområdet, men også værdifuld for den som vil kigge tilbage og vide mere om hvad der blev gjort, og hvordan det blev gjort. For tiden drives Nordisk Ministerråds virksomhed mest i projektform. For at en sådan projektvirksomhed skal være meningsfuld og ikke bare resultere i en række satsninger uden indbyrdes sammenhæng, må den være led i en mere langsigtet strategi og knyttes til en kontinuerlig basisvirksomhed. Hele virksomheden må kunne overskues og analyseres med henblik på både fortiden og fremtiden.

I løbet af 2013 har projektdeltagerne holdt to møder. Den afsluttende redigering af rapporten er blevet noget forsinket, bl.a. afventer gruppen

stadig bidrag fra personer udefra. Rapporten forventes klar i begyndelsen af 2014.

Projektledelsen ligger hos Birgitta Lindgren der er pensioneret medarbejder ved Språkrådet i Sverige.

Klarsprog

I november 2013 afholdt klarsprogsgruppen klarsprogskonference i Helsinki for 90 deltagere fra hele Norden. Titlen og temaet for konferencen var Myndighetstexter ur medborgarperspektiv. Konferencen blev finansieret med midler fra Nordplus Språk, NOSK og Institutet för de inhemska språken i Finland. NOSK bidrog desuden til at planlægge og søge midler til konferencen.

I 2013 udkom konferencerapporten fra klarsprogskonferencen Juridisk sprog i Norden (der blev afholdt i Reykjavik 2011). Rapporten er redigeret af Ari Páll Kristinsson og Jóhannes B. Sigtryggsson og er tilgængelig på NOSKs hjemmeside (http://nordisksprogkoordination.org/temaer/organisationer/organisationer-1/nordiske-sprognaevn/copy2_of_sprognaevnenes-netvaerk/klarsprogspublikationer).

Klarsprogsgruppen arbejder desuden på en projektansøgning til et projekt der har til formål at kortlægge borgernes *holdninger* til myndighedstekster i de forskellige nordiske lande og sammenholde dem med myndighedernes *praksis*, altså tilrettelæggelsen og forankringen af klarsprogsarbejdet i de forskellige nordiske lande. Formålet med projektet er således at undersøge om landenes forskellige organisering af klarsprogsarbejdet påvirker borgerens holdninger til tekster fra offentlige myndigheder – og i givet fald hvordan. Der indsendes en ansøgning om midler til Nordplus Sprog i 2014. Det videre arbejde i 2014 tilrettelægges på grundlag af svaret på ansøgningen.

Projektledelsen for klarsprogsgruppen ligger i Institutet för de inhemska språken, Finland. Projektledelsen for projektet om nordiske borgeres holdninger til myndighedstekster og nordiske myndigheders klarsprogspraksis ligger hos Språkrådet i Norge og Dansk Sprognævn.

Nordisk Miniordbog (tidligere Nordisk barneordbok)

Nordisk Miniordbog henvender sig til børn i alderen +/-12 år og er udarbejdet af erfarne leksikografer og oversættere. Den indeholder 3245 opslagsord med korte forklaringer. Opslagsordene præsenteres samtidig på fire sprog: dansk, norsk bokmål, nynorsk og svensk. Ordbogen er også

søgbar via finsk, færøsk, grønlandsk, islandsk og samisk. Ordbogen kan bruges til at træne både læse- og lytteforståelse. Alle ord med forklaringer findes som lydfiler man kan klikke på og lytte til.

Ordbogen blev lanceret 1. februar 2013 i København. Ordbogen kan tilgås fra netadressen <http://miniordbok.org>, og den er tilgængelig sammen med tekster og arbejdsopgaver på netadressen: www.nordeniskolen.org.

Projektledelsen lå i Språkrådet i Norge.

Nordisk sprogbarometer

Sprogbarometerprojektets mål er at komme frem til fælles målepunkter som man siden kan bede de nationale statistikbureauer om at følge op på i forbindelse med nationale udredninger. Netværket samarbejder med europæiske sprogbarometerprojekter i regi af *European Federation of National Institutions for Language (EFNIL)* som udarbejder sprogstatusrapporter. I løbet af 2013 blev et nyt sæt data indsamlet fra 24 europæiske lande, deriblandt Island, Finland, Danmark, Norge og Sverige. Grønland og Færøerne er også blevet inviteret til at deltage og indlevere data. Der arbejdes nu på at visualisere de indsamlede oplysninger på nettet.

Projektledelsen ligger i Dansk Sprognævn.

Termbase til støtte for nordisk mobilitet

Termbasen vil kunne bidrage til at øge mobiliteten mellem de nordiske lande, særligt for elever, studerende og jobsøgende, idet termbasen vil modvirke sproglige barrierer i Norden og dermed øge mulighederne for nordisk samkvem. Den nordiske sprogkonvention giver nordiske borgere ret til at bruge deres eget sprog i kontakten med myndigheder og andre offentlige institutioner i andre nordiske lande. Dette gælder på vigtige områder knyttet til institutioner som domstole, sundheds-, social- og børneområdet, samt arbejdsmarkeds-, skatte, politi- og skolemyndigheder. De sproglige rettigheder giver imidlertid sprogbrugerne en række udfordringer, særligt hvad angår fagterminologien inden for disse områder. En fælles nordisk termbase vil dermed være et nyttigt hjælpemiddel til at skabe gensidig forståelse af vigtige fagbegreber for nordiske sprogbrugere. Projektet blev tildelt 35.000 DKK af NSN ved årets netværksmøde i Akureyri. Disse midler blev brugt til et diskussionsmøde i Oslo i november 2013. På mødet blev det besluttet at søge midler hos Nordplus og Nordisk Kulturfond i 2014.

Projektledelsen ligger i Norge. Derudover deltager Danmark, Sverige, Finland, Island og terminologiorganisationerne Terminologicentrum TNC, Terminologicentralen TSK og DANTERMcentret. Det kan være at Grønland, Færøerne og det samiske sprogområde på en eller anden måde slutter sig til projektet senere.

Nordsvar.org (tidligere svarbaseharmoniseringsprojektet)

Dansk Sprognævn og Språkrådet i Norge og i Sverige samarbejder om et svarbase-harmoniseringsprojekt: www.nordsvar.org. Her kan medarbejderne søge information fra de to andre landes svarbaser. I løbet af 2013 er programmet blevet finjusteret og testet, og en del tekniske fejl er blevet rettet. Ved projektets afslutning lå der 335 danske, 2407 svenske og 248 norske færdigredigerede svar offentligt tilgængeligt på nordsvar.org. Derudover kan sprognævnsmedarbejdere, forskere mv. med et særligt password få adgang til 1112 ikke-offentlige danske svar.

Der er mulighed for at de øvrige sprognævn kan koble sig på, og det vil koste ca. 50.000 DKK for hvert sprog i tekniske tilpasninger. Den årlige drift er på ca. 35.000 DKK og er blevet finansieret af NSN's projektmidler. Arbejdet med at lægge nye svar ind i basen finansieres af sprognævnene selv.

ASTIN er styregruppe for projektet og vil evaluere projektet. Projektledelsen ligger i Dansk Sprognævn.

Publiceringsplatform

Med midler fra bl.a. Nordplus Sprog- og Kulturprogrammet startede arbejdet med publiceringsplatformen i 2011. Formålet med publiceringsplatformen er at tilgængeliggøre og synliggøre serierne Nordiske Studier i Leksikograf og LexicoNordica (begge udgivet af Nordisk Forening for Leksikografi), Sprog i Norden (udgivet af NSN) og Klart sprog i Norden (udgivet af arbejdsgruppen for klarsprog under NSN). Ideen bag denne omlægning af publiceringspraksissen er at der er tale om en løsning der er miljøvenlig, øger tilgængeligheden for offentligheden, specielt unge brugere som fx studerende, og sikrer et arkiv over tekster der omhandler sprogvidenskabelige og sprogpoltiske temaer. Løsningen er desuden, på længere sigt, billigere.

I 2013 blev den oprindelige arbejdsgang ændret: Oprindeligt var planen at xml-opmærke alle filer for nemt at kunne genere ønskede formater. Det viste sig imidlertid vanskeligt og tidskrævende at få anvendelige

xml-filer ud af materialet, bl.a. fordi det indeholder mange specialtegn, og af økonomiske og tidsmæssige årsager var det nødvendigt at ændre strategi. Det blev i stedet besluttet at skanne alle relevante numre af publikationerne til pdf-filer og uploade dem i et open source-miljø, Open Journal System (OJS). Her får hver publikation sin egen hovedside hvor man kan se de forskellige numre, indholdsfortegnelser og fremsøge, læse eller printe de ønskede artikler. På OJS-plattformen der hostes af Statsbiblioteket i Århus, vil publikationerne både være sikret og eksponeret på bedste vis. Derudover er der, ved at anvende denne platform, mulighed for at samle og systematisere arbejdsgangene i forbindelse med redaktionsprocessen ét sted, og således vil de fremtidige numre kunne blive tilgængelige langt hurtigere end det tidligere har været tilfældet. Der er skannet 1571 artikler. Ca. 1/3 er forsynet med metadata og lagt ind i publiceringsplatformen. Projektet afsluttes i 2014.

Projektledelsen ligger i Dansk Sprognævn.

NoSSinK (Nordisk Sprog & Sprogforståelse i de nye Kommunikationsmedier)

NSN har bevilget støtte til at der etableres et forskningsnetværk med deltagere fra Dansk Sprognævn og universiteterne i Göteborg, Oslo, Bergen og Trondheim. Forskningsnetværket har samlet eksisterende viden på området sprogforandring og sproglig fornyelse i nye kommunikationsmedier i Norden. De konkrete resultater af arbejdet blev lagt frem ved et afsluttende symposium i København 2013, sammen med nationale statusrapporter over forskning på området. Der var ca. 100 deltagere ved symposiet, og arrangementet vakte i øvrigt stor interesse i pressen. Seminaret dokumenterede desuden en del grundlæggende aspekter som vil udgøre et godt fundament for det fremtidige arbejde inden for området. I forbindelse med statusrapporterne har netværket arbejdet med anbefalinger som er blevet oversendt NORIA-Netværket for Nordiske Sprog, Sprog- og Kommunikationskulturer for at gøre området til at satsningsområde for de nordiske forskningsprogrammer.

NSN er blevet holdt orienteret om projektets fremdrift og resultater via bloggen <http://nossink.org/>. Projektet afsluttes i 2014 med udgivelsen af en videnskabelig publikation fra det afsluttende seminar. Publikationen vil ud over et antal videnskabelige artikler også indeholde en redegørelse for de ophavsretslige og databeskyttelsesmæssige problemer der er knyt-

tet til undersøgelsen af sproget i de nye kommunikationsmedier, samt de anbefalinger som netværket har sendt til NORIA-netværket.

Projektledelsen ligger i Dansk Sprognævn.

Sprog i Norden

Sprog i Norden er en årbog der udgives af NSN. I artiklerne behandles aktuelle sproglige emner, og et vigtigt udgangspunkt for årsskriftet er præsentationer og diskussioner på det årlige nordiske sprog møde. Skriftet indeholder også en oversigt over nordiske publikationer om sprog og ord-bøger det pågældende år. De nordiske sekretærer sidder i redaktionen for udgivelsen, og de fleste bidrag er baseret på foredrag fra Sprog mødet det foregående år. I 2013 overtog Torbjørn Breivik, nordisk sekretær fra Sprogrådet i Norge, opgaven som hovedredaktør for Sprog i Norden 2012 og 2013. De to numre af Sprog i Norden blev færdigredigeret i løbet af 2013 og ligger nu tilgængelige på NOSKs hjemmeside (<http://nordisksprogkoordination.org/temaer/tidsskrifter-og-boeger/spraak-i-norden-publikationer>). I løbet af 2014 vil alle numre af Sprog i Norden blive tilgængelige på publiceringsplatformen.

Projektledelsen ligger hos Språkrådet i Norge.

Tegnsprogsnetværket

Tegnsprogsnetværket består af tre arbejdsgrupper som blev nedsat ved et arbejdsseminar i København 2012. Opgaven er at formulere projektansøgninger for tre forskellige projekter:

1. Informationsgruppen fokuserer på at udvikle informationsmaterialer om tegnsprogenes stilling i de nordiske lande. Informationsgruppen havde ikke behov for at mødes fysisk i løbet af 2013. Kontakten foregik i stedet via e-mail og videochat. Derimod ønskede de hjælp til at opstille et budget. Susanna Karlsson fra Språkrådet i Sverige blev adspurgt, hun kom gerne med forslag til fonde, men havde svært ved at afse tid til at opstille et budget. Desuden gik hun på barsel i april 2013. I det videre arbejde har projektgruppen derfor arbejdet med projektansøgningen på egen hånd.
2. Undervisningsgruppen fokuserer på undervisningsmateriale om de nordiske tegnsprog som kan anvendes i døveskoler eller ved tolkeuddannelser. Gruppen har haft svært ved at komme i gang med at skrive

ansøgninger og ønsker støtte til det videre arbejde, men de har ikke specificeret hvilken støtte de har behov for. De finder det også vanskeligt at arbejde fordi de befinder sig i tre forskellige lande, men har ikke anmodet om penge til et fysisk møde.

3. Korpusgruppen arbejder med at udvikle et projekt som har til formål at udvikle korpusser for de nordiske tegnsprog. Korpusgruppen holdt et fysisk møde i Stockholm i marts 2013 og kom frem til at der burde arrangeres et seminar hvor man kunne diskutere anvendelsen af korpusserne i et bredere perspektiv, og hvad man på det grundlag bør tage hensyn til ved opbygningen af korpusserne. Seminariet blev afholdt i København i december 2013 og blev finansieret vha. deltagerbetaling.

Tegnsprogsnetværket har endnu ikke fundet en endelig organisationsform. Indtil videre ligger koordineringsansvaret hos Språkrådet i Sverige. Tommy Lyxell er kontaktperson. og han har også ansvaret for tegnsprogsnetværkets blog Nordiska teckenspråk, <http://nordiskateckensprak.wordpress.com>. I Nordisk Ministerråd, embedsmandskomiteen for uddannelse og forskning (MR-U), pågår der diskussioner om at afsætte særskilte midler til tegnsprogsnetværket, men der foreligger endnu ikke nogen endelig beslutning.

Andre projekter og aktiviteter under Nordisk Sprogkoordinations ansvarsområde i 2013

I 2013 har NOSK især fokuseret på en række opgaver som der gøres nærmere rede for i det følgende.

Koordinering af Program for Sprogkurser

NOSK og projektlederen for Nordiske Sprogpiloter-projektet har jævnligt været i kontakt om planlægning af og støtte til konferencer samt støtte til sprogiloturser.

Som konsekvens af de forestående omorganiseringer af NOSK er koordineringen af programaktiviteter under Program for Sprogkurser især sket i regi af NMR, fx i form af en konference om nordisk sprogforståelse i Stockholm hvor NOSK og medlemmer af NSN sammen med de øvrige nordiske aktører på sprogområdet deltog i diskussionen om tilrettelæggelsen af det fremtidige nordiske sprogsamarbejde.

I samarbejde med Samarbejdsnævnet for Udlandslektorer (SNU), som

ikke formelt har en tilknytning til NOSK, blev der i januar 2013 afholdt et seminar om digital sprogundervisning.

Formidling om aktiviteter i Netværket for Sprognævnene i Norden og Program for Sprogkurser

Fokus for NOSK har i 2013 primært været på afslutning og afvikling af aktiviteter knyttet til perioden 2009-2013, og der har derfor været mindre fokus på opsøgende virksomhed.

NOSK blev dog repræsenteret af Thomas Henriksen i det etablerede benchmarkingnetværk, og NOSK havde også i 2013 observatørstatus i Parallelsprogsnetværket og var repræsenteret ved Parallelsprogsnetværkets konference "Parallelsproglighed og internationalisering på nordiske universiteter - dyrt men dejligt!".

Information og kommunikation om sprogområdet

NOSK informerer sit netværk og offentligheden om sine aktiviteter, primært via hjemmesiden www.nordisksprogkoordination.org. Tilknyttet hjemmesiden er der oprettet en google-kalender hvor der løbende bliver gjort opmærksom på relevante arrangementer. Formålet med NOSKs hjemmeside er at informere om det nordiske sprogsamarbejde og de aktører og aktiviteter som indgår i samarbejdet. I løbet af efteråret 2014 udsendte NOSK derudover et månedligt nyhedsbrev om bl.a. relevante arrangementer og udgivelser i Norden. Netstedet skal derudover være en resurse for undervisere, og link til undervisningsmateriale er derfor indtelt i målgrupper fordelt på alderstrin. Dette sidste formål dækkes også gennem samarbejdet med www.nordeniskolen.org.

Indsamling og nyttiggørelse af erfaringer og resultater fra Nordplus Sprog- og Kulturprogrammet

NOSK indsamler løbende erfaringer og resultater fra Nordplus Sprog- og Kulturprogrammet og fra Nordisk Kulturfond. Projekter som er blevet vurderet som særligt vellykkede, er blevet profileret i forbindelse med arrangementer rettet mod sproglærere.

I 2012 indgik NOSK i planlægningen af og ansøgningen om midler til et nordisk benchmarking-netværk. Netværket blev initieret af den nordiske perle Hanaholmen og fik midler fra Nordplus og Nordisk Kulturfond. Benchmarkingnetværket fokuserer på projekter som har fået sprogpriser eller har udmærket sig på anden vis, og formålet med netværket er at

identificere succesfaktorer og opstille anbefalinger for kommende projekter. Benchmarking-netværket har udskudt deres konference, der oprindeligt var planlagt til 2013, til 2014.

I forbindelse med parallelsprogsarbejdet har flere nordiske universiteter etableret et netværk af seniorforskere, juniorforskere og forskningsadministratorer. NOSK har observatørstatus i netværket og har løbende rapporteret fra netværkets aktiviteter til NSN's ledergruppe, se også <http://nordiskparallelsprogsnet.blogs.ku.dk/>.

Forskningsnetværket NoSSink har indsamlet eksisterende viden på området sprogforandring og sproglig fornyelse i nye kommunikationsmedier i Norden. De konkrete resultater af arbejdet blev præsenteret ved et afsluttende symposium i København 2013, sammen med nationale statusrapporter over forskning på området. I forbindelse med statusrapporterne har netværket udarbejdet en række anbefalinger som er blevet sendt til NORIA-Netværket for Nordiske Sprog, Sprog- og Kommunikationskulturer for at gøre området til et satsningsområde for de nordiske forskningsprogrammer.

Styrkelse af den nordiske sprogforståelse blandt børn og unge

I 2013 arrangerede NOSK i samarbejde med de nordiske sprogpiloter Sprogpilotkonference 2013 der over et 3-dagsprogram bl.a. kunne byde på oplæg om best practice i nabosprogsundervisningen og på forfatterbesøg ved Merete Pryds Helle. NOSK bidrog også til at arrangere Nord-språk-konferencen Det nordiske klasserum. Konferencen var åben, og temaet for konferencen var nordisk fagdidaktik. NOSK bidrog også med midler til sprogpilotkurset i november på Lysebu, herunder midler til selve programmet og rejsemidler til to færøske deltagere. Desuden bidrog NOSK med midler til Norsk Målungsdoms konference Språkmangfaldet i Norden der henvendte sig til unge der interesserer sig for nordiske sprog og sprogsituationen i de nordiske lande.

Nordisk Miniordbog blev lanceret ved et arrangement i København d. 1. februar 2013 i København. Ordbogen ligger sammen med tekster og arbejdsopgaver på netadressen <http://miniordbok.org/>. Miniordbogen er desuden tilgængelig via www.nordeniskolen.org (se i øvrigt beskrivelsen af Miniordbogen under netværkets aktiviteter).

Netværket har desuden overført overskydende midler fra 2012 og foregående år til nordeniskolen.org. NOSK, repræsenteret ved Bodil Aur-

stad, har desuden udarbejdet litteraturhistoriske materialer til www.nordeniskolen.org.

Netværket besluttede i 2013 at iværksætte en ny undersøgelse af børns og unges sprogforståelse.

Møder, konferencer og seminarer om sprogpleje og sprogbevaring, sprogrådgivning og sprogteknologi

Udover konferencen Det nordiske klasserum (arrangeret i samarbejde med Nordspråk, se ovenfor), arrangerede NOSK i samarbejde med NSN's sprog møde 2013 i Akureyri, Island. Temaet for konferencen var sprogteknologi, se i øvrigt beskrivelsen under NSN's aktiviteter.

Rapporter og andre skrifter

NOSK har i samarbejdet med NSN arbejdet videre på en elektronisk publiceringsplatform for nordiske skriftserier om sprog (se beskrivelsen under NSN's aktiviteter).

Sprog i Norden 2012 og 2013 blev færdigredigeret i løbet af 2013 og ligger nu tilgængelige på NOSKs hjemmeside (<http://nordisksprogkoordination.org/temaer/tidsskrifter-og-boeger/spraak-i-norden-publikationer>). Torbjørgh Breivik, nordisk sekretær fra Norsk Sprogråd, har været hovedredaktør af de to numre.

I 2013 udkom også konferencerapporten fra klarsprogs konferencen Juridisk sprog i Norden (der blev afholdt i Reykjavik 2011). Rapporten er redigeret af Ari Páll Kristinsson og Jóhannes B. Sigtryggsson og er tilgængelig på NOSKs hjemmeside (http://nordisksprogkoordination.org/temaer/organisationer/organisationer-1/nordiske-sproгнаevn/copy2_of_sproгнаevnenes-netvaerk/klarsprogspublikationer).

Hjemmeside for Netværket for Sprognævnene i Norden

Information om NSN, netværkets projekter og aktiviteter ligger som en del af NOSKs hjemmeside www.nordisksprogkoordination.org. Fra Sprogkoordinationens hjemmeside linkes der desuden til sprognævnenes egne hjemmesider, og de enkelte sprognævn er kort præsenteret.

Information om konferencer og publikationer m.m. er også sendt ud via nyhedsbreve og mailinglister.

NSN har desuden brugt deres lukkede blog (<http://nordisknyt.org/>) som arkiv og intranet for de nordiske sekretærer, ledere og arbejdsgrup-

per under netværket. På nordisknyt.org kan man bl.a. finde et arkiv over sagspapirer, mødepapirer, grundlagsdokumenter etc.

Anne Kjærgaard var (i samarbejde med Sabine Kirchmeier-Andersen og Anni Renner Mortensen) tovholder for Nordisk Sprogkoordination i 2013.

Sabine Kirchmeier-Andersen er direktør i Dansk Sprognævn.

Det nordiske sprogsamarbejde og Nordisk Sprogkoordination 2014–2018

Hulda Zober Holm

Organiseringen af sprogområdet i perioden 2014-2018 bygger på ønsker om at forenkle organiseringen af sprogområdet, opnå klarere linjer for referenceforhold, styrke Nordisk Sprogkoordination og koordineringen af aktiviteter samt inddrage kultursektoren.

Organiseringen af sprogområdet i perioden 2014-2018 er blevet vedtaget af Embedsmandskomitéen for Uddannelse og Forskning (EK-U) på baggrund af en evaluering af enhederne på sprogområdet, gennemført i årene 2011 og 2012. Evalueringsrapporten ”Med språkförståelsen i fokus - Evaluering av enheterna på språkområdet under Nordiska ministerrådet 2012” blev leveret til Nordisk Ministerråds Sekretariat af evaluator i maj 2012.

Att på allvar främja den nordiska språkförståelsen i enlighet med de politiska ambitionerna är ett omfattande och långsiktigt arbete. Men att skapa en organisation som på bästa sätt kan få ett genomslag hos barn och unga banar också väg för det nordiska samarbetet generellt och aktualiserar dess värde för omvärlden. En koncentration och ökning av språkområdets resurser är enligt evalueringens uppfattning därför nödvändig. (Citat fra evalueringsrapporten)

Fokusområder

EK-U har besluttet, at det nordiske sprogsamarbejde skal fokusere på følgende tre områder fra 2014:

1. De fire arbejdsspørgsmål i Deklaration om nordisk sprogpolitik (Sprogdeklarationen)

- Sprogforståelse og sprogkundskab
- Parallelsproglighed
- Mangesprogethed og flersprogethed
- Norden som sproglig foregangsregion

Sprogdeklarationen blev vedtaget den 1. november 2006 af Nordisk Ministerråd for Uddannelse og Forskning (MR-U). EK-U stadfæstede i 2008, at ansvaret for opfølgningen af Sprogdeklarationen er et nationalt anliggende. Fra 2014 skal det i højere grad undersøges, hvorvidt tiltag på sprogområdet kan understøtte nationale indsatser i opfølgningen af Sprogdeklarationen. Redegørelser for opfølgningen af Sprogdeklarationen er fremlagt på Nordisk Råds session i 2009, 2011 og 2013.

1. Indsatser for at styrke børn og unges sprogforståelse

Der skal være en fokuseret og målrettet indsats for at styrke børn og unges sprogforståelse i dansk, norsk og svensk. EK-U tog beslutningen om et særskilt fokus på arbejdet med børn og unge den 11. september 2008:

Beslutet om språksamarbetet handlar om at organiseringen [der trådte i kraft i 2009] skal bidra till en prioritering av en fokuserad och målinriktad insats för att barns och ungas förståelse för grannlandsspråken skall stärkas och att alla nordbor skall kunna kommunicera med varandra på ett skandinaviskt språk. ... ÄK-U beslöt följande: ... Særskilt fokus skal lægges på arbejdet med børn og unge.

2. Lytteforståelse

I skolerne har der hidtil været fokus på børn og unges læsefærdigheder, også når det gælder den såkaldte nabosprogsundervisning i fagene dansk, norsk og svensk. En undersøgelse af internordisk sprogforståelse har imidlertid vist, at udfordringerne med at forstå hinanden ligger i forståelsen af det talte sprog.

Forholdet mellem uddannelse og kultur

Sprog er et tværgående politikområde. I flere af de nordiske lande har kulturministrene ansvaret for sprogpolitik. Før den forrige omorganise-

ring på sprogområdet, der trådte i kraft den 1. januar 2009, havde MR-U og Nordisk Ministerråd for Kultur (MR-K) et delt ansvar for arbejdet med sprogpolitik, dog lå hovedansvaret for koordineringen af nordisk sprogpolitik i MR-U. Fra 2009 havde MR-U hovedansvaret for sprogområdet i Nordisk Ministerråd, mens MR-K havde et sektoransvar. Fra 2014 bibeholder MR-U hovedansvaret for sprogområdet i Nordisk Ministerråd, mens MR-K bibeholder sektoransvaret.

Skole og uddannelse er et vigtigt område for virkemidler på sprogområdet, men der er også vigtige virkemidler på kulturområdet som for eksempel computerspil, film, litteratur, medier, teater og tv. Det nye forvaltningsorgan skal i perioden 2014-2018 forholde sig til og gøre brug af begge sektors virkemidler med det formål at styrke sprogforståelsen, primært blandt børn og unge.

EK-U og Embedsmandskomiteén for Kultur (EK-K) har besluttet at indgå i en dialog for at styrke samspillet mellem uddannelse og kultur. I den forbindelse er der oprettet en ad hoc arbejdsgruppe for sprogområdet med to medlemmer fra EK-U og to medlemmer fra EK-K. Forvaltningsorganet skal i perioden 2014-2018 holdes orienteret af Nordisk Ministerråds Sekretariat om udviklingen i de to sektors samspil.

EK-U er styregruppe

EK-U overtog fra og med 2013 rollen som styregruppe for virksomheden på sprogområdet fra en af de forhenværende rådgivende grupper under MR-U; Nordisk Skolesamarbejde (NSS). EK-U er fortsat styregruppe for virksomheden på sprogområdet. Styregruppen skal sikre, at forvaltningsorganet administrerer sin virksomhed inden for de aftalte rammer. Styregruppen kan ændre aftalen med forvaltningsorganet, den samlede beløbsramme, sammensætningen af aktører og fordelingen af midlerne mellem aktørerne samt fokusområder. Styregruppen tager stilling til hvorvidt og i hvilken form der skal iværksættes en evaluering.

Foreningerne Nordens Forbund er nyt forvaltningsorgan

EK-U har udpeget Foreningerne Nordens Forbund (FNF) som nyt forvaltningsorgan for det nordiske sprogsamarbejde for perioden 2014-2018, som er finansieret af midler fra MR-Us budget. FNF har fået opgaven på baggrund af et tilbud der beskriver, hvordan FNF vil løse opgaven med at koordinere aktørerne samt hvilke selvstændige initiativer der vurde-

res som relevante for at styrke sprogforståelsen i dansk, norsk og svensk blandt børn og unge.

Skolen skal være omdrejningspunktet for strategiske indsatser på sprogområdet i hele den 5-årige periode. Målet er, at hvert eneste barn skal have en nordisk oplevelse under sin skoletid (Citat fra Foreningerne Nordens Forbunds tilbud)

Nordisk Sprogkoordination er således fysisk placeret i Foreningerne Nordens Forbunds sekretariat i København, og FNF har rekrutteret medarbejderne til Sprogkoordinationen. Sprogkoordinationen har i 2014 som de to vigtigste overordnede mål at opbygge og etablere en god koordinering af aktørerne og at opbygge og etablere gode kanaler for arbejdet med information og kommunikation, herunder en ny hjemmeside der i udtrykket skal ligne FNFs webbaserede undervisningsplatform Norden i Skolen.

FNF blev etableret i 1965 og er en paraplyorganisation for de nationale Foreningen Norden. Foreningen Norden blev stiftet i 1919 i Danmark, Norge og Sverige, og Foreningen Norden er i dag etableret i alle de nordiske lande samt Færøerne og Åland. De medarbejdere i de nationale foreninger der har et særligt ansvar for skoleområdet, vil få opgaver knyttet til de selvstændige initiativer, som FNF iværksætter i regi af Nordisk Sprogkoordination.

Aktører

De aktører som Nordisk Sprogkoordination skal koordinere er følgende:

- Program for Sprogkurser (der består af Nordkurs, Nordspråk og De Nordiske Perler)
- Nordiske Sprogpiloter
- Nordisk gruppe for Parallelsproglighed (midlertidig gruppe)
- Netværket for Sprognævnene i Norden

Ekspertgruppen Nordens Sprogråd

EK-U har besluttet at oprette en faglig referencegruppe for virksomheden på sprogområdet; Ekspertgruppen Nordens Sprogråd (ENS). EK-U har vedtaget et mandat for ENS som gælder i perioden 2014-2016. EK-U udpeger medlemmerne til ENS. Forvaltningsorganet skal anvende ENS til dels at opnå den fornødne faglige rådgivning, der via medlemmerne af referencegruppen dækker hele Norden, og dels til at sikre at der er sammenhæng mellem aktiviteterne på tværs af aktørerne. Nordisk Sprogkoordination har ansvaret for sekretariatsbetjeningen af ENS. Forvaltningsorganet skal fremlægge arbejdsplaner og årsrapporter for ENS og drøfte disse med gruppen inden arbejdsplaner og årsrapporter leveres til styregruppen EK-U via Nordisk Ministerråds Sekretariat. ENS skal udtale sig om arbejdsplanen.

Økonomi

MR-U har i 2014 afsat 6.911.000 DKK til sprogområdet. Beløbet er eksklusive midler til Nordplus Sprogprogrammet. Midlerne fordeles fra år til år. I 2014 fordeles midlerne på sprogområdet således:

Tabel 1

Nordisk Sprogkoordination (inklusive selvstændige initiativer)	2.290.349 DKK
Program for Sprogkurser (Nordkurs, Nordspråk, De nordiske Perler)	3.596.062 DKK
Nordiske Sprogpiloter	374.590 DKK
Netværket for Sprognævne i Norden	250.000 DKK
Nordisk gruppe for Parallelsproglighed	150.000 DKK
Ekspertgruppen Nordens Sprogråd	250.000 DKK

Fordeling af midler på sprogområdet i 2014

Årshjul for planer og afrapporteringer

Ifølge aftalen mellem Nordisk Ministerråd og forvaltningsorganet Forningerne Nordens Forbund (FNF) har forvaltningsorganet det overordnede ansvar for kvaliteten af hver enkelt opgave og aktør; strategiske planer for 2014-2018 samt arbejdsplan og årsrapport. Ifølge aftalen er det også forvaltningsorganet, der ved afslutningen af 2014 afgør, hvorvidt der

skal udarbejdes én samlet arbejdsplan og én samlet årsrapport for forvaltningsorganet og de enkelte tjenesteydelser. Nordisk Sprogkoordination ønsker i udgangspunktet at levere arbejdsplaner og årsrapporter i én samlet arbejdsplan og én samlet årsrapport.

Strategiske planer for 2014-2018

Forvaltningsorganet skal levere forvaltningsorganets og aktørernes strategiske planer 2014-2018 til styregruppen EK-U via Nordisk Ministerråds Sekretariat senest den 1. juni 2014. I den forbindelse har Nordisk Sprogkoordination indkaldt aktørerne til et strategi-seminar i foråret 2014. Forvaltningsorganet kan justere de strategiske planer. Såfremt der er tale om ændringer i substansen, skal styregruppen orienteres.

Arbejdsplaner

Forvaltningsorganet skal levere forvaltningsorganets og aktørernes arbejdsplaner til styregruppen EK-U via Nordisk Ministerråds Sekretariat senest den 1. november. Nordisk Ministerråds Sekretariat har således mulighed for at gå i dialog med forvaltningsorganet i november, inden arbejdsplanen forelægges for styregruppen. Forventede resultater for det kommende år skal fremgå af arbejdsplanen/arbejdsplanerne. Styregruppen skal godkende arbejdsplanen. Forvaltningsorganet kan justere arbejdsplanen. Såfremt der er tale om ændringer i substansen, skal styregruppen orienteres.

Årsrapporter

Forvaltningsorganet skal levere forvaltningsorganets og aktørernes årsrapport til styregruppen EK-U via Nordisk Ministerråds Sekretariat senest den 15. februar årligt. I årets første eller andet kvartal fremlægges årsrapporten for styregruppen. Hvorvidt det forgangne års forventede resultater blev opnået, skal fremgå af årsrapporten. Styregruppen skal godkende årsrapporten.

Nordiske aftaler

Forvaltningsorganet skal have kendskab til og i arbejdet forholde sig til følgende nordiske aftaler og andre dokumenter af betydning:

Helsingforsaftalen, som trådte i kraft den 1. juli 1962, særligt Artikel 8:

Undervisningen og uddannelsen skal i skolerne i de nordiske lændene i læmplig omfang omfatte undervisning om sprog, kultur og almindelige samfundsførelser i de øvrige nordiske lændene, indbefattet Færøerne, Grønland og Åland.

Aftale om kulturelt samarbejde, som trådte i kraft den 1. januar 1972, særligt Artikel 3, d:

(Aftaleparterne skal) fremme undervisningen i de andre nordiske lændenes sprog, kultur og samfundsførelser.

Deklaration om nordisk sprogpolitik fra 2006 (Sprogdeklarationen)

Strategier på undervisnings- og kulturområdet (såfremt sådanne findes)

Formandskabsprogrammer og sektorprogrammer

Chefkonsulent Hulda Zober Holm er leder af Nordisk Sprogkoordination.

Summary

The Nordic languages cooperation and Nordic Languages Coordination 2014-2018. The organization of the language field in the period 2014-2018 was adopted by the Committee of Senior Officials for Education and Research (EK-U). EK-U has decided that the Nordic language co-operation will focus on three areas from 2014: 1) The four issues in the Declaration on a Nordic language policy, 2) Efforts to strengthen children and young people's understanding of languages and 3) Listening comprehension. EK-U is the steering committee for the language field. The steering committee will ensure that the executive agency manages its business within the agreed framework. EK-U has appointed Confederation of Nordic Associations (Foreningerne Nordens Forbund - FNF) as the new execu-

tive body for the Nordic languages cooperation for the period 2014-2018 which is funded by the Council of Ministers for Education and Research (MR-U). Nordic Languages Coordination is physically located in FNFs Secretariat in Copenhagen.

Ny sproglitteratur 2013

Danmark

Jørgen Nørby Jensen

Andersen, Mette Skovgaard og Lisbeth Verstraete-Hansen: Hvad gør vi med sprog? Behov for og holdninger til fremmedsprog i den danske centraladministration i et uddannelsespolitisk perspektiv. Handelshøjskolen i København 2013. 120 s.

Rapporten omhandler behovet for og holdninger til fremmedsprog i centraladministrationen i Danmark. Er engelsk nok? Og er vi gode nok til fremmedsprog i Danmark? Det er blot et par af de spørgsmål som rapporten forsøger at afdække.

Betydning og forståelse. Festskrift til Hanne Ruus. Redigeret af Dorthe Duncker, Anne Mette Hansen, Karen Skovgaard-Petersen. Selskab for Nordisk Filologi. Københavns Universitet 2013. 396 s.

Betydning og forståelse et festskrift til Hanne Ruus, professor ved Institut for Nordiske Studier og Sprogvidenskab ved Københavns Universitet, udsendt i anledning af hendes 70-års fødselsdag. Af festskriftes i alt 29 artikler kan i denne sammenhæng fremhæves *Hans Basbøll*: Ordskemaet og Ikke-Stød, *Tanya Karoli Christensen*: Er nok nok?, *Erik Hansen*: Definit komparativ, *Henrik Galberg Jacobsen*: Filolog og/eller reformator. Filologerne og den ortografiske normering, *Sabine Kirchmeier-Andersen*: Om at faste hunden, gro skæg og vokse hestehale, *Frans Gregersen*: Noter til sprogsynets historie i Danmark fra Rasmus Rask til J. Normann Jørgensen.

Bjerre, Malene: Fjern fejlene. 2. udg. Forlaget Ajour 2013. 82 s.

Fjern fejlene giver svar på de mest almindelige tvivlsspørgsmål når det gælder om at skrive korrekt. Den er opdateret efter Retskrivningsordbogen fra 2012.

Danske Studier 2013. Udgivet af Merete K. Jørgensen og Henrik Blicher under medvirken af Simon Skovgaard Boeck. Universitets-Jubilæets danske Samfund nr. 583. 2013. 228 s.

Af bindets indhold er der i denne sammenhæng grund til at fremhæve *Christian Becker-Christensen*: Nudansk Ordbog 1953-2010, *Liisa Theilgaard*: Retskrivningsordbogen, 4. udgave (anmeldelse), *Henrik Galberg Jacobsen*: Claus Drenghed-Nielsen: Grammatik på dansk (anmeldelse).

Danske Talesprog, bind 12. Udgivet af Nordisk Forskningsinstitut, Afdeling for Dialektforskning. Museum Tusulanums Forlag 2012. 238 s.

Bindet indeholder syv foredrag fra symposiet *SJUSK 2011, Nordisk symposium om naturligt talesprog*, som blev afholdt på CBS 23.-24. november 2011: *Ditte Boeg Thomsen*: fonetisk reduktion og morfosyntaktisk uigen-nemsigtighed, *Jan Heegård*: Funktionel distinkthed, *Jan Heegård og Jacob Thøgersen*: "Her er pressens radioavis", *Holger Juul*: Utdaledistinkthed og stavefærdighed hos danske børn, *Nicolai Pharao*: Lukkelydsreduktioner i københavnsk rigsmål, *Ruben Schachtenhaufen*: Nulrealiseringer af verbalformer i dansk spontantale, *Anja Schüppert, Nanna Haug Hilton, Charlotte Gooskens och Vincent J. Van Heuven*: Stavelsesbortfall i modern danska, *Hans Basbøll, Laila Kjærbaek, Claus Lambertsen og Ditte Boeg Thomsen*: Fra lyd til ord.

Danske Talesprog, bind 13. Udgivet af Nordisk Forskningsinstitut, Afdeling for Dialektforskning. Museum Tusulanums Forlag 2013. 336 s.

Bindet indeholder afhandlingen: *Malene Monka*: Sted og sprogforandring - en undersøgelse af sprogforandring i virkelig tid hos mobile og bofaste informanter fra Odder, Vinderup og Tinglev.

Gode ord er bedre end guld. Festskrift til Henrik Jørgensen. Redigeret af Simon Borchmann, Inger Schoonderbeek Hansen, Tina Thode Hougaard, Ole Tøgeby og Peter Widell. Aarhus Universitet 2013. 540 s.

Gode ord er bedre end guld er et festskrift til Henrik Jørgensen, lektor i dansk sprog ved Aarhus Universitet. Det er udsendt i anledning af hans 60-års fødselsdag og indeholder i alt 29 artikler om forskellige sproglige forhold. Af artiklerne kan nævnes *Lars Brink*: Hvordan udtales en apposition?, *Erik Hansen*: Dativen. Et digt fra den danske romantik, *Henrik Galberg Jacobsen*: 2012-retskrivningen - i historisk perspektiv, *Eva Skafte Jensen*: Af de store blege gule, *Marianne Rathje*: Dropper lige subjektet -

subjektellipser i kommentarfeltet på ekstrabladet.dk, *Ole Ravnholt*: ... auf einem anderen Blatt. Genitiv, gruppegenitiv, s-genitiv i moderne dansk.

Hultgren, Anna Kristina: Parallelsproglighed på danske universiteter: En statusrapport 2013. Københavnerstuder i Tosprogethed. Bind C5. Københavns Universitet, 2013. 72 s.

Hensigten med denne rapport er at dokumentere brugen af engelsk og dansk på universiteterne i Danmark. Undersøgelsen fokuserer på tre områder: publicering, undervisning og antallet af internationale studerende og ansatte. Konklusionen er at det især er inden for forskningspublicering at brugen af engelsk er meget høj, og det gælder især inden for de naturvidenskabelige fag. Undervisningen foregår derimod stadig i overvejende grad på dansk, ligesom andelen af dansktalende studerende og ansatte stadig udgør langt den overvejende del.

Jacobsen, Henrik Galberg og Peter Stray Jørgensen: Håndbog i Nudansk. Politikens Forlag, København 2013, 6. udg. 696 s.

Håndbog i Nudansk har siden 1988 været standardopslagsværket i praktisk sprogbrug. Den giver svar på hvad der korrekt og god skriftlig og mundtlig sprogbrug i mange forskellige situationer. 6. udgave er en grundigt revideret udgave af 5. udgave fra 2005, og revisionen har primært haft til formål at bringe bogen i nøje overensstemmelse med den nyeste udgave af *Retskrivningsordbogen* fra Dansk Sprognævn. I en række af artiklerne er der ligeledes indsat informative bokse med overskriften "Nyt i 2012-retskrivningen" der i koncentreret form giver en oversigt over nyhederne inden for det emne eller problem som artiklen handler om. I lighed med tidligere udgaver indeholder 6. udgave desuden en udførlig litteraturfortegnelse. Der er kommet en række nye titler ind, og der er strøget en del ældre titler som er vanskelige at få fat på. Listen indeholder nu i alt ca. 500 henvisninger.

Laursen, Katja Årosin: "Det er sprogligt – selv hvor du ikke lægger mærke til det". – En empirisk undersøgelse af sproglige og faglige vanskeligheder hos farmaceutstuderende med dansk som andetsprog på Københavns Universitet. Københavnerstuder i Tosprogethed. Bind C4. Københavns Universitet, 2013. 115 s.

I denne bog undersøges følgende spørgsmål: 1) I hvor høj grad kan de vanskeligheder som de studerende med dansk som andetsprog på

School of Pharmaceutical Sciences ved Københavns Universitet oplever, betegnes som hhv. faglige og sproglige? 2) Hvad er argumenterne for og imod støttetiltag, som specifikt er rettet mod dansk som andetsprog? For at besvare spørgsmålene har forfatteren gennemført seks interviews med hhv. to studerende der deltog i det sprogstøttende kursus, en studievejleder, studielederen og den kursusansvarlige for faget Lægemedelformulering samt en danskunderviser fra Center for Internationalisering og Parallelsproglighed. Derudover indgår også en spørgeskemaundersøgelse besvaret af 94 farmaceutstuderende på 3. årgang.

Mål & Mæle. 35.+ 36. årgang. 2 numre. Redigeret af Kasper Boye, Carsten Elbro, Ken Farø, Erik Hansen og Jørn Lund 2013. 32 s. + 32 s.

Foruden brevkassen *Sprogligheder* med spørgsmål om sprog indeholder årgangen artikler af bl.a. *Rikke Franch: Gør du automatisk en tekst personlig?*, *Erik Hansen: Mangt om meget*, *Lars Trap-Jensen: Retskrivningsordbogen 2012 – en bedre ordbog?*, *Lars Brink: Nyere danske lydlove. Udtaleændringer siden 1955-generationen.* For 2012 har *Mål og Mæle* kåret *akutjob* til ”årets ord”.

Ny forskning i grammatik 20. Redigeret af Lisbeth Falster Jakobsen, Kirsten Jeppesen Kragh, Eva Skafte Jensen og Erlings Strudsholm. Udgivet af Institut for Sprog og Kommunikation, Syddansk Universitet, Odense. 2013. 219 s.

Ny forskning i grammatik giver læseren et indblik i den nyeste nationale sprogforskning inden for dansk og en lang række fremmedsprog. Af bindets 11 artikler kan nævnes *Kasper Boye: Sagforhold, propositioner og den danske indikativ*, *Lars Heltoft: Om neksusbegrebets nødvendighed*, *Henrik Jørgensen: Sætningsspidsens udfyldning – observationer fra et talesprogs materiale.*

NyS. Nydanske Sprogstudier. 44-2013. Redigeret af Tanya Karoli Christensen, Jan Heegaard, Torben Juel Jensen, Janus Mortensen, Nicolai Pharao, Marianne Rathje og Jann Scheuer. Udgivet af Dansk Sprognævn 2013. 108 s.

Bind 44 er et temanummer om *sproglig normering*. Det indeholder 3 artikler og 2 anmeldelser: *Niels Davidsen-Nielsen: Principia Orthographica*, *Anna Kristina Hultgren: Sproglig normering i callcentre og på universiteter – hvordan og hvorfor?*, *Dorthe Lønsmann: Normering eller nor-*

mer? Sprogpolitik og sprogvvalg i en dansk virksomhed med engelsk som concernsprog, *Kirsten Rask*: RO – Rettidig omhu? Anmeldelse af Dansk Sprognævn: *Retskrivningsordbogen*, 4. udg., *Tore Kristiansen*: Anmeldelse af Dansk Sprognævn: *Retskrivningsordbogen*, 4. udg.

NyS. Nydanske Sprogstudier. 45-2013. Redigeret af Tanya Karoli Christensen, Jan Heegaard, Torben Juel Jensen, Janus Mortensen, Nicolai Pharao, Marianne Rathje, Jørgen Schack og Jann Scheuer. Udgivet af Dansk Sprognævn 2013. 150 s.

Af indholdet i *Nys 45* kan nævnes *Jørgen Schack*: Dig og mig og vi to. Synspunkter på kasus i moderne dansk, *Anita Ågerup Jervelund* og *Jørgen Nørby Jensen*: Replik til Kirstens Rasks anmeldelse af RO 2012.

Nyt fra Sprognævnet 2013 nr. 1-4. 40 s. + 16 s. + 16 s. + 16 s. Dansk Sprognævn

Nummer 1 er et udvidet temanummer om Retskrivningsordbogen fra 2012. Det indeholder i alt 11 artikler skrevet af ordbogens redaktører og af Sprognævnets direktør. I nummeret kan man læse om ord der er kommet ind i ordbogen, og om ord der er gået ud, om ændrede staveformer, ordformer og bøjningsformer, om ændringer i retskrivningsreglerne, om arbejdet med ordbogen og om selve det lovgrundlag der fastlægger arbejdet. Af artiklerne kan nævnes *Anita Ågerup Jervelund*: En ny og bedre Retskrivningsordbog – det endelige resultat, *Jørgen Nørby Jensen*: Nye opslagsord i Retskrivningsordbogen 2012, *Jørgen Schack*: Substantiver på *-el*, *-en* og *-er*, *Margrethe Heidemann Andersen*: Bindestregsændringer i Retskrivningsordbogen 2012, *Sabine Kirchmeier-Andersen*: Retskrivningsordbogen i fremtiden – analog eller digital?

De sidste tre numre indeholder et udvalg af sproglige spørgsmål til Dansk Sprognævn, bogomtaler og artikler om sproglige forhold, fx *Margrethe Heidemann Andersen*: Superligakager. Om muffins, macarons, whoppies og andre kager (nr. 2), *Jørgen Schack*: Auten(ti)citet (nr. 3), *Anna Kristiansen*: Unges holdninger til stavefejl på Facebook (nr. 4).

Ord til hinanden. Modersmål-Selskabets årbog 2013. Modersmål-Selskabet 2013. 102 s.

Årbogen indeholder en række tekster om måden vi taler til hinanden på – i bredeste forstand. Af bogens artikler kan nævnes *Kristian Tvilling*:

Fucking høflige, Anne Kjærgaard: *Breve fra det offentlige*, Anders Fogh Jensen: *Tale der virker. Godt*.

Ordbog for sygeplejestuderende. Af Lissi Hansen. Gyldendal Akademisk, København 2013. 155 s. Illustreret.

Politikens Retskrivnings- og Betydningsordbog. Redigeret af Erik Høvring. 7. udg. Politikens Forlag, København 2013. 785 s.

Rask, Kirsten: Korrekturlæserens grundbog. 2. udg. Grafisk Litteratur 2013. 127 s.

Bogen indeholder kapitler om grammatik, retskrivning og tegnsætning samt over 700 øvelser. Den er opdateret efter 4. udgave af Retskrivningsordbogen fra 2012.

Rask. Internationalt tidsskrift for sprog og kommunikation. Vol. 37, april 2013 og vol. 38, august 2013. 127 s. + 424 s. University of Southern Denmark, Institute of Language and Communication.

Af indholdet i de to bind kan nævnes *Hans Götzsche*: Jag har haft ringet til lægen. Eller: Hvad skal vi med en referencegrammatik? En kritik af visse, mest formale, aspekter ved Hansen & Heltoft: *Grammatik over det Danske sprog* (vol. 37), *Thomas Hestbæk Andersen & Alexandra Holsting*: "Tag du bare til EM, Eriksen" – om det eksplicitte Subjekt i den danske imperativsætning (vol. 38), *Niels Davidsen-Nielsen*: Tense and aspect in English – What did we learn by cooperating? (vol. 38), *Flemming Smedegaard*: Organisationer som leksikogrammatik (vol. 38).

Reichstein, Tina: Skriveguide - værdiskabende professionel kommunikation. Forlaget Frydenlund, Frederiksberg 2013. 104 s.

Bogen er en guide til god skriftlig kommunikation – både på nettet, i e-mails, på sms, de sociale medier og i traditionelle papirtekster. Den lægger vægt på at der skal være en synlig sammenhæng mellem en organisations værdier og dens skriftlige kommunikation. Bogen henvender sig til alle der skriver professionelle tekster, enten på jobbet eller i forbindelse med studiet, og den giver råd om hvordan man skal disponere en tekst, hvordan man undgår de mest almindelige stavfejl, hvordan man forklarer fagord bedst muligt, og hvordan man kommunikerer effektivt.

Schoonderbeek Hansen, Inger, Tina Thode Hougaard og Peter Widell (red.): 14. Møde om Udforskningen af Dansk Sprog. Aarhus Universitet 11.-12. oktober 2012. Århus 2013. 516 s.

Siden 1986 har sprogforskere fra ind- og udland hvert andet år mødtes i Århus for at sætte fokus på udforskningen af det danske sprog. På det 14. møde deltog 44 forskere med i alt 34 foredrag. En række af dem er trykt i den foreliggende rapport sammen med et antal resumeer af bidragydere der fx har fået publiceret deres artikel andetsteds. Af rapportens artikler kan nævnes *Anna Kjærgaard*: Hvad mener du SKAT? – Modtageranalyser af breve fra SKAT, *Marianne Rathje*: ”sku allgvel ha købt penis forlænger creme :)” Afvigelser fra retskrivningsnormen i kommentarfeltet på ekstrabladet.dk, *Ole Ravnholt*: Arbejdsdelingen mellem appellativer og pronominer, *Jacob Thøgersen*: Det flade a i prestige- og i hverdagsprog. Har ”det flade a” sejret?

Sprogforum. Tidsskrift for sprog- og kulturpædagogik. Årgang 19, 2013, nr. 56 + 57. 100 s. + 96 s. Aarhus Universitetsforlag.

Nr. 56 er et temanummer om *indholdsbasert sprogundervisning* og indeholder bidrag af bl.a. *Michael Svendsen Pedersen*: Buschaufførprog, *Birgit Henriksen* og *Anne Sofie Jakobsen*: Akademiske ord og fraser – et nordisk projekt. Nr. 57 sætter fokus på *subjektivitet og flersprogethed* og indeholder artikler af bl.a. *Karen-Margrethe Frederiksen*: At finde sin stemme på dansk, *Jesse-Lee Costa Døllerup* og *Tanya Lindkvist*: Kunsten at lære sprog – Hvad sker der, når sprogindlæring flytter ud af klasseværelset og ind på museet?

Sprog & Samfund. Nyt fra Modersmål-Selskabet. 31. årgang, 2013, nr. 1-4. 16 s. + 16 s. + 16 s. + 16 s. Redigeret af Michael Blædel.

Foruden anmeldelser af sproglitteratur og omtaler og referater af sproglige arrangementer indeholder numrene kortere artikler om forskellige sproglige forhold, fx *Georg Søndergaard*: Dansk i middelalderen (nr. 1), *Kirsten Rask*: Den engelske syge til behandling (nr. 2), *Jørgen Chr. Wind Nielsen*: Er rigsdansk et ideal? (nr. 4).

Thurah, Thomas og Martin Kristiansen (red.): Sproget i de elektroniske medier. Hans Reitzels Forlag 2013. 280 s.

Denne bog handler om sproget i de elektroniske medier. Det er bogens bærende idé at sproget udgør kernen i det journalistiske fag. Bogen

indeholder to indledende kapitler om sproget, hhv. den basale grammatik og det journalistiske sprog, samt en række kapitler om sproget på de forskellige digitale platforme, radio, tv, nettet og de sociale medier. Sproget i de elektroniske medier henvender sig til de studerende på landets journalist- og kommunikationsuddannelser.

Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer. Bind 11 (kurv-lindorm). Udgivet af Afdeling for Dialektforskning. Nordisk Forskningsinstitut. Københavns Universitet. Syddansk Universitetsforlag 2013. 473 s.

Jørgen Nørby Jensen er informationsmedarbejder i Dansk Sprog­ nævn, Danmark.

Finland

Anna Maria Gustafsson

Folkmålsstudier 51. Red. Hanna Lehti-Eklund, Camilla Lindholm och Caroline Sandström. Meddelanden från Föreningen för nordisk filologi r.f. Helsingfors 2013. 141 s.

I nummer 51 av tidskriften Folkmålsstudier diskuterar Caroline Sandström den svenska dialektologins historia i Finland från 1800-talet till i dag. Artikeln *Från folkmålsstudier till interaktionell dialektologi* behandlar trender inom dialektologin i Finland: utveckling av metoder, teorier och centrala begrepp. Viktiga avhandlingar och andra verk inom dialektologin tas upp i artikeln.

Den finska dialektologen Nina Kamppi diskuterar hur polysema ord kan analyseras i ordboksarbete i artikeln *På jakt efter metoder för lexikografisk betydelseindelning*, medan Tomas Lehecka skriver om förhållandet mellan semantiska och syntaktiska preferenser hos adjektiv importerade från engelskan. Därtill ingår en artikel av Mats Landqvist där han studerar samtal till Sjukvårdsrådgivningen i Sverige som har lett till klagomål.

Folkmålsstudier publicerar numera Lectio praecursoria-föredrag, alltså inledande föredrag som hålls av disputanden på doktorsdisputationer.

Även detta nummer innehåller ett sådant: Jennica Thylin-Klaus lectio praecursoria med rubriken ”*Den finländska svenskan*” 1860–1920. *Tidig svensk språkplanering i Finland ur ett idéhistoriskt perspektiv*.

Haasteena näkökulma, Perspektivet som utmaning, Point of view as challenge, Perspektivität als Herausforderung. Red. Maria Eronen & Marinella Rodi-Risberg. VAKKI-symposium XXXIII. Vasa 2013. 412 s.

Publikationen innehåller 33 artiklar som baserar sig på föredrag från det VAKKI-symposium som hölls i Vasa den 7–8 februari 2013. VAKKI står för Forskarföreningen för översättning, fackspråklig kommunikation och flerspråkighet vid Vasa universitet. Temat för seminariet 2013 var ”Perspektivet som utmaning”.

I artiklarna diskuteras temat ur filosofiska, sociologiska, kulturella, pedagogiska, yrkesmässiga och historiska synvinklar. Artiklarna har referentgranskats enligt double blind-principen. Publikationen är flerspråkig och den finns också tillgänglig i elektronisk form: www.vakki.net/publications/no2_sve.html

Hakulinen, Kerkko & Sirkka Paikkala: Pariisista Papukaijannokkaan. Suomenkieliset ulkomaiden paikannimet ja niiden vieras-kieliset vastineet. [Från Paris till Papukaijanokka. Utländska ortnamn med finska namnformer och deras motsvarigheter på andra språk] Kotimaisten kielten keskuksen julkaisuja 172. Institutet för de inhemska språken. Helsingfors 2013. 492 s.

Uppslagsverket innehåller 2 500 namn på orter i världen som har ett finskt namn eller en finsk namnform som avviker från den form namnet har i det lokala språket, exempelvis *Pariisi* (sv. *Paris*) och *Pöytävuori* (sv. *Taffelberget*; afrikaans *Tafelberg*, eng. *Table Mountain*). I verket anges också för varje namn vad det är för en plats, var platsen finns och vad den heter på det lokala språket och på engelska. I många fall ges också det franska och det tyska namnet.

Namnen är alfabetiskt inordnade dels enligt sin finska namnform, dels enligt den form namnet har på det lokala språket eller på engelska. Förutom namnförteckningarna innehåller boken skrivregler för finska namn på orter i utlandet, samt en genomgång av hur man genom åren har sett på stavningen av denna typ av namn i finskan. Därtill finns ett avsnitt om internationell namnvård.

Heinonen, Tarja Riitta: Idiomien leksikaalinen kuvaus kielenkäytön ja vaihtelun näkökulmasta [En lexikal beskrivning av idiom med betoning på språkbruket och variation] Institutionen för moderna språk vid Helsingfors universitet. Helsingfors. 238 s.

Tarja Riitta Heinonen har i sin doktorsavhandling studerat hur idiomatiska uttryck varierar i språkbruket och i vilken form man letar efter dem i ordböcker. För undersökningen har hon använt sig av korpus- och tidningsmaterial. I idiom kan t.ex. verbet bytas ut mot ett annat verb med samma eller nästan samma betydelse. Idiomen varierar också ofta så att den form som är bekant för alla språkbrukare modifieras så att något centralt ord byts ut i ett visst sammanhang. I avhandlingen belyser författaren de här variationssätten och diskuterar möjligheterna att beakta dem i ordböckerna. Detta ger läsaren en god inblick i diskussionen om ordböckernas normativitet överlag. I ordböcker ingår vanligen endast vedertagna, allmänna former, men inga variationer av produktiva uttryck, vilket gör att ordboksanvändaren i värsta fall uppfattar de varierade formerna som icke-önskvärda. Avhandlingen finns även som elektronisk publikation: helda.helsinki.fi/bitstream/handle/10138/37918/heinonen_vaitoskirja.pdf?sequence=1

Henricson, Sofie: Svenska i finsk miljö: Interaktion, grammatik och flerspråkighet i samtal på svenska språköar i Finland. Helsingfors universitet, humanistiska fakulteten, Finska, finskugriska och nordiska institutionen, Nordiska språk. Helsingfors. 259 s.

I avhandlingen studeras arrangerade gruppsamtal bland ungdomar i svenska gymnasier på fyra s.k. svenska språköar i Finland, nämligen Björneborg, Kotka, Tammerfors och Uleåborg. I de här finska städerna utgör de svenska befolkningsgrupperna om ungefär 400–1200 individer endast mellan 0,2 och 1 procent av totalbefolkningen. Syftet med studien är att utöka kunskapen om finlandssvenskt samtalsspråk med den starkaste förankringen i finska samhällen. Materialet utgörs av cirka elva timmar video- och audioinspelade gruppsamtal, som har transkriberats enligt samtalsanalytiska metoder. Grupperna består av mellan tre och åtta 16–18-åriga deltagare.

Genomgående noteras ortvisa skillnader i samtalen inspelade i de fyra städerna, om än skillnader mellan inspelningsorterna snarast kan anses hänga ihop med samtalssituationen, hur väl samtalsdeltagarna känner varandra och vilken gemensam referensram de har.

Tillgänglig som pdf-version: helda.helsinki.fi/bitstream/handle/10138/40361/henricson_avhandling.pdf

Hyvä sanakirja: tieteellisiä, käytännöllisiä ja eettisiä näkökulmia leksikografiaan. [En bra ordbok: vetenskapliga, praktiska och etiska synpunkter på lexikografi] Red. Laura Tyysteri & Kaisa Lehtosalo. Publications of the Department of Finnish and Finno-Ugric Languages of the University of Turku 3, Turun yliopiston suomen ja suomalais-ugrilaisen kielentutkimuksen oppiaineen julkaisuja 3. Åbo. 212 s.

Syftet med den här artikelsamlingen är att ge en översikt av den samlade tysta kunskap som finns på ordboksområdet i Finland. De frågor som diskuteras i artiklarna gäller principerna för ordboksarbete och de val som lexikografen ställs inför i det praktiska arbetet. Ett övergripande tema för de tio artiklarna är lexikografins etik. Artikelförfattarna är ordboksarbetare och forskare inom området lexikografi. Artiklarna behandlar bland annat normativa och deskriptiva ordböcker samt arbetet med ordböcker som riktar sig till minoritetspråkstalare och invandrare.

Häggman, Larserik: Det nordiska Finland. Pohjola-Norden. Helsingfors. 88 s.

Rapporten klargör i vilken utsträckning finländska institutioner och organisationer är medvetna om och utnyttjar det nordiska samarbetet, ägnar sig åt det samt i vilken utsträckning nordiska kontakter hör till normal rutin. I rapporten kommer författaren också med förslag till utvidgat nordiskt samarbete och utnyttjande av nordiskt samarbete speciellt när det gäller ungdomar.

Boken finns även på finska och är tillgänglig som pdf-version på adressen: www.pohjola-norden.fi/filebank/3519-PN_Pohjoismainen_Suomi_raportti_SVE_12092013_PAINO.pdf

Kielikello. Kielenhuollon tiedotuslehti. [Finska språkbyråns tidskrift] Institutet för de inhemska språken. Helsingfors.

Tidskriften *Kielikello* utkom med fyra nummer också under 2013. I årets första nummer togs vetenskapens språk upp. I samma nummer informerades läsarna om en ny form av termarbete: en termbank som sammanställts med hjälp av talkoarbete, så kallad crowdsourcing. Vetenskapsbanken i Finland är ett tvärvetenskapligt projekt vars syfte är att bygga upp en öppen termdatabas som täcker alla vetenskapsgrenar i Finland

och som riktar sig till både forskarsamhället och allmänheten. Termbanken fungerar enligt wiki-principen, så att experter kan lägga in termer från sitt eget fackområde vilka sedan kan bearbetas vidare utifrån diskussionen på plattformen. Projektet inleddes med ett pilotprojekt inom botanik, språkvetenskap och juridik 2011. Finansiärer är Finlands Akademi och Helsingfors universitet.

Synpunkter på talspråkets olika former och forskningen på området presenterades och diskuterades i en artikel som redogjorde för arbetet med att bygga upp en omfattande artikelsamling om talspråket i Helsingfors. Artiklarna i samlingen har skrivits i anknytning till ett forskningsprojekt vars syfte är att belysa hela spektret av det språk som man hör i 2000-talets Helsingfors. Enligt forskarna är helsingforsiskt och helsingforsdrag inte enbart ett regionalt, utan också ett socialt begrepp. För forskarna är det också intressant att studera bl.a. vem som upplever sig själv som helsingforsare och vem som har rätt att se sig själv som "en äkta helsingforsare" i andras ögon.

Förhållandet mellan talspråket och skriftspråket var ämnet för en av artiklarna där skillnaderna mellan dessa språkformer diskuteras. Artikel författarna konstaterar att diskussionen och analysen av skillnaderna har hjälpt forskarna att se vad det är som är karaktäristiskt för talat respektive skrivet språk. Å andra sidan konstaterar de att trots de konstaterade skillnaderna är det också mycket som är gemensamt.

Extra krydda till ämnet gavs Kielikellos läsare genom artikel som behandlar ord som beskriver ljud och rörelser, s.k. imitativa eller onomatopoetiska interjektioner, t.ex. *räks*, *roiskis*, *pam*. Sådana här ord har vanligen förknippats med språket i tecknade serier, men en närmare titt på dem visar att de också är ett viktigt stilmedel i skriftspråk. De har många funktioner i olika slags texter, också i sakprosa. De ingår i allmänhet i konstruktioner, dvs. i vanliga strukturer i språket, och därmed blir de en del av ett mångbottnat nätverket av betydelser och konstruktioner.

Namnfrågor behandlades också i Kielekello under året, bland annat användningen av utländska ortnamn i finskan och olika aspekter på förnamn, t.ex. vilka nya förnamn som är på väg in i almanackan och hur namn med främmande ursprung anpassar sig till finskan, dvs. hur språkbrukarna böjer dem. Detta har bland annat undersökts genom en enkät. Dessutom ingick en artikel om hur vi språkbrukare avgör om ett namn bärs av en kvinna eller en man. Det finns en hel del namn där tolkningen inte är given.

I vanlig ordning innehöll varje nummer av *Kielikello* också svar på språkfrågor, bland annat sådana som berörde verbböjning, pronomen och syftningsförhållanden. Därtill presenterades språkvården på en stor dagstidning, nyord, bakgrunden till vissa fraser och dialektala fenomen. Dessutom publicerades rekommendationer från finska språknämnden.

Laurent, Linda: Svenska språköar och finska utskär. Magma-Studier 3/2013. Tankesmedjan Magma. Helsingfors. 118 s.

Författaren jämför i rapporten de så kallade finlandssvenska språköarnas verksamhet, styrkor och svagheter med varandra. Som svenska språköar i Finland betraktas framför allt städerna Björneborg, Kotka, Tammerfors och Uleåborg, som samtliga har en livskraftig svensk minoritet. De här "traditionella språköarna" har kompletteras med ett tiotal "nya" där svenskan på senare tid blivit attraktiv och där ett intresse för svenskspråkig verksamhet har uppstått. De svenska språköarna fungerar dels som ambassadörer för svenskt och tvåspråkigt arbete, dels som platser där barn oavsett språklig bakgrund får en möjlighet att få dagvård och gå i skola på svenska. I rapporten synliggörs också de människor och organisationer som arbetar för språket.

Rapporten ger också en översikt över "finska rum" i enspråkigt svenska kommuner och över "samiska rum" som ligger utanför samernas hembygdsområde. Tillgänglig på www.magma.fi/images/stories/reports/ms1303_sprakoar.pdf

Namn i stadsmiljö. Handlingar från NORNA:s 42 symposium i Helsingfors den 10–12 november 2011. Red. Leila Mattfolk, Maria Vidberg & Pamela Gustavsson. Skrifter 7. Norna-rapporter 90. Institutet för de inhemska språken. Helsingfors. 269 s.

Symposierapporten från det nordiska namnforskarsymposiet i Helsingfors i november 2011 innehåller artiklar om de olika slag av namnskick som finns och har funnits i våra städer, t.ex. personnamnsskicket i två svenska städer på 1600-talet, kvinnors namnfraser i Helsingfors i slutet av 1800-talet och föräldrars val av namn på sina barn i Göteborg på 2000-talet. Här ingår även texter om t.ex. namngivning av befästningar i Finland i slutet av 1700-talet eller hur 2000-talets mäklare i Stockholm marknadsför bostäder med attraktiva namn på bostadsområden. Därtill kan nämnas en artikel om vilka följder dagens kommunreformer i Finland kan ha på namnplaneringen.

Ordbok över Finlands svenska folkmål. Institutet för de inhemska språkens webbpublikationer 33. Helsingfors. Institutet för de inhemska språken. Nätpublikation: kaino.kotus.fi/fo/

Delar av den finlandssvenska dialektordboken, *Ordbok över Finlands svenska folkmål*, publicerades under året på nätet. Det första avsnittet som publicerades omfattar I–K och senare publiceras avsnitten A–H och L–O. Ordboken kommer att uppdateras fortlöpande.

Ordboken beskriver dialekterna i Österbotten, på Åland i Åboland och i Nyland och ger uppgifter om uttal, böjning och betydelser för alla socknar. Nätordboken har försetts med en mångsidig sökfunktion som gör det lätt för användaren att hitta i ordboksartiklarna. Man kan t.ex. söka på uttal och exempelfraser eller välja specifika dialektområden, orter eller långgivande språk.

Ordboken bygger på ett material som består av cirka 1 miljon orduppgifter som samlats in under åren 1860–1970.

Rapautuva kieli. Kirjoituksia vähemmistökielten kulumisesta ja kadosta. [Språk i förvittring. Texter om minoritetsspråk som förbrukas och försvinner] Red. Kimmo Granqvist & Päivi Rainò. Suomalaisen kirjallisuuden seuran toimituksia 1404. Helsingfors. 254 s.

Boken tar upp en mängd olika orsaker till språkdöd, både på individplan, men också på samhällsnivå. Orsakerna till attrition, dvs. försämrade språkkunskaper, kan vara till exempel historiska och politiska. Boken innehåller åtta artiklar som belyser attrition i Finland. Utsatta minoritetsspråk är sådana som naturligt nog drabbas av attrition, och boken diskuterar för första gången attrition som språkligt fenomen i Finland. De språk som behandlas är romani, jiddish, ingermanländska, samt finskt och finlandssvenskt teckenspråk.

Slotte-Lüttge, Anna & Liselott Forsman: Skolspråk och lärande. Guider och handböcker 2013:5. Utbildningsstyrelsen. Tammerfors. 249 s.

Boken är ett resultat av forsknings- och utvecklingsprojektet Skolspråk, som i sin tur är en del av den finländska Utbildningsstyrelsens projekt "Att möta flerspråkiga elever i skolan". Projektet har genomförts i samarbete med årskurserna 7 till 9 i fyra svenskspråkiga skolor i Finland. Boken bygger i hög grad på autentiska klassrumssituationer och skolvardag i olika språkmiljöer. Den beskriver hur språkinriktat och språkoptimistiskt arbete kan se ut inom ramen för skolans dagliga verksamhet – i alla ämnen.

Boken innehåller en stor mängd tips och verktyg för lärare med två- eller flerspråkiga elever i sina klassrum på olika skolstadier.

Boken finns också tillgänglig som pdf-version på www.oph.fi/download/150655_Skolsprak_och_larande.pdf

Språkbruk, tidskrift utgiven av Institutet för de inhemska språken. Helsingfors.

Språkbruk utkom i vanlig ordning med fyra nummer under år 2013. Charlotta af Hällström-Reijonen var chefredaktör och Pia Westerberg redaktionssekreterare. Monica Äikäs besvarade språkfrågorna i tidningen.

Årets första nummer innehöll två artiklar med medicinsk anknytning; Rune Skogbergs *Hjärta tack och sömnakne – om förvrängningar av medicinska ord* samt Fredrik Olssons artikel *Säsongsinfluensan i svenska sociala medier* där Olsson förklarar hur det går att upptäcka säsongsinfluensans styrka via sociala medier. Namnvårdaren Sirkka Paikkala skrev om valet av efternamn i äktenskapet. I en intervju berättade komikern André Wickström bland annat om språkliga utmaningar på den nordiska ståuppscenen.

I nummer 2/2013 diskuterade professor Leena Huss tvåspråkiga skolor, Leif Höckerstedt undrade om humanisterna räddar världen och Monica Äikäs tog upp betydelseglidningar i översatt text. Författaren och översättaren Robin Valtiala skrev i en artikel om processen när han experimentöversätter arabisk poesi till svenska. Språkbruks chefredaktör Charlotta af Hällström-Reijonen beskrev den finlandssvenska satsflätan, och huvudredaktören för Ordbok över Finlands svenska folkmål Caroline Sandström om ordet latadöuv (att vara latadöuv betyder ungefär 'att slå dövörat till').

I Språkbruk 3/2013 redogjorde Karin Milles för hen-debatten i Sverige och Finland, och om ordets etablering i de båda länderna. Namnforskaren Emilia Aldrin diskuterade hur föräldrar motiverar sina namnval när de väljer förnamn till sina barn. Översättarna vid Statsrådets translatorsbyrå Anna Gref och Diana Nyberg berättade om användningen av översättningsminnen vid translatorsbyrån. Martina Ollas redogjorde för hur djur låter i artikeln *Mjau! Djurlåten på olika språk*.

I Språkbruks sista nummer för 2013 redogjorde Anna Slotte-Lüttge och Liselott Forsman för hur en språkstödsmentor kan stärka svenskan i skolan. Eeva-Liisa Nyqvist skrev om hur finskspråkiga elever tillägnar sig de svenska substantivformerna i en viss logisk ordning, även om un-

dervisningsordningen i skolorna dessvärre inte följer denna naturliga inlärningsgång. Maja Skanding tog läsaren *På språkreise i Skavlandinavia*. Artikeln handlar om skandinavisk språkanpassning. I detta sista nummer 2013 inleddes också samarbetet med bildkonststuderande i Tölö specialiseringsgymnasium. Elever från olika årskurser kommer att bidra med Språkkonst i tidskriften.

I varje nummer ingick också en ledare, recensioner av ordböcker och annan språkvetenskaplig litteratur, språkfrågor och svar, notiser, rapporter från språkseminarier och språkkonferenser, samt annat redaktionellt material. Varje nummer innehöll därtill en krönika. Krönikörer under 2013 var Henrik Huldén, Maria Sundblom Lindberg, Katarina von Numers och Sven-Erik Klinkmann.

Sundman, Marketta: Tvåspråkiga skolor? En analys av fördelar och risker med införandet av skolor med svenska och finska som undervisningsspråk. Magma-Studier 4/2013. Tankesmedjan Magma. Helsingfors. 93 s.

I ljuset av internationella undersökningar klarar sig elever i tvåspråkiga skolor i medeltal anmärkningsvärt bra. I Finland saknas genuint tvåspråkiga skolor med undervisning på både svenska och finska för samma elever, och frågan har debatterats intensivt både bland pedagoger och på tidningarnas insändarsidor de senaste åren. I rapporten diskuterar författaren tvåspråkiga skolor som en fördel för tvåspråkiga elever och hon resonerar kring majoritetsspråkets eventuella dominans i sådana skolor. Sundman reder också ut oklarheterna kring begreppet tvåspråkig skola. Studien finns tillgänglig i pdf-version på www.magma.fi/images/stories/reports/ms1304_webb.pdf

Tandefelt, Marika: Prima vara! Språk- och stilhistoriska studier i finlandssvenska och svenska varuhusannonser under 1900-talet. Skrifter utgivna av Svenska litteratursällskapet i Finland nr 779. Studier i Nordisk filologi 85. Helsingfors. 221 sidor.

I boken analyserar författaren svensk språkutveckling under 1900-talet genom att studera tidningsannonser som varuhuset Stockmann i Helsingfors och Nordiska Kompaniet i Stockholm har publicerat. Hur förändras språket i tidningsannonserna i jämförelse med den allmänna språkutvecklingen? Förändras språket i olika takt i Sverige och Finland? Studierna visar bland annat hur rättstavningsreformen 1906 långsamt börjar tilläm-

pas i annonsspråket, hur verben förlorar sin pluraländelse och när varuhuset börjar dua sina kunder. Boken ger en bild av dessa förändringar ur ett jämförande sverige- och finlandssvenskt perspektiv.

Tiittula, Liisa & Pirkko Nuolijärvi: Puheen illuusio suomenkielissä kaunokirjallisuudessa. [Talspråksillusionen i finsk skönlitteratur] Suomalaisen Kirjallisuuden Seuran Toimituksia 1401. Helsingfors. 637 s.

Boken ger en heltäckande bild av återgivningen av talat språk i den finskspråkiga skönlitteraturen från 1800-talets slut fram till i dag. I boken beskrivs de medel som författarna och översättarna använt vid olika tidpunkter för att skapa en talspråksillusion i texterna. Numera strävar man efter att framför allt i tidningsintervjuer låta intervjupersonens röst synas genom avvikelser från den övriga normenliga texten. Också inom övriga genrer finns ett behov att återge tal som en del av en text som i övrigt representerar standardspråk. Boken ger redaktörer och övriga intresserade en heltäckande bild av olika stilar och de medel som använts för att passa in i dem.

Anna Maria Gustafsson er nordisk sekretær i den svenske avdelingen i Institutet för de inhemska språken, Finland.

Færøerne

Marius Staksberg og Zakaris Svabo Hansen

Fellows-Jensen, Gillian: The elusive feminine personal names or When the saints go marching out. I: Zakaris Svabo Hansen et al. (red.), s. 69-81.

Sammenligning mellem kvindenavne fra middelalderen i England og på Færøerne.

Hansen, Zakaris Svabo, Jógvan í Lon Jacobsen, Tina K. Jacobsen, Kristin Magnussen og Turið Sigurðardóttir: Frá Sturlungu til West Venture. Heiðursrit til Anfinn Johansen á 60 ára degnum. Tórshavn: Fróðskapur.

Festskrift til navneforskeren Anfinnur Johansen på 60 års fødselsdagen.

Hansen, Zakaris Svabo: Overva og overvan. I: Zakaris Svabo Hansen et al. (red.), s. 117-119.

Semantisk behandling af verbet overva og de afledte substantiver overvan og overving.

Heycock, C., A. Sorace, Z. S. Hansen & F. Wilson: Acquisition in Variation (and Vice Versa): V-to-T in Faroese Children. I: *Language Acquisition*. 20, 1, s. 5-22.

Artikel om variation i ordstilling hos færøske børn.

Jacobsen, Jógvan í Lon: Málmaðurin í útvarpinum 3. Veturin 2010-2011 – og so nøkur hornaspørk. Tórshavn: Sprotin.

Grundstammen i bogen er korte artikler i forbindelse med et sprogprogram i færøsk radio, hvor forfatteren svarede på sprogspørgsmål fra lytterne.

Jacobsen, Jógvan í Lon: Navnapurisma í føroyskum. I: Zakaris Svabo Hansen et al. (red.), s. 120-130.

Om purisme i personnavnelovgivningen på Færøerne.

Louis-Jensen, Jonna: „Tað sær hvørt bern“. Om nogle af mælkebøttens danske og nordatlantiske navne. I: Zakaris Svabo Hansen et al. (red.), s. 177-181.

Petersen, Hjalmar P.: Vera og hava + høvuðssagnorð. Ein fyribilskanning. I: Zakaris Svabo Hansen et al. (red.), s. 195-206.

Redegørelse for en foreløbig undersøgelse af distributionen af *vera/hava* + hovedverbum.

Sigurðardóttir, Turið: Nøvn í føroyskum barnabókum, upprunaligum og týddum. I: Zakaris Svabo Hansen et al. (red.), s. 207-212.

Om proprier i færøske børnebøger – oprindelige og oversatte.

Simonsen, Marjun Arge: Fólkanøvn í staðarnøvnum – uppruni og aldur. I: Zakaris Svabo Hansen et al. (red.), s. 213-225.

Om personnavne i færøske stednavne, og et forsøg på at bestemme navnenes oprindelse og alder.

Simonsen, Marjun Arge: Málráð. Hvussu orða vit okkum á rættan hátt. Tórshavn: Sosialurin.

Bogen indeholder korte artikler med rådgivning om korrekt sprog, der har stået trykt i avisen Sosialurin.

Særheim, Inge: Støpilin **på Mykines og Ståbelen i Rogaland. Nokre namn som er laga til norr.** støpull **m.** I: Zakaris Svabo Hansen et al. (red.), s. 261-270.

Behandler norske, færøske og islandske stednavne, som har sin oprindelse i norrønt *støpull*.

Vikstrand, Per: Finns det förkristna sakrala ortnamn på Färöerna? I: Zakaris Svabo Hansen et al. (red.), s. 271-277.

Weyhe, Eivind: Skála, Skálabotnur og Skálafjørður. I: Zakaris Svabo Hansen et al. (red.), s. 278-284.

Om tre stednavne på Eysturoy på Færøerne.

Weyhe, Eivind: Skørðini í Norðradali. *Fróðskaparrit* 60: 73-82.

Artikel om visse stednavne.

Marius Staksberg er ansatt i sekretariatet i Færoyska málráðið, Færøylene.

Zakaris Svabo Hansen er formann i Færoyska málráðið, Færøylene.

Island

Jóhannes B. Sigtryggsson

Guðrún Kvaran. Glíman við orðin. Ritgerðasafn gefið út í tilefni sjötugsafmælis höfundar 21. júlí 2013. [‘Kampen ved ordene. Artikelsamling udgivet i anledning af forfatterens fødselsdag 21. juli 2013’]

Redaktion: Gunnlaugur Ingólfsson, Jóhannes B. Sigtryggsson, Þóra Björk Hjartardóttir. Rit 86. Stofnun Árna Magnússonar í íslenskum fræðum, Reykjavík. 374 s.

I dette festskrift (70-års-fødselsdag) er udvalgt af artikler fra Guðrún Kvaran's vidtfavnende forskning. Artiklerne handler bl.a. om islandsk leksikografi, personnavne, sproghistorie og etymologi.

Tilbrigði í íslenskri setningagerð. 1. Markmið, aðferðir og efni-viður. [Variation i islandsk syntaks. 1. Målsætninger, metoder og materiale] Red. Höskuldur Þráinsson, Ásgrímur Angantýsson og Einar Freyr Sigurðsson. Málvísindastofnun Háskóla Íslands, Reykjavík. 127 s.

Det er første bind af tre med resultater fra et stort forskningsprojekt (2005–2007) om variation i islandsk syntaks. Store undersøgelser blev lavede i forbindelse med projektet. I dette bind fortælles om målsætningene med projektet, metoderne og materialet.

Skíma. Mál gagn móðurmálskennara [‘Islandsklærerernes tidsskrift’]. Skíma kom i et nummer i 2013. Den kan hentes her: <http://www.modurmal.is/pdf/skima2013.pdf>.

Orðasafn í líffærafræði. I. Stoðkerfi. Enska – latína – íslenska. Red. Jóhann Heiðar Jóhannsson. Redaktion: Jóhann Heiðar Jóhannsson, Hannes Petersen, Ágústa Þorbergsdóttir. Stofnun Árna Magnússonar í íslenskum fræðum, Reykjavík. 56 s.

En terminologisk liste over anatomiske termer på engelsk, latin og islandsk. Også udgivet elektronisk på www.emma.is.

Tölvuorðasafn. 5. udg. [‘Ordbog over datalogi’]

En ny udgave af denne ordbog kom ud 2013. I 4. udgave af ordbogen var ca. 6500 termer og i 5. udgave er de ca. 3,8% flere. Ordbogen er nu kun udgivet elektronisk (www.sky.is).

Czarnecki, Przemyslaw. Słownik polsko-islandzki islandzko-polski. [‘Polsk-islandsk islandsk-polsk ordbog’] Konsulent: Jakub Chymkowski. Level trading, Kraków. 445 s.

Haukur Þorgeirsson. Hljóðkerfi og bragkerfi. Stoðhljóð, tónkvæði og önnur úrlausnarefni í íslenskri bragsögu ásamt útgáfu á

Rímum af Ormari Fraðmarssyni. Hugvísindastofnun Háskóla Íslands, Reykjavík. 409 s.

Dette er en doktorafhandling ved Islands Universitet (Háskóli Íslands) af Haukur Þorgeirsson. Afhandlingen handler om forbindelsen mellem det fonologiske system og versemålssystemet i islandsk sproghistorie.

Jóhannes B. Sigtryggsson er nordisk sekretær for Islands Sprognævn.

Norge

Ole Kristian Våge og Torbjørg Breivik

Abdi, Guleed: Den nye norsken. Slangordbok, Kagge forlag, 110 s.

Mange nye slangord og -uttrykk har opphav på austkanten i Oslo, og dei spreier seg ofte til andre delar av byen og resten av landet. Forfatteren, som er fødd og oppvaksen i Oslo aust, har samla dei 200 vanlegaste slangorda som vert nytta i det daglege miljøet han lever i. Oppslagsorda har bruksdøme og forklaring på opphavet.

Berg, Ivar, A. Dalen og K.F. Seim (red.): Runestudier. Festskrift til Jan Ragnar Hagland, Akademika forlag, 285 s.

Fleire av Jan Ragnar Haglands kollegaer heidrar han på 70-års dagen ved å gje ut ei samling artiklar om runer. Hagland er professor i norrønt språk ved NTNU og har ei lang forskar- og omsetjarkarriere bak seg.

Caprona, Yann de: Norsk etymologisk ordbok, Kagge forlag, 1920 s.

Ordboka inneheld det historiske opphavet til om lag 12 000 ord, og ho er delt inn i sju hovudtema: "Mennesker", "Naturen", "Dyrelivet", "Hverdagen", "Samfunnet", "Noen yrker og fag" og "Språk". Desse er igjen delt inn i 57 kapittel. Både den tematiske oppdelinga og det ledige språket gjer det underhaldande å lese frå ord til ord. Samstundes er innhaldet fagleg fundert, sjølv om forfatteren ikkje er filolog. Salssuksessen tyder på at boka er populær langt utover den tradisjonelle lesarkrinsen for språkbøker. Den imponerende ordboka vann Brageprisen i open klasse i 2013.

Botheim, Guri Sørungård og andre: Det fleirspråklege Noreg, Skald, 271 s.

I høve Språkåret 2013 har denne dokumentarboka kome ut. Gjennom 30 portrett gjev ho innblikk i korleis det fleirkulturelle Noreg har blitt eit fleirspråkleg land. Vi møter menneske med mellom anna eritreisk, fillipinsk, marokkansk og samisk-kvensk opphav, og får høyre korleis dei opplever det å vere fleirspråkleg i kvardagen. Boka er illustrert med bilete av Oddleiv Apneseth, og ho inneheld ein fagdel skriven av Øystein Vangnes.

Fretland, Jan Olav og A. Søyland: Rett og godt. Handbok i nynorskundervisning, Samlaget, 210 s.

Dette er ei praktisk bok for dei som skal undervise i nynorsk. Boka er todelt, og første del tek for seg undervisningsmetodikk. Metodikken er differensiert etter målgruppe, anten det er snakk om elevar som skal lære med utgangspunkt i eigen dialekt, eller bokmålselevar og framandspråklege som skal lære nynorsk. Andre del tek for seg den nye nynorsknormalen frå 2012. Boka gjev også råd om ordval på nynorsk.

Fridtun, Kristin: Nynorsk for dumskallar, Samlaget, 109 s.

Nynorsk for dumskallar er ei lettfatteleg og humoristisk innføring i nynorsk historie og kultur, ordtilfang, grammatikk og målføring. Boka passar både for dei som er nyfikne og vil lære litt om nynorsk, og for dei som må lære nynorsk mot sin eigen vilje.

Grepstad, Ottar: Historia om Ivar Aasen, Samlaget, 376 s.

Ivar Aasen reiste lenger enn Marco Polo, hadde bøker på minst 20 språk og hugsa alt han hadde lese éin gong. Han grunnla den nynorske diktekunsten og blei ein internasjonalt respektert språkforskar. Berre 22 år gamal bestemte Ivar Aasen seg for å finne det språket som alt var på folkemunne, og gjere dette til det norske språket. Offentleg sa han det først 16 år seinare og laga seg ein grenselaust ambisiøs arbeidsplan. Lesaren blir med Aasen landet rundt og livet igjennom i arbeidet hans med å oppfylle planen. Rekk han det han må få gjort?

Biografien får fram korleis Ivar Aasen vidareførte ideala frå opplysningstida. I boka til Ottar Grepstad trer lingvisten og språkforskaren fram som visjonær og like uthaldande som ein maratonløpar. Grepstad krydrar bio-

grafien med sitat frå Aasens dagbøker, og "Historia om Ivar Aasen" gjer han til eit levande menneske med sterke og svake sider.

Grepstad, Ottar: Språksansen. Kalender om all verdas språk, Samlaget, 403 s.

Kva skjedde i all verdas språk nett i dag? Kven har sagt noko lurt om språk og litteratur som du kan bruke på bloggen, i lunsjpausen eller i talen? Denne språkkalenderen har svar for alle dagar året igjennom. Kvar dag har ei forteljing frå språkhistoria verda rundt, opplysningar om ein forfattar som blei fødd eller døydde, og eit ord for dagen om språk. Det heile er lagt til rette med overblikk og språkglede, sans for detaljar og glimt i auga. Forteljingane rommar 5000 år og nærmare 300 språk. Som dagane går tek eit fargerikt bilete form, av språk som blir til og vinn posisjonar, endrar seg og tek uventa vendingar.

Språksansen er ei overraskande, lærerik og underhaldande bok for fleire enn den språkinteresserte.

Haugen, Odd Einar (red.): Handbok i norrøn filologi. 2. utgåve, Fagbokforlaget, 797 s.

Handbok i norrøn filologi har vore eit standardverk frå ho kom i 2004. Forfattarane er spesialistar på kvar sine felt. Boka gjev lesarane ei innføring i sentrale emne som handskrifter, teksttolking, runer, språkleg utvikling og variasjon, namn og nemne, dikt og soger og den gamle gudelæra. Boka er også ei rik kjelde til norsk og islandsk kultur- og teksthistorie i vid forstand, og gjev oss innblikk i den mangfaldige norrøne litteraturen og brytinga mellom den gamle åsatrua og kristendommen. Den tilgjengelege framstillingsforma og alle fargeillustrasjonane gjer boka til eit lesarvennleg møte med den norrøne kulturen.

Andreutgåva av *Handbok i norrøn filologi* er gjennomgåande revidert, og tre nye kapittel har kome til. Det finst supplerande materiale og oppdateringar på nettstaden nofihandbok.no.

Hellevik, Alf, M. Rauset og A. Søyland: Ordbok for grunnskulen, nynorsk, 5. utgåve, Samlaget, 172 s.

Ordboka er i samsvar med den nye norma for nynorsk som blei vedteken i 2012, og er laga for barnetrinnet i grunnskulen. Boka har over 11 000 oppslagsord med fulle bøyingsformer av substantiv, adjektiv og

verb. Ordboka inneheld elles ordforklaringar, ein minigrammatikk og øvingsoppgåver i korleis ein kan bruke ordboka.

Hårstad, Stian og Toril Opsahl: Språk i byen. Utviklingslinjer i urbane språkmiljøer i Norge, Fagbokforlaget, 200 s.

Språk i byen handler om hvordan språket utvikler seg i urbane språkmiljøer. Med utgangspunkt i studier som er gjort blant annet i Trondheim og Oslo, diskuterer forfatterne hvordan språket blir brukt til både å skape og å opprettholde sosiale kategorier og grenser. Med en nyere forståelse av begrepet stil belyser de hvordan ungdom i byen iscenesetter seg selv gjennom språklig praksis.

Jónsson, Sigurður, C. Laurén, J. Myking og H. Picht: Parallelspråk og domene, Novus forlag, 196 s.

Boka er skriven av nestorane innan fagspråkforskning i dei nordiske landa, og ho freistar å diskutere og klargjere to sentrale språkpolitiske omgrep i dag, parallelspråklegheit og domenedynamikk. Bakgrunnen er presset frå engelsk som bruksspråk på ei rekkje samfunnsområde, især i universitets- og høgskolesektoren. Forfattaren kjem også med framlegg om ulike tiltak mot engelsk språkbruk, som gradvis fortrengrer nasjonalspråka i Norden. Boka tek for seg situasjonen i dei nordiske landa og ser kritisk på fleire språkpolitiske initiativ.

Kjøll, Georg: Språkfilosofi, Fagbokforlaget, 207 s.

Forfattaren tek for seg grunnleggjande filosofiske problemstillingar i språkstudiar. Han nyttar klassiske teoriar for å diskutere ulike spørsmål knytte til kva meining eigentleg er for noko, og ikkje minst tilhøvet mellom meining og tenking. Kan tenking eksistere utan språk, og korleis kan vi vite at ord har same meining for ulike menneske? Boka freistar å nå eit publikum utanfor fagfeltet språkfilosofi.

Kleggetveit, Ingunn: Eg eller jeg? Masteroppgåve, Universitetet i Agder, 142 s.

Oppgåva er ein studie av forhold som kan påverke valet mellom nynorsk og bokmål hos ungdom i Indre-Agder. Analysen av argumentasjonen til informantane viser at elevar med bokmål som hovudmål i større grad har eit instrumentelt forhold til skriftspråket, medan nynorskinformantane har ei overvekt av argument som kan knytast til identitet, este-

tikk og språkpolitikk. Studien tyder på at drivkrefter som ligg bak valet mellom tradisjonelle og meir modererte talemålsformer i stor grad også gjer seg gjeldande i valet mellom nynorsk og bokmål.

Klyve, Arne og J. Severud: Ordbok for underklassen, Spartacus forlag, 125 s.

Boka er meint som hjelp til å forstå byråkratispråk og moteord i arbeidslivet. Boka har eit alfabetisk oversyn over nyord som invaderer kvardagen vår. Alle orda er henta frå aviser eller dokument skrivne for folk flest. Med skarpt blikk og humor analyserer forfattarane ord det går inflasjon i og viser at dei ofte er resultat av tanketomme hersketeknikkar.

Kulbrandstad, Lars Anders, O. Veka og B. Øygarden: Ordrett, ordliste for barnesteget, 3. utgåve, Fagbokforlaget, 265 s.

Ordlista er retta inn mot barnesteget i grunnskulen og har knappe 12 000 oppslagsord. Ordlista følgjer den nye norma for nynorsk som blei vedteken i 2012, og ordlista er godkjend av Språkrådet.

Kulbrandstad, Lars Anders, O. Veka og B. Øygarden: Ordrett, ordliste for ungdomssteget, 2. utgåve, Fagbokforlaget, 408 s.

Ordlista er retta inn mot ungdomssteget i grunnskulen og har omlag 18 000 oppslagsord. Ordlista følgjer den nye norma for nynorsk som blei vedteken i 2012, og ordlista er godkjend av Språkrådet.

Kulbrandstad, Lars Anders, O. Veka og B. Øygarden: Sidemålsordlista, nynorsk, 2. utgåve, Fagbokforlaget, 381 s.

Ordlista er retta inn mot heile grunnskulen og er fagleg og pedagogisk lagd til rette for undervisninga i nynorsk som sidemål. Ordlista følgjer den nye norma for nynorsk som blei vedteken i 2012.

Moi, Toril: Språk og oppmerksomhet, Aschehoug, 96 s.

I denne essayboka vil litteraturviteren Toril Moi undersøke korleis ein kan finne eit daglegspråk som gjer oss merksame på verda rundt oss. Merksemd som omgrep er knytt til både språk og det blikket vi nyttar når vi ser rundt oss. Boka er utforskande og byggjer både på fiksjon og filosofi. Moi nyttar mange eksempel frå litteraturen, frå Henrik Ibsen til Vigdis Hjort. Eit kapittel tek også for seg 22. juli-hendingane og rettssaka et-

terpå. Boka var først ute i ein serie frå forlaget i høve 100-årsjubileet for kvinneleg røysterett.

Nesse, Agnete: Innføring i norsk språkhistorie, Cappelen Damm, 232 s.

Denne innføringa i språkhistorie er hovudsakleg mynta på studentar på lågare grad, men òg andre som er interesserte i temaet. Boka tek for seg både talemål og skriftspråk i ulike historiske periodar, frå runeskrift til samtidsspråk. Ei rekkje autentiske tekstdøme er nytta for å syne korleis språket har utvikla seg. Vi møter språket til grupper som m.a. bønder, prestar, kjøpmenn og forfattarar, barn og eldre, kvinner og menn opp gjennom historia.

Nybø, Bjørg Nesje: Skriftkulturen i ei språkleg revitalisering, Universitet i Oslo, 262 s.

Dette er ei doktoravhandling som samanliknar framveksten av irsk og nynorsk skriftkultur. Nybø ser på korleis irsk, som var i ferd med å døy ut på 1800-talet, blei teke i bruk att gjennom rettsleg vern og ei oppbløming av litteratur. Parallellane til nynorsk er fleire, og Nybø analyserer både likskapar og ulikskapar i avhandlinga.

Nymo, Steinar: Ikkje så galle? Kolofon forlag, 426 s.

Ikkje så galle? Er ei alfabetisk etymologisk ordliste frå senjamålet og området ikring Senja. Boka tek lesaren med på ei oppdagingsferd i tid og rom frå moderne norsk og tilbake til "den ordskapande tanken". Boka inneheld 1776 oppslagsord og 438 kommentarar i form av sluttnotar.

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet. Band 11, Grønvik, Oddrun, Lars. S. Vikør og Dagfinn Worren (red.), Samlaget 2013, 1608 spalter.

Det ellefte bindet av i alt tolv bind av Norsk Ordbok har kommet. Det begynner på substantivet *stø* og går til adjektivet *tåvрут*. Som i tidlige bind har hver artikkel både ordforklaring og kildehenvisning. Innholdet ligger for øvrig søkbart på Norsk Ordboks hjemmeside. Ordboksprosjektet skal slutføres til grunnlovsjubileet i 2014 og har som uttalt målgruppe språkforskere, lekfolk og elever og lærere i skoleverket. Norsk Ordbok blir utarbeidet ved Institutt for lingvistiske og nordiske studium ved Universitetet i Oslo.

Ongstad, Sigmund (red.): Nordisk morsmålsdidaktikk, Novus forlag 2012, 332 s.

Boka er ei samling av artiklar som tek for seg morsmålsdidaktikk som forskingsfelt og som fagfelt i dag. Med andre ord gjev boka ei oppdatering på morsmålsdidaktikk, ikkje berre i Noreg, men også i andre nordiske land.

Schmidt, Tom (red.): Målblomar til Margit. Veneskrift til Margit Harsson på 70-årsdagen den 9. juni 2013, Novus forlag, 181 s.

Margit Harsson er førsteamanuensis ved Universitetet i Oslo, og i dette veneskriftet har fleire kollegaer skrive artiklar om hovudsakleg stad- og personnamn, men òg andre tema.

Sigmond, Ellen M.O., I. Bryhni og K. Jorde: Norsk geologisk ordbok, Akademika forlag, 496 s.

Ordboka inneheld over 10 000 oppslagsord innanfor geologi, ofte med etymologisk forklaring. Kvart oppslagsord er definert og har fått omsetjing til engelsk. Ifølgje forfattarene, som har lang fartstid i faget, er ordboka eit nyttig oppslagsverk for alle som arbeider med eller er interesserte i geologi.

Språknytt er tidsskriftet til Språkrådet, og i 2013 kom det ut fire nummer på til saman 136 sider. Språknytt er gratis, har eit opplag på 12 000 og er lagt ut elektronisk på www.sprakradet.no/spraknytt. Bladet har dei faste spaltene *Lesarspørsmål*, *Nyord*, *Kort*, *Med andre ord*, *Språkbrukaren* og *Historia bak*.

Nummer 1 har spørsmålet "Har du hjernefluer?" på framsida, og inne i bladet får vi ein artikkel om ei spesiell norsk-russisk ordbok laga av ein russisk filolog. Mange av orda er omsette til heilt ukjende eller feil norske ord, men ordboka er framleis i sal. Vidare er det ein artikkel om norsk scenespråk og eit intervju med tidlegare kulturminister Hadia Tajik. Nummeret har også artiklar om søking og om årets ord i 2012 (*nave*).

Nummer 2 handlar i stor grad om Ivar Aasen sidan det 5. august var 200 år sidan han blei fødd, og fleire skribentar tek for seg Aasens arbeid frå ulike vinklingar. I tillegg er det eit intervju med Per Egil Hegge, som har vore journalist, redaktør og korrespondent i Aftenposten. Han har skrive fleire bøker, mellom anna om språk og språkbruk.

Intervjuet i nummer 3 omhandlar spesielle norske ord sett utanfrå,

mellom anna *lakenskrekk*, ein artikkel er om striden som har vore om norsk scenespråk, og ein artikkel om det å tolke og bli tolka.

Det siste nummeret i 2013 opnar med eit intervju med Yann de Caprona, som har skrive Norsk etymologisk ordbok. 200-årsjubilantane Asbjørnsen og Moe blir markerte med artikkelen ”Språklig eventyr” der arbeidet deira blir sett med språklege augo. Ein artikkel omhandlar krydder i kvardagen, kvar nemningane kjem frå og kva namna eigentleg tyder.

Statsspråk – bladet for godt språk i staten – blir utgitt av Språkrådet og kommer ut med fire nummer i året. Bladet har et opplag på 17 500 og distribueres gratis til statsansatte og andre interesserte. Bladet kan også lastes ned som pdf-dokument fra www.sprakradet.no/statssprak. Statsspråk har fire sider og inneholder en hovedartikkel, et større oppslag med praktiske råd og tips (”På plakaten”) og dessuten notiser om relevante arrangementer, språktips og et dikt.

Tema for nummer 1 i 2013 var flerspråklig valginformasjon og en gjennomgang av hvordan en bruker ”i forhold til”. Nummer 2 hadde ”språkets gyldne middelvei” som tema, og omhandla bruk av fagspråk, stammespråk og krav til språk når en skal formidle informasjon til ”folk flest”. Nummer 3 tok for seg statlig språkmangfold gjennom en omtale av hvilke minoritetsspråk vi har i Norge. Det siste nummeret i 2013 presenterte historien om hvordan nynorsk ble statsspråk. Spalten ”På plakaten” har i dette nummeret tips og råd om hjelpemidler som skribenter kan bruke for å unngå språklige feil og feller når de skriver (ordbøker, nettsteder og liknende).

Stokkeland, Jostein og T. D. Bjerga: Engelsk–nynorsk / Nynorsk–engelsk ordbok. 2. utgåva 2013, Samlaget, 468 s.

Ordboka er utvida med nyord både på engelsk og norsk, og tilfanget er no oppe i 36 000 oppslagsord. Ordboka er i tråd med den nye rettskrivingsnormalen for nynorsk.

Stokkeland, Jostein og T. D. Bjerga: Engelsk–bokmål / Bokmål–engelsk ordbok. 2. utgåva 2013, Samlaget, 468 s.

Ordboka er utvida med nyord både på engelsk og norsk, og tilfanget er no oppe i 36 000 oppslagsord.

Termposten er eit elektronisk nyhendebrev frå Språkrådets terminologiteneste. Det kjem ut tre gonger i året og informerar om kva som skjer

på terminologifronten i Noreg og andre stader. Det inneheld også aktuelt stoff for alle som arbeider med eller interesserer seg for terminologi. Nyhendebrevet er gratis, og ein kan abonnere ved å gå til Språkrådets nettside www.sprakradet.no/termposten.

Vangsnes, Øystein A.: Språkleg toleranse i Noreg – Norge, for faen!, Samlaget, 101 s.

Forfattern set spørsmålsteikn ved statusen for språkmangfald i Noreg, og han tek opp motstanden mot nynorsk og samisk blant mange. Det vert ofte hevda at dialektane har ein høg status hos oss, men stemmer det? Øysten Vangsnes tek også opp motstanden mot at minoritetsspråklege born skal få tilbod om morsmålsopplæring. Boka slår eit slag for språkleg toleranse, både for kvar enkelt og for samfunnet som heilskap.

Vågslid, Torgils: Nytt om gamle namn, Kolofon forlag, 314 s.

Nytt om gamle namn ser på norske stadnamn med eit skarpt blikk på landskap og lokalitet ifølgje forfattern sjølv.

Aarønes, Lars: Skriv godt, J.M. Stenersens forlag AS, 185 s.

Skriv godt gjev tips og gode råd som inspirerer til å skrive betre. Boka omhandlar korleis du kan gjere ulike tekstar som jobbsøknader, CV, rapportar, presentasjonar o.l. betre. Du får også eit oversyn over dei vanlegaste språkfeila vi gjer, og korleis vi kan rette dei opp.

Ole Kristian Våge er seniorrådgiver i Språkrådet, Norge.

Torbjørge Breivik er nordisk sekretær og seniorrådgiver i Språkrådet, Norge.

Sápmi

Marko Marjomaa

Kejonen, Olle: 101 julevsáme verba = 101 lulesamiska verb. Vulkan, Stockholm. ISBN: 978-91-87669-83-5.

På svenska och lulesamiska. Korrektur och språkvård: Ann Christin Kuoljok Skoglund.

Dån tsamáda tjavvga = You whisper loudly.

Information om samiska språk. Text på engelska, sydsamiska, ume-samiska, pitesamiska, lulesamiska, nordsamiska, enaresamiska, skoltsamiska, kildinsamiska, tersamiska, akkalasamiska. Riddu riddu davvi álb-mogiid guovddáš, Samuelsberg.

http://riddu.no/sites/riddu.no/files/132114_pl_sprakhefte.pdf

Pope, Kirsten och Sárá, Máret: Eatnigiella : giellaoahpu váldogirji. [Grammatik, på nordsamisk]. Davvi girji, Kárásjohka. ISBN: 978-82-7374-923-9.

Perssen, Tor et. al.: Fievrrut = Fievro = Vuajastahkh. Davvi Girji, Kárásjohka. ISBN: 978-82-7374-707-5.

Fievrrut Fievro Vuajastahkh er en bok som er tilpasset for barn i begynneropplæringen, men denne boka passer også for barn som er språklig sent utviklet. Til hvert bilde er det skrevet et ord på nord-, lule-, sørsamisk og tegn-til-tale.

Skum, Marja-Kristin: Giellamolsun sámi boazodoalus. [Språkbyte inom renskötseln.] Sámi allaskuvla, Guovdageaidnu. ISBN: 978-82-8263-123-5, ib. s. 88-91 og s. 92-94.

Angéus-Kuoljok, Susanna och Tuolja, Karin: Giellaj hilá. [Lærebok i lulesamisk for voksne.] 2. reviderade uppl. Samernas utbildningscentrum, Jokkmokk.

„Go ealáska, de lea váttis dápmat”: davvisámegiela etnolingvist-talaš ceavzinnávccaid guorahallan guovtti gránnjágielddas Deanus ja Ohcejogas 2000-logu álggus. [Avhandling (ph.d.), sociolinguis-

tic dissertation, which examines the ethno-linguistic vitality of North Sámi in the two neighbouring municipalities of Deatnu/Tana in Norway and Ohcejohka/Utsjoki in Finland. UIt Norges arktiske universitet. ISBN: 978-82-8244-108-7.

Rahka, Sandra Nystø: Grammatihkalasj synonymija: Lånudallam bájkálasj kásusij ja bájkálasj adposisjåvnåj gaskan. [Grammatisk synonymi: växling mellan plats kasus och plats adpositioner]. Bacheloroppgave, Universitetet i Nordland, Bodø.

Juuso, Jane, Sparrok, Sylvia och Fjellgren, Patricia: Jag tar tillbaka mitt språk. Sametinget, Kiruna. ISBN: 978-91-977657-8-7.

Boka är en handbok för att ta tillbaka sitt språk. Den handlar om fenomenet "språkspärr" som kan vara en anledning till att personer inte talar ett språk fastän de har kunskapen.

Kintel, Lars Theodor: Julevsáme kausatijva værbbasuorgudimij syntávsalassj doajmma. [Lulesamisk kausativ verbavledningars syktaktisk funktion]. Bacheloroppgave, Universitetet i Nordland, Bodø.

Jupmelen Sjugnede. [Sørsamiske bibeltekster i utvalg]. Bibelselskapet, Oslo. ISBN 978-82-541-0871-0.

Ågård, Anita Dunfjell: Lilja stoerre åabpine sjædta. [Barnebok på sørsamisk]. Gielem nastedh, Snåsa. ISBN: 978-82-992976-4-6.

Lägesrapport, De samiska språken i Sverige 2012. Samiskt språkcentrum, Sametinget i Sverige.

<http://www.sametinget.se/54352>

Todal, Jon: Man ollu rievddada sámegeiela dillii Norggas? [Kvantitative endringer i den samiske språksituasjonen i Noreg, Nordsamisk og norsk tekst]. Sámi logut muitalit 6 - Guovdageaidnu : Sámi allaskuvla, cop. 2013 - Raporta / Sámi allaskuvla ; 1/2013. s. 17-56

http://www.regjeringen.no/upload/FAD/Vedlegg/SAMI/Samiske_tall_forteller_6.pdf

Jacobsen, Anna, Åsta Vangberg og Nora Marie Bransfjell: Mijjen gieles. [Sørsamisk grammatikk bok]. Red. Maja Dunfjeld. Sijti Jarnge, Hattfjelldal. ISBN: 978-82-92700-04-4.

Jåma, Ellinor Marita: Må bruke den samisken vi kan. [Intervju med sametingsrepresentant Ellinor Jåma om situasjonen for sørsamisk språk. Parallell norsk og sørsamisk tekst].

Eit annonsevedlegg for Språkåret 2013. Inger Johanne Sæterbakk. s. 14-15.

Kintel, Ingrid, Mikkelsen, Inga och Mikkelsen, Mathias: Rádnastallam ájnna njuolgadus le sámástit/ Rádnastallams eneste regel er å snakke samisk I: Bårjås - populærvitenskapelig tidsskrift. Artikkel (lulesamisk tekst med norsk sammendrag). Báhkko forlag, Drag.

Olthuis, Marja-Liisa, Kivelä, Suvi och Skutnabb-Kangas, Tove: Revitalising indigenous languages : how to recreate a lost generation. [The book tells the story of the Indigenous Aanaar Saami language (around 350 speakers) and cultural and language revitalisation in Finland]. I: Multilingual Matters, Bristol. ISBN: 9781847698889, ib.

Eikjok, Jorunn och Røe, Ola: Same i Russland. [Intervju med Semen Bolsjunov og Nina Kaneva fra Loparskaja på Kola], álliidLágádus, Kárášjohka.

Skåden, Asbjørg och Skåden, Sigbjørn (red.): Samiske stedsnavn i Skánik/Skånland. Skániid girjie, Evenskjer. ISBN: 978-82-91973-52-4.

Nystad, Arne, Solbakken, Alf A. och Boyne, Hans Petter ja earát: Sámi dovdosánit : sánit miela ja láhttema birra j.n.v. ja sánit mat leat jávkan dahje jávkamin sámegielas ; sámi-dáru sátnelista [Emosjonelle begrep på samisk, Nordsamisk-norsk og norsk-nordsamisk ordliste]. Forfatterne, Karasjok.

Svonni, Mikael: Sátnegirji : davvisámegiela-ruo agiela, ruo agiela-davvisámegiela. [Ordbok: Nordsamisk – svensk, svensk – nordsamisk ordbok]. álliidLágádus, Karasjok. ISBN: 978-82-8263-122-8.

Allt i Kartong – sydsamiska. TV-serie [Elektronisk resurs] :

<http://www.ur.se/Produkter/178327-Allti-Kartong-sydsamiska-Maskinen>

<http://www.ur.se/Produkter/178114-Allti-Kartong-sydsamiska-Regn>

<http://www.ur.se/Produkter/178323-Allti-Kartong-sydsamiska-Potatis>

<http://www.ur.se/Produkter/178325-Allti-Kartong-sydsamiska-Monster>

<http://www.ur.se/Produkter/178326-Allti-Kartong-sydsamiska-Manen>

<http://www.ur.se/Produkter/178324-Allti-Kartong-sydsamiska-Fart>

<http://www.ur.se/Produkter/178328-Allti-Kartong-sydsamiska-Fron>

<http://www.ur.se/Produkter/178115-Allti-Kartong-sydsamiska-Bajs>

<http://eatneme.no>

Rydving, Håkan: Words and varieties : lexical variation in Saami.

Suomalais-ugrilaisen seuran toimituksia - (bind-/heftenr. 269). Helsinki : Société Finno-Ougrienne. ISBN: 978-952-5667-49-3.

www.gielese.no

Eit Aajege-prosjekt for å laga eit språklæringsprogram for sørsamisk for nettbrett og mobiltelefonar, i samarbeid med UiT / Divvun. Nettside av Aajege i Röros samisk språk- og kompetansesenter.

Marko Marjomaa är projektleder vid Sáfá2-projekt/Sámi Giellagáldu.

Sverige

Andrea Daleflod-Jontegen

Alvehus, Johan: Skriva uppsats med kvalitativ metod: en handbok. Liber, 144 s.

Denna uppsatshandbok ger på ett brett och lättillgängligt sätt svar på frågor som "Hur går man tillväga för att få till uppsatsen så bra som möjligt?" och "Vad spelar metoden för roll i det hela?". Under det inledande arbetet med att skriva uppsats uppstår många frågor. Hur ska man veta att man har formulerat en relevant problemställning och att man valt rätt empiriskt material? Vad är en fallstudie och hur gör man observationer?

Och hur ska egentligen en akademisk text struktureras? Boken ger på ett översiktligt sätt svar på dessa frågor och många fler. Den kan användas både under en introduktionskurs, där man som student kanske för första gången stöter på frågor om metod, och ända upp till examensarbetet, där det både ges en överblick över olika befintliga metoder och visas på frågeställningar, som kan diskuteras på seminarier och i handledningsprocessen. Boken består av fyra huvuddelar: metod, skrivandet, fältarbetet och skrivbordsarbetet.

Bylin, Maria: Aspektuella hjälpverb i svenskan. Institutionen för svenska och flerspråkighet vid Stockholms universitet, 204 s. Ak. avh.

Denna avhandling behandlar några aspektuella hjälpverb med infinitivkomplement och tar även upp hur de har fått hjälpverbsegenskaper i svenskans historia. De undersökta verben är *bruka, tendera, börja, sluta, fortsätta och hota*. Avhandlingen ger en beskrivning av de sex hjälpverbens individuella historier, som tillsammans tecknar en övergripande bild av hjälpverbiseringsprocessen. Den ger också ett historiskt perspektiv på hjälpverbstesten och det hjälpverbstypiska i de egenskaper testen ska identifiera. På så sätt belyser den också gränslandet mellan verb och hjälpverb. Ett viktigt begrepp i detta avseende är skiktning, som tydliggör hur gränsen mellan verb och hjälpverb beror av samspelet mellan nyare och äldre användningar.

Carlsdotter Wallin, Ylva: Så skriver du för unga. Ordfront förlag, 212 s.

Handbok som riktar sig till personer som vill skriva skönlitterärt för unga läsare. Författaren och skrivläraren Ylva Carlsdotter Wallin ger en guide in i barnlitteraturens värld, delar med sig av skrivverktyg, inspiration och tips, hela vägen från manusidé tills det är dags att kontakta ett förlag. Hon ger konkreta råd kring allt från persongestaltning, dramaturgi och vanliga nybörjarfel till hur man hanterar skrivkramp. Här finns dessutom en mängd skrivövningar att utgå ifrån. I boken får vi också ta del av tankar och idéer från barnlitteraturforskare och framgångsrika författare som Martin Widmark (LasseMajas detektivbyrå), Lena och Olof Landström (Bu och Bä) och Jenny Jägerfeld (Här ligger jag och blöder).

Englund Hjalmarsson, Helena och Mattson, Anki: Juristens skrivhandbok. Jure, 132 s.

Handboken fungerar både som skrivhandledning och uppslagsbok för alla som skriver texter med juridisk anknytning eller juridiskt innehåll. Den ger de grundläggande reglerna för disposition och språklig stil i modernt juridiskt skrivande samt praktiska tips och konkreta exempel om allt från ordval till typografi. Ett utförligt kapitel med skrivregler kompletterar de allmänna skrivreglerna med sådant som är specifikt för juridiskt skrivande. Vissa vanliga texttyper och dokument behandlas i egna avsnitt och ger specifika råd för vad man ska tänka på när man skriver definitioner, domar, myndighetsbeslut, avtal, författningar och PM.

Från ett språk till ett annat. Om översättning och tolkning. Språkrådet och Norstedts, 268 s.

Boken är en lägesrapport där tolkar, översättare, forskare, utbildare och statliga tjänstemän belyser olika aspekter av översättning och tolkning. Ämnen som tas upp är bl.a. utbildning och arbetsmarknad, översättning i ett historiskt perspektiv, översättning och tolkning som forskningsområde, teckenspråkstolkning, terminologi och arbetet inom EU. Bland bidragen kan nämnas *Lars Wollin: Den svenska översättningskulturen – hur gammal är den?*, *Cecilia Wadensjö: Tolkutbildning i Sverige*, *Riina Heikkilä: Film- och nyhetsöversättning – likheter och skillnader*.

Gibbons, Pauline: Stärk språket, stärk lärandet: språk- och kunskapsutvecklande arbetssätt för och med andraspråkselever i klassrummet. 3 uppl. Hallgren & Fallgren, 230 s.

Sedan första svenska upplagan 2006 har denna bok blivit en av de viktigaste böckerna för alla som undervisar elevgrupper med varierande språkkunskaper. Denna tredje upplaga har kopplats till Lgr 11 och vänder sig därmed främst till lärare i grundskolan. Författaren utgår från aktuell forskning och ger konkreta modeller för hur man kan arbeta både språk- och kunskapsutvecklande i olika ämnen. Kunskap om samverkan mellan språk- och kunskapsutveckling blir allt viktigare. Den nya läroplanen ställer höga krav på elevers förmåga att tillägna sig och själva producera text i olika ämnen. Lgr 11 är synbart influerad av internationell forskning som betonar att ämne och språk går hand i hand och utvecklas parallellt. Gibbons framhåller fördelarna och möjligheterna med att arbeta ämnesövergripande med både första- och andraspråkselever i klassrummet och visar att en sådan undervisning är språkutvecklande för alla elever. Alla exempel samt uttalsavsnitten har så långt som möjligt anpassats till svenska för-

hållanden. På Nationellt centrum's webbplats www.andrasprak.su.se finns studiehandledningar till boken.

Hedlund, Anneli: Klarspråk lönar sig. 2 uppl. Språkrådet och Norstedts Juridik AB/Fritzes, 76 s.

Den nya utgåvan är uppdaterad och innehåller bland annat helt nya exempel på klarspråksarbete. Här finns råd om metoder och arbetssätt samt exempel på hur ett antal myndigheter och kommuner arbetar med klarspråksfrågor och organiserar det långsiktiga arbetet. Den svenska språklagen säger att det offentliga språket ska vara vårdat, enkelt och begripligt. Det innebär att offentliga organisationer måste se över sin språkanvändning och arbeta aktivt med klarspråk. Inte minst är det viktigt av rättssäkerhets- och demokratiska skäl. Några av frågorna som diskuteras i boken är: Vad har man att vinna på att arbeta med klarspråk och vad får det för konsekvenser om kommunikationen inte fungerar? Hur arbetar man med klarspråk och hur skapar man de bästa förutsättningarna? Boken fokuserar på klarspråksarbete i kommuner, landsting och statliga myndigheter, men råden och metoderna är givetvis lika användbara i privat verksamhet och vänder sig till alla som skriver i tjänsten.

Holmes, Philip och Ian, Hinchliffe: Swedish. A comprehensive grammar. 3 uppl. Routledge, 744 s.

Första upplagan belönades med Erik Wellanders språkvårdspris. Nu, i dess tredje upplaga har texten fått en omfattande uppdatering för att anpassas till nya språknormer och för att spegla svenska Språkrådets rekommendationer. Den nya upplagan erbjuder även en förbättrad layout, ett helt reviderat innehåll och en mer användarvänlig disposition. Betoningen ligger på svenska i den dagliga kommunikationen och bygger på modernt korpusmaterial, media och Internet för att tillhandahålla högfrekventa ord och uttryck. Boken inkluderar även en detaljerad behandling av grammatiska strukturer och ordklasser, många exempel från dagens svenska med särskild uppmärksamhet på de områden som skapar förvirring och svårigheter för elever, tydliga kontraster mellan svenska och engelska samt hela bokkapitel om ordbildning, ortografi och interpunktion.

Hur lång är en stund? Exakta Printing, 196 s.

Boken är ett samarbete mellan Språktidningen och Språkrådet och

består av drygt 300 frågor från Språktidningen, där Språkrådets experter svarar på läsarnas frågor. Boken är full av roliga och lärorika artiklar om slang, språkvård, dialekter, talesätt, lånord och mycket mer. Detta är Språkrådets och Språktidningens andra bok med språkfrågor. Språkfrågor ger inte bara vägledning i kniviga skrivsituationer, utan är även både allmänbildande och underhållande läsning.

Hyltenstam, Kenneth och Lindberg, Inger: Svenska som andraspråk: i forskning, undervisning och samhälle. Studentlitteratur, 868 s.

Andraspråksinläring och andraspråksanvändning är ett livligt och mångfacetterat internationellt forskningsområde, där även andraspråksundervisningens politiska, kulturella och sociala dimensioner har fått stor uppmärksamhet. Att språket spelar en avgörande roll för skolframgång, liksom för integration och delaktighet i samhället, är ett välkänt faktum. Därför är det av stor betydelse att all utbildning genomsyras av språklig medvetenhet och ett mångkulturellt förhållningssätt. I denna andra, omarbetade upplaga av boken presenteras ett brett och aktuellt urval studier samlade kring tre områden, som belyser andraspråksforskningens olika frågor och perspektiv: andraspråk och individ, andraspråk och samhälle samt undervisningspraktik i det flerspråkiga klassrummet. Boken fungerar som ett viktigt kunskapsunderlag och stöd för alla de lärare, skolledare och andra utbildningsansvariga som strävar efter att åstadkomma bättre förutsättningar för flerspråkig utveckling och framgångsrik andraspråksanvändning.

Janson, Tore: Germanerna – Myten, historien, språken. Norstedts, 238 s.

Begreppet germaner är laddat med rasbiologi och nazism, men vilka var de egentligen? Stora vilda kämpar med svärd och hjälmar och underliga gamla kläder eller rakryggade, raknästa och ljushåriga unga män i elegant uniform från 1940-talet? I en vindlande historia som löper över tvåtusen år visas det hur en berättelse från antiken förändrades och deformerades till en ren förfälskning av det förflutna. Finns det germaner nu, eller fanns de förr? Ser germaner ut på något särskilt sätt? Är svenskar germaner? Konstigt nog är det svårt att hitta klara svar på så enkla frågor. Forskaren Tore Janson berättar om vilka de ursprungliga germanerna var och hur de kom att bli symboler för händelser och idéer som har spelat en

stor roll och fortfarande är laddade med farliga spänningar. Vidare söker författaren beskriva vad som faktiskt hände i antiken, alltså om dem som ursprungligen kallades germaner. Deras historia och deras relationer till romarna har mycket att göra med oss i Norden och formade hela Europas utveckling ända fram till nutiden. Deras språk gav upphov till bland annat svenska, engelska och tyska. Vi i Sverige, liksom de flesta människor i norra Europa, använder ett språk som bygger på det germanerna talade i antiken. Men boken handlar också om hur vi som lever nu ska förhålla oss till vad som skedde förr med samhällen och med språk. Behöver vi veta det? Är det nyttigt eller onödigt eller farligt att se bakåt? Augustnominerad till årets svenska fackbok.

Jensen, Per och Ulleberg, Inger: Mellan orden: Kommunikation i praktiken. Studentlitteratur, 400 s.

Boken handlar om möten mellan yrkesutövare och brukare av samhällets tjänster och det som sker genom orden, och inte minst, det som sker mellan orden. Hur kan vi förstå kommunikation mellan människor? Kan kunskap om kommunikation bidra till att förbättra vår professionella praxis? Kan sådan kunskap bidra till att förändra fastlåsta situationer och se nya möjligheter? Professionalitet innebär att utveckla bra relationer och att prata med människor på bra och hjälpsamma sätt. I detta professionella arbete är samtal grundläggande – med patienter, klienter, barn och deras föräldrar och med kollegor och samarbetspartners. Del 1 introducerar ett grundläggande kommunikationsperspektiv och beskriver olika traditioner inom kommunikationsteori. Del 2 fördjupar sig i systemisk kommunikationsteori och begrepp som utvecklats av Gregory Bateson. Del 3 har fokus på professionell praxis. Genom hela boken vänder sig ofta författarna direkt till läsaren och inviterar till egna reflektioner. I första hand vänder sig boken till studenter inom olika grundutbildningar – sociologer och socialpedagoger, skötare och sjuksköterskor, lärare, förskollärare, fritidspedagoger och barnskötare, men boken är också aktuell för andra yrkesgrupper inom hälso- och sjukvården, socialtjänsten och skolan.

Kieliviesti är svenska Språkrådets finskspråkiga språkvårdstidskrift som sedan 2009 även innehåller artiklar på meänkieli. Kieliviesti utkom med tre nummer år 2013. Varje nummer har en huvudartikel och år 2013 handlade huvudartiklarna om språk och identitet i sverigefinsk litteratur,

om texterna i finsk rap-musik och om ord som uttrycker släktskap i finskan och svenskan. De flesta nummer innehåller även en presentation av nyord i finskan, frågor och svar-spalt samt en kortare ordlista. Några av artiklarna publiceras också på Språkrådets webbplats www.sprakradet.se.

Lagerholm, Per: Muntligt och skriftligt i Sverige och svenskan. Språk- och litteraturcentrum, Lunds universitet, 117 s.

I denna studie undersöks förhållandet mellan skriftspråket och talspråket i svenskan. Lagerholm använder sig av olika perspektiv inom tidigare svensk forskning på området och diskuterar bland annat begreppet muntlighet. Även tidigare forskning kring muntlighet och kultur presenteras. Till sist ges ingående förslag på hur vidare forskning skulle kunna bedrivas. Syftet med studien är att ge teoretiskt och metodiskt underlag till framtida forskning i frågan om tal- och skriftspråkets förhållande. Diskussionen gäller bland annat huruvida talet och skriften under det senaste århundradet har närmat sig varandra eller inte – den svenska språkvetenskapen har nämligen länge vidhållit att så är fallet. Slutsatsen i studien är att den diskussionen snarast blir alltmer meningslös. Det beror bland annat på att nya kommunikationsmedier gör att gränsen mellan tal och skrift upplöses. Lagerholm menar att man i stället bör ”fokusera på språkfunktioner och kommunikativa förutsättningar”.

Lyxell, Tommy: Se språket – barns tillgång till svenskt teckenspråk. Skrifter från Språkrådet, 62 s.

Tommy Lyxell, språkvårdare i teckenspråk på Språkrådet, beskriver i denna rapport barns tillgång till det svenska teckenspråket och vilka åtgärder som krävs för att stärka språkets ställning. Barns möjlighet att lära sig svenskt teckenspråk har minskat dramatiskt under det senaste decenniet. Antalet teckenspråkiga förskolor har blivit färre och antalet elever i specialskolan, som bedriver undervisning på svenskt teckenspråk, har nästan halverats. Risker är att fler förskolor och skolor läggs ner. Skolmyndigheten har känt till denna utveckling i årtal utan att effektiva motåtgärder har vidtagits. Viktigt är att utveckla föräldrastödet så att föräldrarna bättre kan förstå sina barns språkliga situation. Den svenska språklagen säger att samhället ska skydda och främja det svenska teckenspråket. Döva och hörselskadade barn ska ges möjligheter att lära sig, utveckla och använda det svenska teckenspråket. För att detta ska kunna ske måste samhället värna om de miljöer där det svenska teckenspråket används.

Det finns också en stor grupp teckenspråkiga barn som sällan uppmärksammas, nämligen hörande barn till döva föräldrar eller syskon till teckenspråkiga barn. Även dessa barn omfattas av språklagen. Hörande barn måste också ges tillträde till teckenspråkiga miljöer i förskola och skola.

Läsning! Svenskläraryöreningens årsskrift 2013. Red. Gustaf Skar och Michael Tengberg. Natur & Kultur. 2013. 224 s.

Boken ger en bred bild av fenomenet läsning och om skolans möjligheter att ge alla elever så goda förutsättningar som möjligt att utvecklas som läsare. Läsning är ett ämne med vad som verkar vara en permanent aktualitet, inte bara inom skolans värld utan även i den allmänna samhällsdebatten. En bra utgångspunkt för kategorisering av läsning kan vara att ställa sig frågor som vem läser, vad läses och i vilket syfte man läser. Svenskläraryöreningens årsskrift 2009 tog upp både skönlitteratur och sakprosa. I 2013 års tidsskrift har perspektivet på läsning breddats ytterligare, och de olika artiklarna behandlar läsningens plats och funktion i skolan, med utgångspunkt i både empiriska studier och i teoretiskt och kritiskt grundade resonemang. Boken erbjuder många olika infallsvinklar på frågor som relaterar till läsning i ett utbildningsperspektiv och är en angelägen läsning för blivande och redan verksamma svensklärare.

Marginaized and Ignored: National Minority Children's Struggle for Language Rights in Sweden 2013. Rapport från Svenska kyrkan.

Samiska barn och andra minoriteters barn får inte tillgång till sitt språk i skolan. Det innebär att Sverige bryter mot EU:s skydd för nationella minoritetsspråk, enligt en ny rapport. Det menar Svenska kyrkan i en nyskriven rapport om nationella minoritetsbarns rättigheter. Det största problemet är att skolans regelverk inte är anpassat till minoritetsbarnens situation. Möjligheterna att lära sig sitt nationella minoritetsspråk är i dag mycket små, vilket går emot de skyldigheter Sverige förbundit sig till genom att skriva under Europarådets konvention, enligt Svenska kyrkan. Rapporten har tagits fram tillsammans med ett antal minoritetsorganisationer och presenterades i samband med Europarådets granskningsbesök i Sverige den 17-19 december 2013. Under besöket undersökte en särskild kommitté hur Sverige lever upp till sina skyldigheter enligt stadgan.

Melin, Lars: Money talks: Om kronor, ören, tillgångar, cash flow, likvider, ltro, qe och annan bullshit. Morfem. 2013. 152 s.

Boken, som är av populärvetenskaplig karaktär, tar på ett upplysande och underhållande sätt upp svenska modeord inom det ekonomiska språket och ger en beskrivande bild av nutidens aktiemarknadsjargong, kris-terminologi, finansmarknadstyper och managementfloskler. Man får bland annat veta vad aktierådgivarna egentligen menar, hur politiker pratar pengar, hur man pratar bort ett dåligt resultat eller pratar upp en aktiekurs, vilka som är de senaste modeorden och vilka ord som har floskelvarning. Exempel på modeord som tas upp är: tillväxt, räntegap, cash flow, LTRO, QE. Boken ger inga direkta råd, utan är mer filosofisk. Det är inte en bok för den som jagar aktie- och fondtips, utan snarare för den som vill förkovra sig i språket kring våra pengar.

Nettelbladt, Ulrika och Salameh, Eva-Kristina – Språkutveckling och språkstörning hos barn. (Del 2). Pragmatik - teorier, utveckling och svårigheter. Studentlitteratur, 536 s.

Behovet av en bok på svenska som samlar olika forskningsperspektiv och ger en översikt av kommunikativa svårigheter hos barn har länge varit stort. Denna bok utgör andra delen i serien Språkutveckling och språkstörning hos barn. Boken är uppdelad i fyra block. Block I ger en historisk översikt av hur pragmatiska svårigheter, inklusive autism, beskrivits och väcker frågor om normalitet. Block II sätter beskrivningar av pragmatiska svårigheter i relation till grundläggande teorier om pragmatik. Block III redovisar forskning om pragmatisk utveckling hos barn med typisk utveckling. Teorier och begrepp inom barnspråksforskning med relevans för pragmatisk utveckling och svårigheter diskuteras också. Block IV behandlar pragmatiska svårigheter i relation till språkstörning och andra funktionsnedsättningar. I slutet presenteras en övergripande teoretisk modell för att förstå hur pragmatiska svårigheter kan uppstå. Blocket ger också pedagogiska och kliniska implikationer för intervention. Boken riktar sig främst till logopedstudenter, logopeder och specialpedagoger men kan även läsas av studenter i psykologi, pedagogik och språkvetenskapliga ämnen. Den kan också användas som fördjupningslitteratur i forskarutbildning inom dessa ämnen. Även andra yrkesgrupper som kommer i kontakt med barn med olika slag av pragmatiska svårigheter kan ha nytta av boken.

Sheikhi, Karin: Vägar till förståelse (andraspråkstalare i samtal med studie- och yrkesvägledare). Institutionen för svenska språket, Göteborgs universitet, 313 s. Ak. avh.

När vuxna andraspråkstalare går till studie- och yrkesvägledare är det inte kunskaperna i svenska som avgör hur lyckat samtalet blir. Denna avhandling visar att det snarare beror på bägge deltagarnas samarbetsvilja och lyhördhet. Författaren har undersökt videofilmade samtal mellan vuxna andraspråkstalare och studie- och yrkesvägledare. Materialets andraspråkstalare är relativt nyinflyttade i Sverige och har begränsade svenskkunskaper. Men de språkliga tillkortakommandena utför inget större hinder i samtalen: "Övergripande präglas samtalen av samarbete som i hög grad leder till att man kan förstå varandra, i synnerhet i fråga om begränsade frågor, så kallad lokal förståelse", säger författaren. Det är alltså ett gemensamt arbete för att uppnå förståelse och lyhördhet från förstaspråkstalarens sida som är avgörande för ett lyckat samtal, snarare än rena språkkunskaper. Resultaten pekar på vikten av förstaspråkstalarens ansvar att underlätta för andraspråkstalaren på olika sätt, till exempel genom att göra fler och längre pauser i samtalet och genom att ibland använda en mer vardaglig samtalsstil.

Språktidningen. Utgiven av Språktidningen AB.

Tidningen har under 2013 utkommit med 8 nummer. Alla nummer var på 84 sidor, utom ett specialnummer som var på 100 sidor. Språkrådet har medverkat bl.a. med avdelningen *Frågor och svar*. Andra stående inslag har varit spalter skrivna av Östen Dahl, professor i lingvistik, Catharina Grünbaum, språkvårdare och skribent, Olle Josephson, professor i nordiska språk samt Fredrik Lindström, programledare och språkvetare. Nummer 1 innehöll bl.a. Språkrådets årliga nyordsartikel och en artikel om retorik. I nummer 2 finns en artikel om att allt fler svenskar byter till ett unikt efternamn samt en intervju med författaren Åsa Larsson. I nummer 3 reds begreppet särskrivningar ut ordentligt och det finns även en artikel om baskiska. Nummer 4 var ett samarbete med tidsskriften *Modern psykologi* där språk och psykologi sammansvetsades i ett extra tjockt nummer. Nummer 5 handlar om hur vi uttalar ord fel och texten i våra populära snapsvisor analyseras. I nummer 6 berättar skribenten och författaren Göran Greider hur han undviker att få skrivkramp. Nummer 7 handlar om felsägningar och om svenska för yrkesutbildade invandrare. Nummer 8 innehåller en artikel med Sara Danius, ny medlem i Svenska Akade-

mien och vi får också läsa om musikälskarna som blev så inspirerade att de lärde sig svenska.

Statens offentliga utredningar (SOU): Begripliga beslut på migrationsområdet. 2013. 365 s.

Utredningens uppdrag har varit att kartlägga hur Migrationsverkets beslut är utformade samt hur de uppfattas av enskilda parter, överinstanser och andra myndigheter. I uppdraget har ingått att, mot bakgrund av kartläggningen, lämna förslag till hur utformningen av Migrationsverkets beslut kan utvecklas. Utredningen har dessutom haft till uppgift att ta ställning till om det finns anledning att, utöver det arbete som inletts, ytterligare utveckla hur migrationsdomstolarnas och Migrationsöverdomstolens domar och beslut utformas. Utgångspunkten vid denna bedömning har varit kartläggningen och förslagen beträffande Migrationsverkets beslut. Utredningen ska enligt direktiven presenteras på ett sådant sätt att resultaten kan ligga till grund för Migrationsverkets och Sveriges Domstolars fortsatta arbete med dessa frågor.

Strömquist, Siv: Skiljeteckensboken: skiljetecken, skrivtecken och typografiska grepp. Morfem. 2013. 189 s.

Utan skiljetecken, ingen text. Åtminstone inte någon särskilt lättläst text. Det är med skiljetecknen, skriftspråkets allra minsta byggstenar, som texter tar form, blir överskådliga och begripliga. Det kan vara svårt att hålla reda på reglerna för alla skilje- och skrivtecken. När måste man sätta ut kommatecken? Hur använder man egentligen semikolon? Ska det vara stor eller liten bokstav efter kolon? Är det någon skillnad på svenska och engelska citattecken? Ska det vara mellanslag före procenttecken? Var hittar man alla små krumelurer i ord som *crème fraîche*, *Barça*, *señor* och *koda*? Svaren finns i denna överskådliga och innehållsrika handbok som tar upp allt från punkt och komma till matematiska tecken, smilisar och hur man använder kursiveringar och andra typografiska grepp. Här finns också intressanta och mindre kända skiljetecken som interrobang och ironitecknet. Siv Strömquist är docent i nordiska språk vid Uppsala universitet och författare till flera välkända böcker om skrivande och språkriktighet.

Såld på engelska? Om språkval i reklam och marknadsföring. Red. Per-Åke Lindblom, Frank-Michael Kirsch och Arne Rubensson. Språkförsvaret, 179 s.

För första gången publiceras en bok på svenska, som analyserar engelskans roll i reklam och marknadsföring i Sverige och andra länder, där engelska inte är officiellt språk. Ett flertal skribenter, kulturpersonligheter och forskare från Sverige, Nederländerna, Norge och Tyskland belyser frågor om språk och marknadsföring från olika utgångspunkter. Några av de övergripande frågor som behandlas är: Finns det några vetenskapliga rön, som motiverar användningen av engelska i reklam och marknadsföring i länder, där engelska inte är huvudspråk eller officiellt språk? Kan det finnas andra motiv och i så fall, vilka? Vilka föreställningar och myter finns det om det engelska språket inom marknadsföring? Vilka konsekvenser har användningen av engelska i reklamen i Sverige för det svenska språkets ställning? Bland författarna kan nämnas *Kjell Albin Abrahamsson*: Houston, Houston! We got a problem!, *Jonathan Smith*: Copy, cats & copycats. En rundtur i reklamrymden, *Per-Åke Lindblom*: Varför används engelska i reklamen?

Tandefeldt, Marika: Prima vara. Språk- och stilhistoriska studier i finlandssvenska och svenska varuhusannonser under 1900-talet. Svenska litteratursällskapet i Finland, 221 s.

Vad säger språket i tidningsannonser om språkbruket och språkets utveckling? I fem delstudier analyseras svensk språkutveckling under 1900-talet utifrån tidningsannonser som varuhuset Stockmann i Helsingfors och Nordiska Kompaniet i Stockholm har publicerat. Hur förändras språket i tidningsannonserna jämfört med den allmänna språkutvecklingen? Förändras svenskan i olika takt i Sverige och Finland? Studierna visar bl.a. hur rättstavningsreformen 1906 långsamt börjar tillämpas i annonspråket, hur verben förlorar sin pluralisändelse och när varuhuset börjar tilltala sina kunder med "du". Också balansen mellan olika ordklasser och användningen av adjektiv ägnas varsin studie.

Wedin, Åsa och Hedman, Christina: Flerspråkighet, litteracitet och multimodalitet. Studentlitteratur. 2013. 234 s.

Många människor, såväl barn som tonåringar och vuxna, lever i flerspråkiga sammanhang som skiljer sig stort från den enspråkiga miljö som ofta tas för given i skola och vuxenutbildning. Där har man i stället många

gångerna haft svårt att förhålla sig till flerspråkighet och ofta utgått ifrån ett bristperspektiv. Detta gäller inte minst skriftspråksundervisningen, där elevernas egna erfarenheter av en flexibel flerspråkighet utanför klassrummet alltför sällan tas tillvara. I den här antologin presenterar ett antal forskare exempel på skriftspråkliga erfarenheter, både flerspråkiga och multimodala, som sällan uppmärksammas i skolan. I sina studier har de sett en mängd goda exempel på hur lärare arbetar i klassrummet för att gynna skriftspråksutvecklingen hos flerspråkiga elever – deras multilitteracitet. En sådan medvetenhet hos läraren kan t.ex. handla om allt från att ha kunskap om effektiva läsförståelsestrategier till att förstå hur multimodalitet i läroböcker och multimodalt textskapande i klassrummet stärker lärande och flerspråkig kompetens. Boken vänder sig främst till blivande och verksamma lärare i skola och vuxenutbildning.

Andrea Daleflod-Jontegen er praktikant i Språkrådet, Sverige.

Adresser språknemndene 2014

Nordisk språkkoordinasjon,
www.fnfnorden.org/kontakt-os
Hulda Zober Holm
Foreningerne Nordens Forbund
Malmøgade 3
DK-2100 København Ø
Danmark

Dansk Sprognævn, www.dsn.dk
Worsaaesvej 19, 4. sal
DK-1972 Fredriksberg C
Danmark

Føroyska málnevndin, www.fmn.fo
Málstovan
V.U. Hammershaimbs gøtu 16
FO-100 Tórshavn
Færøyene

Institutet för de inhemska språken, www.sprakinstitutet.fi
Berggatan 24
FI-00100 Helsingfors
Finland

Stofnun Árna Magnússonar í íslenskum fræðum, www.arnastofnun.is
Skrifstofa Árnagarði
Við Suðurgötu
IS-101 Reykjavík
Island

Oqaasileriffik www.oqaasileriffik.gh
Postboks 980
DK-3900 Nuuk
Grønland

Språkrådet www.sprakradet.no
Postboks 8107 Dep
NO-0032 Oslo
Norge

Språkrådet,
www.sprakochfolkminnen.se/sprakradet
Box 20057
SE-104 60 Stockholm
Sverige

Sámi giellagáldu www.giella.org
Sajos
FI-99 870 Anár
Finland



norden

Nettverket for
språknemndene i Norden

Språk i Norden er et årsskrift som språknemndene i Norden publiserer. Aktuelle språklige temaer omtales her, og presentasjonene og diskusjonene på det årlige nordiske språkmøtet er utgangspunktet for publikasjonen. I dette nummeret er temaet de nordiske språkenes framtid i informasjonsteknologien. De fleste artiklene er baserte på innlegg som ble holdt på det nordiske språkmøtet Akureyri i august 2013, men noen er også inviterte artikler, som er relaterte til temaet. Siste delen er en oversikt over språklitteratur utgitt i 2013 i de nordiske landene.

Språk i Norden
Språkrådet,
Postboks 8107 Dep
NO-0032 Oslo
Telefon: +47 22 54 19 61 / +47 926 18 874
E-post: Torbjorg.Breivik@sprakradet.no