

tidsskrift *sb.*, -et, -er, i sms.  
 tidsskrift- el. tidsskrifts-, fx  
 tidsskrift(s)artikel.

tidsspild *sb.*, -et el. tids-

spilde *sb.*, -t.

# Nyt fra Sprognævnet

spørgsmål, *bf.* pl. ene

tidsstemple *vb.*, -ede.

tidsstempling *sb.*, -en, -er.

## Den der skriver *d* i *mand*

Af Eva Skafte Jensen

Ortografi er disciplinen om ret skrift, altså retskrivning. Den rette skrift er ikke given, men funderet i et sæt af konventioner vi igennem århundrederne har bestemt os for, typisk uden først at holde møder om dem. Det er ikke biologisk nødvendigt at vi skriver ordene på bestemte måder, men jo vældig praktisk, fordi man så ikke behøver at skulle opfinde skrivemåden forfra hver gang, og fordi man antagelig lettere kan afkode det andre skriver, hvis det skrives på samme måde hver gang. I visse skriftsprog ligger retskrivningen meget tæt på udtalen, typisk nævnes italiensk og finsk som eksempler på dette, i andre sprog er der en større afstand mellem udtale og ortografi, her er de oftest nævnte eksempler engelsk og dansk.

I dansk er der nemlig som bekendt ikke et ligefremt forhold mellem udtale og de rette skrivemåder. *Ost* skrives *o-s-t* skønt vokalen i standarddansk snarere udtales som en *å*-lyd, og den sidste konsonant snarere som en *d*-lyd.

Ikke desto mindre er der skrivemåder der på subtil vis viser noget om udtalen, skønt indirekte. Skrivemåden for *mand* (*m-a-n-d*) er et godt eksempel. Det skal denne artikel handle om.

### Oprindelige og uoprindelige stumme bogstaver

Det sidste bogstav i *mand*, altså *d*, er uoprindeligt. Det har aldrig været udtalt.

En del ord i dansk har skrivemåder med stumme bogstaver, hvor disse stumme bogstaver ikke altid har været stumme. De stumme *h*'er foran *v* og *j* i ord som *hvad* og *hjem* har engang været udtalt, men er i løbet af historien blevet svagere og svagere udtalt for tilsidst helt at forsvinde i udtalen. I vores nabo-sprog er det ikke nødvendigvis gået sådan med udtalen af *h*. På svensk hedder *hjem* således *hem*; her er udtalen af *h* bibeholdt (mens udtalen af *j* er svundet). I visse norske dialekter er udtalen af *h*'et faktisk blevet styrket og blevet til en *k*-lyd, *kva* betyder *hvad*. >

De stumme bogstaver i dansk kan altså sagtens afspejle en gammel udtale, men det er slet ikke sikkert de gør det. I stedet kan andre faktorer have gjort at de er til stede.

I middelalderen havde man ikke en centralt fastlagt ortografi, men skrev i overensstemmelse med den praksis man havde, der hvor man var. Langt de fleste håndskrifter blev skrevet i klostrene rundt omkring - det var der de skriftkyndige befandt sig - og det menes at de enkelte klostre på visse områder har haft lokale konventioner for hvordan man skulle skrive, og at de enkelte skrivere også kan have haft foretrukne måder at skrive på. I det 20. århundrede var det en yndet disciplin blandt forskere i gammeldansk at stedfæste de gamle håndskrifter til forskellige steder i landet, blandt andet gennem studierne af lokale skriftkonventioner.

Manglen på en fastlagt ortografi koblet med udviklingen af udtalen resulterede blandt andet i at man vænnede sig til at visse bogstaver ikke behøvede at være begrundet i en udtale for at være med. De stumme *h*'er er et godt eksempel på dette. Man vænnede sig til at de var der, og skrev dem med i de ord hvor de skulle være og 'altid' havde været. Nogle gange skrev man dem dog også ind i andre ord; der er mange eksempler på at man *føjede h*'er til i ord hvor der aldrig havde været udtalt et *h*, for eksempel *hoc* for *og* og *hvth* for *ud* (se fx Brøndum-Nielsen 1932: 307). Desuden har man mange eksempler på at de stumme bogstaver blev placeret anderledes i rækkefølgen af bogstaver, enten før eller efter det sted de stod, dengang de afspejlede en udtale, fx *ihona* i stedet for *hiona* (som her betyder *tyende*).

Studierne i udtale i 1300-, 1400- og 1500-tallet bygger blandt andet på disse

uoprindeligt stumme bogstaver og på de omvendte skrivemåder. Når man i 1400-tallet i mange håndskrifter ser omvendte skrivemåder og stumme bogstaver i ord som ikke sproghistorisk set bør have dem, antager man at der er sket udviklinger i udtalen. Relevant for skrivemåden af *mand* er at man i 1300- og 1400-tallet tilsyneladende var holdt op med at udtale *d* når det kom lige efter *n* eller *l*. Dette er netop noget man slutter ud fra omvendte skrivemåder og brugen af uoprindelige *d*'er.

### Hvornår er *man* en *mand* i gammeldansk?

I gammeldansk blev ordet *mand* skrevet *m-an*, altså ligesom man nu til dags skriver pronomenet *man*. Der var ikke noget *d*, hverken i udtalen eller i skriften.

Når man derfor møder *man* i gammeldanske tekster, kan man ikke på skrivemåden se om det skal læses som moderne *man* eller moderne *mand*. I gammeldansk havde man heller ikke ubestemte artikler (*en/et*), og derfor kan man ikke af formen alene afgøre om man i eksemplet nedenfor skal oversætte det første *man* til *en mand* eller til *man*, jf. de to oversættelser a. og b.:

Brytær man up dør annærs manz

a. bryder en mand en anden mands dør op

b. bryder man en anden mands dør op

Nogle gange kan man få hjælp fra den større kontekst. I visse tilfælde efterfølges *man* af *han* i en senere sætning, ergo skal *man* oversættes til *en mand*. I andre tilfælde vil *man* blive fulgt op af flere *man*'er, i disse tilfælde er der snarere tale om pronominalbrug. I en del

tilfælde må man bare leve med uklarheden, og det går såmænd også ganske fint. Se bare næste eksempel:

Far man kunu

Det er ikke afgørende om man skal oversætte dette til *får en mand en kone* eller *får man en kone*. Mangel på entydighed behøver ikke være et problem.

### Stødet

I moderne dansk angiver det stumme *d* efter *n* (og *l*, jf. *vild*, *kold*) at der er stød på *n*'et i ord som *hund*, *hånd*, *land*, (*Øre*)*sund*, *kind*, *mund*, *tand* og i *mand*. Men det er kun i de fire første af disse ord at der engang har været en *d*-udtale; i de efterfølgende ord inkl. *mand* er *d*'et uoprindeligt og afspejler ikke en gammel udtale med *d*.

Det vides ikke nøjagtigt hvornår dansk udviklede stød. Ifølge fonetikerens Hans Basbøll (2008: 150) findes der en svensk beskrivelse af dansk udtale fra slutningen af 1500-tallet, hvor man får at vide at det særlige ved os danskere er at vi vender ordene et par ekstra gange i struben og så først derpå klemmer dem ud - som om vi hoster. Måske er det en

Betegnelsen *gammeldansk* dækker traditionelt over perioderne *olddansk* (også kaldet *runedansk*), *ældre middeldansk* og  *yngre middeldansk*

(URNORDISK)	inden 800
OLDDANSK/RUNEDANSK	800-1100
ÆLDRE MIDDELDANSK	1100-1350
YNGRE MIDDELDANSK	1350-1525
(ÆLDRE NYDANSK)	1525-1700
(YNGRE NYDANSK)	1700-1950
(MODERNE DANSK)	efter 1950



Eva Skafte Jensen (f. 1966)  
er seniorforsker i Dansk  
Sprognævn.

beskrivelse af stødet. Hvis det er tilfældet, har stødet været etableret og vidt brugt i 1500-tallet. Udviklingen må altså have fundet sted tidligere end det, men hvor meget tidligere, vides ikke. Der er hypoteser om at stødet går helt tilbage til 1100-tallet, se fx Skautrup 1944: 256 ff.

Det er fristende at se en kobling mellem udviklingen af stødet og den udtaleudvikling hvorved visse *d*'er blev stumme, men det vil være meget svært at bevise, og i virkeligheden er det faktiske forløb nok mere komplekst end som så.

Uanset hvordan udviklingen er gået, kan man i senmiddelalderlige og efterreformatoriske skrifter se vældig mange uoprindelige stumme *d*'er, også i ord uden stød på *n*, *hand* for *han* er for eksempel forholdsvis hyppigt. Uoprindeligt *d* i *kunde*, *skulde* og *vilde* er også mere og mere almindeligt i denne periode, og som bekendt etablerer disse sidstnævnte skrivemåder sig som den rette måde at skrive datidsformerne på, indtil de igen bliver erstattet i 1948 af de oprindelige former uden *d*: *kunne*, *skulle*, *ville*. I argumentationen for denne ændring henviste man netop til at de stumme *d*'er i disse ord ikke afspejlede en ældre udtale.

Det stumme *d* i *mand* er altså uoprindeligt, det afspejler ikke og har aldrig afspejlet en udtale med *d*. Til gengæld ville det være >

forkert at sige at det slet ikke afspejler en udtale overhovedet. Det angiver jo netop at der er stød på *n*, og på den måde kan man i skriften holde *mand* og *man* adskilt (og også *hund* og *hun*). Udtaleforskellen har bare ikke noget med lyden *d* at gøre. Der er i stedet tale om en ortografisk konvention der bruges til at vise stødet på *n*.

#### Henvisninger

- Johs. Brøndum-Nielsen (1932). *Gammeldansk Grammatik*, bd. II. Kbh.: J.H. Schultz Forlag.
- Hans Basbøll (2008). Stød, diachrony and the non-stød model. *Nowele* 54-55. s. 147-189.
- Peter Skautrup (1944). *Det danske Sprogs Historie*, bd. I. Kbh.: Gyldendal & Nordisk Forlag.

## Flygtning eller migrant – vær opmærksom på dit ordvalg

Af Sabine Kirchmeier

De mange mennesker som kommer ind over grænserne til Europa i denne tid, har skabt debat ikke blot om hvordan vi skal forholde os til dem, men også om hvilke ord vi skal bruge når vi omtaler dem. Det gælder især ordene *flygtning* og *migrant*.

Det er ikke nogen ny situation at verden forandrer sig, og at vi forsøger at finde ord til at beskrive denne forandring. Men ord er ikke kun neutrale betegnelser for objektive forhold, de er også med til at give udtryk for hvordan vi forholder os til verden omkring os. Somme tider finder vi nye ord, andre gange giver vi eksisterende ord et nyt indhold

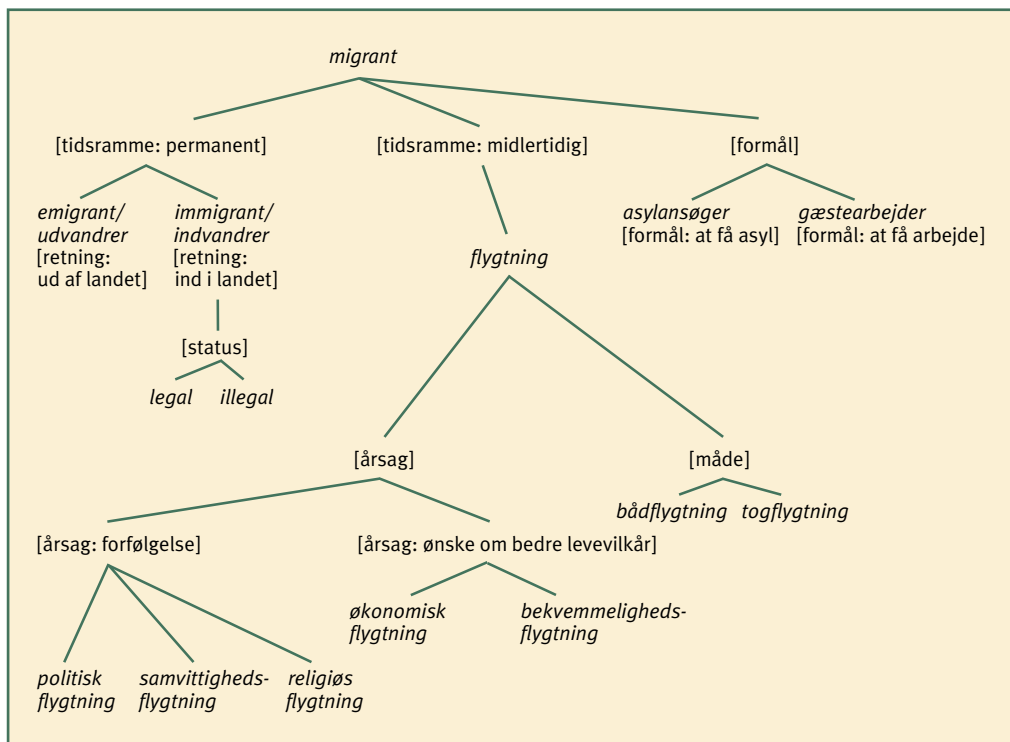
Lige nu er betegnelsen *migrant* under forandring. Ordet er forholdsvis nyt. Vi kender det fra ca. 1975, hvor det blev introduceret som et nyt ord der dækker både emigranter og immigranter<sup>1</sup>. I Den Danske Ordbog er en

*migrant* beskrevet som en 'person der flytter fra ét sted til et andet, fx for at søge arbejde og ofte over landegrænser'.

I fagkredse bliver *migrant* brugt som en fællesbetegnelse der dækker over alle mennesker der bevæger sig over grænserne for at tage korterevarende eller permanent ophold i et andet land uanset hvad årsagen til dette måtte være. Ordene dækker forskellige perspektiver: Flygtninge og emigranter er mennesker som er på vej væk fra et land, asylansøgere og immigranter er mennesker som ønsker at tage ophold i et nyt land.

Der er mange forskellige begreber der hænger sammen med ordet *migrant*, fx *emigrant*, *immigrant*, *indvandrere*, *udvandrere*, *flygtning*, *politisk flygtning*, *økonomisk flygtning*, *be-kvemmelighedsflygtning*, *bådflygtning*, *tog-flygtning*, *asylansøger* og *gæstearbejder*. Man kan skabe et overblik over relationerne imellem dem i nedenstående diagram:

1 Jf. *Nye Ord i Dansk 1955 til i Dag* ([www.dsn.dk/noid](http://www.dsn.dk/noid)).



*Migrant* er overbegrebet over dem alle, og derefter specificerer de forskellige ord forskellige typer af migranter som kan opdeles efter forskellige kriterier. Af diagrammet kan man se at de tidligere meget brugte ord *immigrant* og *emigrant* er synonyme med ordene *indvandrer* og *udvandrer*, og at de netop adskiller sig fra *flygtning* ved at de typisk sigter mod et mere permanent ophold i et andet land. Derimod ligger det i ordet *flygtning* at man som udgangspunkt ønsker et midlertidigt ophold for derefter at kunne vende tilbage til hjemlandet. Et andet perspektiv på migrationen kan være formålet med at rejse til et andet land. Det kan fx være at søge asyl eller at søge arbejde. Begrebet *flygtning* kan opdeles efter fx årsagen til flugten, fx politiske, religiøse eller andre forhold, og måden der flygtes på, fx med båd eller med tog.

Sådanne skematiske fremstillinger kan være nyttige til at få overblik over hvordan ordene og deres indhold forholder sig til hinanden, og hvilke betydningselementer der indgår. Det ville således sprogligt set være ganske logisk at kalde de mennesker som for tiden bevæger sig ind i Europa, for migranter så længe de er undervejs uanset på hvilken måde, af hvilken årsag, med hvilket formål og for hvilket tidsrum, for man ved ikke på forhånd hvad grunden til deres rejse er, og man kan ikke fastslå deres status før deres sag har været behandlet af myndighederne.

Men sådan er det ikke i den aktuelle debat. *Migrant* bruges nu ofte om en person som har forladt sit hjem for at opnå en bedre levestandard. En migrant betragtes ikke som forfulgt og har derfor ikke krav på >

beskyttelse. Ordet kan i den aktuelle debat siges at være blevet en samlebetegnelse for økonomiske flygtninge, bekvemmeligheds-flygtninge eller illegale immigranter og står nu som en slags modsætning til den gruppe af flygtninge som har krav på beskyttelse.

Det går så at sige på tværs af den hidtidige forståelse af begrebet og er pludselig svært at passe ind i diagrammet. Dermed kan ordet ikke længere uden risiko for misforståelser bruges som overbegreb for alle grupper der bevæger sig over grænserne.

Denne udvikling ses ikke kun i Danmark, men også i fx USA og England. Det er ikke utænkeligt at den danske debat er påvirket af den udenlandske debat om ordet *migrant*.

Vi har derfor ikke nogen entydig fællesbetegnelse for de mennesker der bevæger sig igennem Europa. Uanset om man vælger ordet *flygtning* eller *migrant*, risikerer man med sin sprogbrug at tage stilling til personens formål med rejsen og retten til at opholde sig her i landet.

I Sverige bruges ordet *migrant* som neutralt overbegreb. Svensk Røde Kors skriver fx

”Migration innebär att människor av olika anledningar rör sig över landsgränser. Frivilligt eller ofrivilligt”.

Den svenske udlændingestyrelse hedder *Migrationsvärdet*. Så her fungerer ordet *migrant* som et overbegreb. I Norge mener Språkrådet (det norske sprognævn) at ordet *flyktning* bør kunne bruges:

”Ordet *flyktning* tyder ifølge Nynorskordboka «person som er på flukt; person som har flykta frå heimstaden, landet». Ordet i allmennspråket viser såleis til ein som har rømt



Sabine Kirchmeier-Andersen (f. 1955) er direktør for Dansk Sprognævn.

frå noko, helst i fare. Dermed bør ein kunne bruke ordet om alle dei som no er på flukt frå Syria og landa rundt.”

Språkrådet i Norge påpeger at ordet *flyktning* udmærket kan bruges og bliver brugt i en mere rummelig betydning end den der fremgår af FN's flygtningekonvention, som definerer en flygtning som:

”en person der som følge af velbegrunnet frygt for forfølgelse på grund af sin race, religion, nationalitet, tilhørsforhold til en særlig social gruppe eller sine politiske anskuelser befinder sig uden for det land, i hvilket han har statsborgerret, og som ikke er i stand til - eller på grund af velbegrunnet frygt ikke ønsker - at søge dette lands beskyttelse”

Også i dansk bruges ordet *flygtning* i en bredere betydning end i flygtningekonventionen, og ligesom i norsk findes der en mere almen betydning af *flygtning*. Den Danske Ordbog har som betydning 2:

”person der skynder sig væk for at undslippe nogen el. for at komme væk fra noget ubehageligt el. farligt; person på flugt”.

Flere politikere og medier har valgt at bruge vendingen *flygtninge* og *migranter* som side-

ordnede begreber. Det at de to ord nu bruges sideordnet, forstærker yderligere tendensen til at ordet *migrant* ikke længere kan bruges som øverste fællesbetegnelse eller overbegreb. Diagrammet bryder sammen, og der opstår forvirring.

Ser man på hvilket af de to ord der bruges hyppigst, fx i nævnets samling af danske dagblade fra 2009 til i dag, har *flygtning* en klar førerposition. Ordet *flygtning* (og dets

bøjede former) forekommer 11.929 gange mens ordet *migrant* (og dets bøjede former) kun forekommer 1058 gange i hele perioden. *Flygtning* er med andre ord 11 gange hyppigere end ordet *migrant*. I 2014 var *flygtning* 15 gange hyppigere. Ser man på første halvdel af 2015, forekommer *flygtning* 4053 gange, mens *migrant* forekommer 502 gange svarende til at *flygtning* nu kun er 8 gange hyppigere end *migrant*.

	i alt 2009-2015		2014		2015	
migrant	1058	8%	189	6%	502	11%
flygtning	11.929	92%	2865	94%	4053	89%
i alt	12.987		3054		4555	

Statistikken viser at *migrant* indtil for nylig har været ganske sjældent brugt og har været ukendt for de fleste, mens *flygtning* har været det hyppigst anvendte udtryk. Det er formentlig en af årsagerne til at det er gået så hurtigt med nyfortolkningen af ordet *migrant*. For mange har det i forvejen været uklart hvad ordet egentlig dækker.

Tallene for 2015 skal imidlertid tages med et stort forbehold, idet en stor del af forekomsterne i 2015 ikke afspejler hvilke ord der faktisk bliver brugt, men snarere den aktuelle diskussion om hvilke ord der kan bruges, hvilket denne tekst jo også bidrager til.

Man kan læse mere i nedenstående artikler, hvor Sprognævnets medarbejdere Anita Ågerup Jervelund og Jørgen Nørby Jensen forklarer den udvikling der lige nu sker med ordet *migrant*:

<http://videnskab.dk/sporg-videnskab/en/migranter-og-flygtninge-hvad-er-forskellen>

<http://politiken.dk/kultur/medier/ECE2850215/flygtningekrise-aendrer-danske-ords-betydning/>

Om den udenlandske debat kan man læse i denne artikel fra BBC:

<http://www.bbc.com/news/magazine-34061097>

Om ordene *migrant* og *flygtning* i Norge:

<http://www.sprakradet.no/Vi-og-vart/hva-skjer/Aktuelt-ord/flyktning/>

# Hverdagens sprogteknologi nu og i fremtiden

Af Philip Diderichsen, Peter Juel Henriksen,  
Sabine Kirchmeier og Bolette Sandford Pedersen

I de fire foregående artikler er vi dykket ned i nogle eksempler på hverdagens sprogteknologi: stavekontrol, talesyntese og talegenkendelse, maskinoversættelse osv. Som det er fremgået, bygger de i høj grad på de samme principper og metoder: automatisk analyse ved hjælp af regelbaserede eller statistiske modeller af sproget. De forskellige sprogteknologiske værktøjer bruger i vid udstrækning de samme basiskomponenter. I udviklingen af sprogteknologi er der derfor rige muligheder for at slå flere fluer med ét smæk ved at forbedre de sprogteknologiske basiskomponenter så de bliver bedre – især til at håndtere dansk – og det ligger der et spændende potentiale i. Basiskomponenterne vil i deres forskellige varianter kunne udnyttes på områder såsom sprogstøtte, oversættelse, den vigtige grundforskning i sproget selv i store tekstkorpusser, kundetilfredshedsundersøgelser ved ”Big Data”-analyse og ”text mining” (dvs. statistiske undersøgelser af store mængder af fx tweets), sagsbehandling (fx diktering af journaler), ordblindestøtte, sprogundervisning af flygtninge og andre immigranter, hjælpemidler (fx udvikling af høreapparater), menneske-maskine-interaktion (fx talestyrede velfærdsteknologier og industrirobotter), oplevelseselementer – og nye emner som ingen havde set.

Men det er dyrt at lave de nødvendige sprogteknologiske basiskomponenter, og erfarin-

Sidste artikel  
i serien om sprogteknologi  
i din hverdag

gen har vist at Danmark er så lille at danske sprogteknologifirmaer har svært ved at trives på rene markedsvilkår. Derfor er meget af den sprogteknologi vi kender fra hverdagen, blevet udviklet af store internationale firmaer, der ikke prioriterer relativt små sprog som dansk så højt, og som derfor heller ikke nødvendigvis benytter sig af danske sprogeksperter. Vi plæderer i denne artikel for at der er behov for en sprogstrategi der tilgodeser dansk sprogteknologi på danske præmisser.

## Hvorfor er sprogteknologi dyr?

Sprogteknologi er altid baseret på analyse og opmærkning af sprogligt input.

Tag fx talesyntese: Skal *bande* have en udtale der rimer på *vande*, eller en der rimer på *vante*? Det kan løses ved hjælp af grammatisk opmærkning. Hvis *bande* enten får etiketten ’verbum’ eller ’substantiv’, giver udtalen sig selv. Hvis de teknologiske sprog tjenester skal kunne afgøre om ordet *skade* refererer til en fugl eller til en beskadigelse, må der derimod semantisk opmærkning til.

Der er overordnet set to måder at opmærke på: via regler formuleret ud fra sproglig grundviden og via sandsynligheder beregnet statistisk ud fra andre data der er opmærket i forvejen – dvs. i hånden. Begge er dyre.

Den statistiske tilgang er den fremherskende i øjeblikket, og der er derfor fokus på indsam-



ling og opmærkning af store mængder grunddata. Selvfølgelig er de store datamængder blevet nemmere at få fat på, for når først et system som fx Google Oversæt eller Apples Siri kører, genererer det en konstant brusende strøm af sproglige data. Og de statistiske opmærkningsalgoritmer raffineres hele tiden. Efterhånden som systemerne ser flere og flere data, bliver de gode til at håndtere hyppigt forekommende ord- og sætningsmønstre, hvorimod mindre hyppige mønstre vil kræve flere data. Udviklingen går i retning af at systemerne selv hjælper til med at udvælge data de har brug for at lære mere om, og som der derfor skal opmærkes flere af; en tilgang der går under navnet *aktiv læring*.

Men materialet skal stadig opmærkes i hånden. Det er den menneskelige indsats der koster penge – både de sprogeksperter der opmærker materialet og de datalingsvister der programmerer de sprogteknologiske værktøjer. Den regelbaserede tilgang er selvsagt også meget dyr på grund af de mange sprogeregler der skal formuleres af leksikografer og lingvister.

## Der er et alternativ til de multinationale

Gennem de seneste fem-seks år har en gruppe på omkring 25 kommuner benyttet taleteknologien til at styrke deres tekstproduktion, give aktiv sprogstøtte (til nydanske funktionærer, blinde og svagtsende, ordblinde mv.) og

mindske arbejdsskader (musearm og anden nedslidning). Men desværre har teknologien ikke været stærk nok til at kunne sikre den ønskede produktivitet. For at komme videre kræves et kvalitetsløft og dermed investeringer.

Den offentlige sektor er en af de største – og i fremtiden mest afgørende – brugere af danskudviklede IT-systemer, ikke mindst sprogteknologi. Kommuner og regioner har tilsammen et enormt økonomisk volumen og står for næsten halvdelen af det danske bruttonationalprodukt. Den enkelte kommune er imidlertid, økonomisk set, en svag agent, som sjældent tør stå alene med udvikling af software og IT-værktøj. Derfor har over halvdelen af de danske kommuner sluttet sig sammen i en interessekreds kaldet Det Offentlige Digitaliseringsfællesskab, ofte omtalt som OS2. Blandt OS2's undergrupper (dem er der mange af) er OS2Talk særligt interessant i et sprogteknologisk perspektiv. Gruppen består blandt andet af de kommuner der har praktisk erfaring med automatisk talegenkendelse.

Særligt de danske sprogkomponenter har været problematiske i de talegenkendere kommunerne har kunnet vælge imellem, som alle er udviklet af det amerikanske firma Nuance Communications. OS2Talk blev derfor oprettet med det formål at undersøge hvordan et alternativ kan skabes i et bredt samarbejde mellem de mange interessenter.

### Artikelserie: sprogteknologi

Dette er den sidste artikel i Nyt fra Sprognævnets serie om sprogteknologi. (Det skrev vi også sidst, men nu mener vi det!). Artikelserien handler om hvordan sprogteknologi egentlig virker. Nævnet vil i de kommende år beskæftige sig mere med

sprogteknologi, ikke bare med fokus på dens betydning for retskrivningen, men også med henblik på at sikre dansks status i alle sammenhænge – nu og i fremtiden.

At lave god talegenkendelse som kan nedskrive danske talte sætninger korrekt, kræver tre komponenter: Et stort, velstruktureret og velannoteret talesprogskorpus (fonetisk transskriberede taleprøver, helst mindst 500 timers tale i alt), et fonetisk transskriptionsværktøj som kan lydskrive selv dagligdags dansk (ikke bare pænt oplæste ordlister o.l.), samt en sprogmodel (en statistisk tabel over samforekomster af ord) baseret på mindst 500 mio. ord (inddelt i sektioner der afspejler de fagområder som den færdige genkender skal fungere i). Disse sprogkomponenter udgør en engangsinvestering da de kan genbruges når først de er udviklet.

### Robotterne kommer!

På det netop afholdte World Economic Forum i Davos, Schweiz, var temaet i år "Den fjerde industrielle revolution": en vision om den accelererende udvikling indenfor robotik, IT, bioteknologi, nanoteknologi m.m. og hvordan vi som samfund skal håndtere den. Det er også relevant for Danmark – og for sprogteknologien.

Faktisk er Danmark i øjeblikket et brændpunkt for robotteknologi, med særligt fokus på samarbejdet mellem menneske og maskine. Ved den internationale First Robotics Competition 2016 vil et elitehold af danske gymnasieelever bl.a. præsentere en robot som forstår dansk. Robottens arm styres med talegenkendelse, gennem kommandoer såsom *Ræk mig lige den skruetrækker, Hæld vandet op i tepotten når det koger* osv. Projektet ledes af en gruppe ingeniører og forskere og er sponsoreret af firmaet Universal Robots i Odense.

Kontakt gerne Peter Juel Henriksen (pjh.ibc@cbs.dk) for mere information.

Hvis den slags sprogkomponenter var offentligt tilgængelige, ville kommunerne være i en helt anden situation. Der ville være flere produkter på markedet fra danske producenter, og de udenlandske firmaer ville formentlig være mere interesserede i at integrere de danske komponenter i deres produkter. Dermed ville kvaliteten, og dermed også produktiviteten, ligge på et langt højere niveau end før.

Hvis vi kan få denne solidariske tilgang op i national skala, vil der være lagt et godt fundament for brugen af dansk i mange nye, digitale sammenhænge med både kulturelle og økonomiske gevinster til følge.

### Den sprogteknologiske værktøjskasse

I Nederlandene har man igennem flere år arbejdet med en strategi for udvikling af en åben, dvs. offentligt tilgængelig, værktøjskasse med basale sprogteknologiske komponenter.

Det lykkedes førende forskere i Nederlandene og Belgien at formå myndighederne til at afsætte sprogteknologiske forsknings- og udviklingsmidler<sup>1</sup> via forskningsprogrammet STEVIN, som løb fra 2004-2009<sup>2</sup>. Ønsket var at stimulere virksomheder til at udvikle sprogteknologiske produkter for nederlandsk og flamsk.

Det lykkedes langt hen ad vejen at virkeliggøre ideerne, og programmet fik international anerkendelse. I dag optræder nederlandsk typisk på niveau med større sprog som fransk, tysk og spansk mht. sprogteknologisk dækning, fx i kortlægningen af det sprogteknologi-

1 Se Binnenpoorte et al. (2002). A Field Survey for Establishing Priorities in the Development of HLT Resources for Dutch. Proceedings of LREC-2002, Las Palmas, Spanien.

2 Se <http://tst-centrale.org/stevin/english>.

ske niveau gennemført af organisationen Multilingual Europe Technology Alliance (META). Rapporten fra Nederlandene konkluderer meget klart: Kun gennem dedikerede programmer som STEVIN blev det muligt at skabe de sprogteknologiske basiskomponenter der gjorde det mere attraktivt for virksomheder at udvikle produkter og tjenester på nederlandsk.

## Den sprogteknologiske håndværker

En åben sprogteknologisk værktøjskasse er nødvendig for at løfte sprogteknologien, men den gør det ikke alene. Der er også brug for sprogteknologiske specialister der kan håndtere værktøjerne. Samtidig er der selvfølgelig brug for den samme slags specialister for løbende at udvikle og vedligeholde de forskellige komponenter og ressourcer.

Selv om der i dag forskes i sprogteknologi på et højt internationalt niveau i Danmark, uddanner vi meget få sprogteknologer. Der findes ikke en egentlig uddannelse i sprogteknologi, kun valgfag og delmoduler ved enkelte uddannelser, som i kandidatuddannelsen IT & Cognition ved Københavns Universitet. Men vi har brug for eksperter der dels er gode til computerprogrammering, dels har en god indsigt i dansk sprog. Der er et behov for en mere grundlæggende uddannelse i sprogteknologi på bachelor- og kandidatniveau, og aftagerne for en sådan uddannelse er der netop nu.

## Hvor bør man så satse?


Fordi håndopmærket sprogligt materiale er så dyrt at udvikle, er det vigtigt at det kan deles og bruges til flere forskellige formål. Vi er i Danmark blevet langt bedre til at genanvende elektroniske ordbøger, dvs. både til almindelig opslagsbrug og til brug i mange forskellige

sprogteknologiske tjenester, fx bruges Ret-skrivningsordbogens grunddata i dag både i bibliotekernes informationssøgning, i Microsoft Words stavekontrol og i Wordfeud.

Men uanset hvor gode vi er til at udnytte ressourcerne effektivt, ændrer sproget sig løbende, og behovet for at sprogteknologien kan tilpasse sig denne udvikling ligeså. Der bør derfor sættes vedholdende og aktivt på en levende og opdateret værktøjskasse, allerhelst i en åben sprogbank der kan sørge for den nødvendige kontinuitet.

Det er der heldigvis rig mulighed for, eftersom forskellige sprogteknologiske værktøjer som nævnt deler mange principper og metoder. Så man vil i princippet til hver en tid kunne fokusere på én anvendelse og samtidig hjælpe flere andre.

Et løft af dansk sprogteknologi kunne fx begynde med stave- og grammatikkontroller. For at stavekontroller skal fungere godt og også kunne hjælpe med grammatik, faste udtryk og måske endda hele tekstens opbygning, kræves der mere omfattende modeller af det danske sprog. De kan indbefatte forskellige komponenter. En navnegenkender kan fx afgøre at et navn med flere led (fx bogtitler, for- og efternavne eller flerleddede firmanavne) kun skal opfattes som ét 'ord' i sammenhængen. Hvis der er adgang til en komponent med de relevante termer og synonymer, vil systemet også kunne fange "forkerte" eller ukonventionelle faste udtryk, sikre den korrekte terminologi eller foreslå danske alternativer til engelske udtryk.

Flere af disse komponenter er også relevante for maskinoversættelse, talegenkendelse osv. Og der findes allerede komponenter der kan udføre nogle af dem, fx udviklet ved 

Center for Sprogteknologi ved Nordisk Forskningsinstitut, Københavns Universitet, der netop har dette som en af sine opgaver.

Der findes altså stadig lommer af ekspertise hvor der udvikles og vedligeholdes sprogteknologiske ressourcer og værktøjer. Der findes også repositorier, dvs. hjemmesider hvor man kan deponere både data og værktøjer med en selvvalgt grad af offentlighed (CLARIN er et sådant repositorie, META-SHARE et andet<sup>3</sup>). Der findes altså allerede muligheder for at udvikle og distribuere de sprogteknologiske komponenter. Det der således primært mangler, er muligheden for systematisk at gøre disse ressourcer gratis og offentligt tilgængelige.

## Hvordan støtter man dansk sprogteknologi?

Det bedste man kan gøre for at støtte dansk sprogteknologi, er at sørge for at de mest basale komponenter til de forskellige sprogteknologiske værktøjer bliver udviklet, gjort tilgængelige og vedligeholdt løbende som en del af den offentlige infrastruktur på samme måde som fx geodata eller oplysninger om flyafgange, togtider mv. i dag stilles gratis til rådighed for udviklere af mobiltjenester så de kan levere en bedre service til borgerne. Herfra vil små som store, offentlige som private virksomheder kunne bruge deres ressourcer på kreativ (videre)udvikling og raffinering af sprogteknologien frem for at bruge tid og kræfter på basiskomponenterne. Det vil give en højere kvalitet og en hurtigere udvikling og skabe en underskov af firmaer med direkte ekspertise og interesse i dansk – en dansk sprogteknologisk industri

som får mulighed for at konkurrere med de internationale IT-giganter.

Ligesom den danske vindmølleindustri havde brug for en strategisk offentlig indsats for at opnå sin nuværende stjernestatus på verdensmarkedet, skal der offentlige forsknings-, udviklings- og vedligeholdelsesbevillinger til for at stimulere udviklingen af en dansk sprogteknologisk industri. Her har sprogteknologien i Danmark generelt haft det svært i de senere år. Der har ikke været en strategisk satsning på dette område siden udviklingen af dansk talesyntese i 90'erne, og de bevillinger der er givet, er ikke blevet givet efter en samlet strategisk plan, som man fx har set det i Nederlandene.

## En åben sprogbank

Sprogteknologien bliver lige så stille integreret mere og mere i vores hverdag, og der er store muligheder for spændende og nyttige anvendelser. Desværre styres udviklingen i øjeblikket i høj grad af store multinationale selskaber. Hvis vi ønsker en sprogteknologi af høj kvalitet der kan udvikles hurtigt af offentlige såvel som private virksomheder med blik – og øre – for det danske sprog, er der brug for at det offentlige understøtter den danske sprogteknologi. Der er behov for at få konsolideret den udvikling og vedligeholdelse der allerede foregår, i en offentlig, åben sprogteknologisk værktøjskasse – allerhelst i form af en åben sprogbank som man allerede har det i fx Norge, Sverige og Finland.

Sproget er krumtappen i vores videnssamfund. Det er kernen i al vidensoverførsel og al kommunikation – og for sprogteknologien er det et råstof med et stigende potentiale for at skabe værdi for hele samfundet.

Alle forskningsfelter har brug for basisres-

<sup>3</sup> Se hhv. <http://clarin.dk> og <http://metashare.cst.dk>.

sourcer. Indenfor medicin kan det fx være vævsprøver, men en enkelt vævsprøve kan man ikke finde meget i. For at se de store mønstre og udvinde ny viden skal der mange flere prøver til, og de administreres og håndteres bedst af specialiserede institutioner som fx Statens Seruminstutts store og internationalt misundte biobank.

Sproget er den eneste ressource der bliver mere af jo mere vi bruger den. Og behovet for central koordinering og udvikling, fx i en sprogbank e.l., er måske endda større. Sproget skal indsamles, bearbejdes og tilgængeliggøres for at gøre gavn via sprogteknologien og via sprogforskernes korpusundersøgelser.

Forskningsministeren har flere gange bebudet at der vil komme en samlet national sprogstrategi mens han er minister. Det kan man kun være taknemmelig for. Vi har med denne artikel villet gøre opmærksom på at en sådan sprogstrategi bør indeholde en strategi for hvordan man udvikler, vedligeholder og anvender værktøjer som støtter at dansk bliver brugt smidigt og raffineret på alle de områder hvor det er uundværligt, ikke mindst den offentlige sektors mange servicefunktioner rettet mod den almindelige borger.



**Philip Diderichsen (f. 1977)**  
*er videnskabelig medarbejder i Dansk Sprognævn.*



**Peter Juel Henriksen (f. 1962)**  
*er lektor i sprog- og taleteknik på CBS.*



**Sabine Kirchmeier-Andersen (f. 1955)**  
*er direktør for Dansk Sprognævn.*



**Bolette Sandford Pedersen (f. 1962)**  
*er professor i sprogteknologi ved Københavns Universitet.*

### **Donér dine tekster til forskningen!**

Sprogforskere fra Københavns Universitet og Dansk Sprognævn har på "græsrodsniveau" taget initiativ til en indsamling af hverdagstekster fra den danske befolkning. Forskerne er interesserede i uredigerede tekster såsom fødselsdagskort, kærestebreve, dagbøger, huskesedler, henvendelser til det offentlige, facebookopdateringer, e-mail, sms'er osv.

Send gerne fotos eller scanninger af tekster med

dato, skribentens køn, fødeår, fødeby og nuværende geografiske tilhørssted, uddannelse og jobtitel til [tekstbase@hum.ku.dk](mailto:tekstbase@hum.ku.dk). Alle tekster bliver anonymiseret og håndteret forsvarligt, og der tages hensyn til hvis der er tekster som kun ønskes tilgængelige for forskerne, ikke offentligt tilgængelige.

Se også <http://dsn.dk/nyt/nyheder/2016/hverdagens-skriftsprog> for mere information.

# Jørgen Schack – 25 år i Dansk Sprognævn



Seniorforsker Jørgen Schack har fejret 25-års-jubilæum i Dansk Sprognævn.

Jørgen Schack blev ansat i Sprognævnet den 1.1.1991 og har været en af de vigtige drivkræfter bag udgivelsen af Retskrivningsordbogen i 1996, 2001 og 2012.

I sin forskning, som dækker et bredt udsnit af dansk morfologi og syntaks, har Jørgen Schack bl.a. beskæftiget sig med danske nominalers struktur og semantik og med ordmærkers (varemærkers) opbygning og karakteristika.

Han er medlem af opgavekommissionen for dansk læsning og retskrivning (Undervisningsministeriet) og står i spidsen for Sprognævnets omfattende undersøgelse af stavfejl i folkeskolens afgangsprøver. Endvidere arbejder han på at styrke Sprognævnets empiriske grundlag i et projekt om automatisk ortografisk rangering af tekster.

Jørgen Schack er en af de forskere der ikke kan lide løse ender og halvgjort gerning, og han kaster sig gerne over problemer i sprogets krinkelkroge som andre har opgivet at løse, bl.a. orddannelse i forbindelse med varemærker, nominalers struktur, adverbialer, attributivkonstruktioner og kommaer.

Sprognævnet ønsker tillykke med jubilæet og ser frem til mange flere års samarbejde!

## Jørgen Schacks bidrag til *Nyt fra Sprognævnet (Nfs)*

*Nye stater og områder.* NFS 1992/1, s. 8-9.

*Den indre svinehund.* NFS 1992/3, s. 1-2.

*Tjekkiet og Slovakiet.* NFS 1992/4, s. 1-2.

*Statsnavne på dansk og nordisk.* NFS 1993/1, s. 6-8.

*Om ordet neger.* NFS 1995/4, s. 1-5.

*Imponerende 10,3 sekunder.* NFS 2001/3, s. 10-14.

*Ændringer i opslagsords stavform eller ordform.* NFS 2001/4, s. 12-17.

*Ændringer i retskrivningsreglerne.* NFS 2001/4, s. 18-20.

*Anbefalede Plantenavne.* NFS 2003/3, s. 1-4.

*Skal vi have former af typen gymnasie i Retskrivningsordbogen?* NFS 2009/4, s. 20-23.

*Adverbium + præposition – ét eller to ord?* NFS 2011/3, s. 1-7.

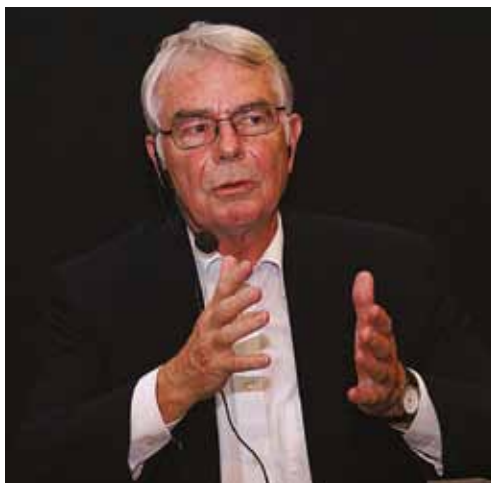
*Substantiver på -el, -en og -er.* NFS 2013/1, s. 20-23.

*Auten(ti)citet.* NFS 2013/3, s. 10-11.

*Rigmøde - historien om et ord.* NFS 2014/2, s. 11-13.

*Jagten på den gode og sikre sprogruger* (s.m. Philip Diderichsen). NFS 2015/3, s. 1-8.

# Jørn Lund 70 år



Den 30. januar 2016 fyldte Jørn Lund 70 år. Jørn Lund blev personligt medlem af Dansk Sprognævn i 1980, og det har han været lige siden. Han har siddet i nævnets arbejdsudvalg og repræsentantskab, ligesom han har været medlem af nævnets publikationsudvalg. I 2013 blev han formand for Sprognævnet.

Jørn Lund var fra 1968 til 1980 underviser og forsker på Københavns Universitet og blev herefter professor på Danmarks Lærerhøjskole. I 1991 blev han chefredaktør for Den Store Danske Encyklopædi og var desuden direktør for forlaget Gyldendal i perioden 1995-2002. Fra 2002 til 2011 var han direktør for Det Danske Sprog- og Litteraturselskab. Siden 1989 har han været medlem af Det Danske Akademi, i en årrække som dets sekretær.

Jørn Lund har skrevet en lang række bøger, afhandlinger og artikler om dansk og nordisk sprog, litteratur og kultur.

Sprognævnet ønsker tillykke med den runde dag og ser frem til samarbejdet i de kommende år!

## Generelle abonnementsvilkår

### Levering

Nyt fra Sprognævnet udkommer 4 gange om året, i marts, juni, september og december. Hvis du midt i måneden endnu ikke har modtaget bladet, bedes du henvende dig til Dansk Sprognævn for at få tilsendt et erstatningseksemplar.

### Betaling

Abonnementet betales forud, og opkrævning sker i januar, enten ved at Dansk Sprognævn sender dig et indbetalingskort (til betaling på posthuset, i banken eller via netbank) eller via BetalingsService (BS).

Som ny abonnent modtager du de numre der er udkommet i indværende år, samt et indbetalingskort for et fuldt årsabonnement.

### Pris

Prisen er 120 kr. om året. Du betaler for et år ad gangen, så selvom du opsiger dit abonnement midt på året, vil du stadig modtage og betale for resten af årets udgivelser.

### Varighed og opsigelse

Abonnementet løber indtil du opsiger det skriftligt, enten per brev eller mail. Seneste frist for opsigelse er 10. december. Det er ikke gyldig opsigelse blot at undlade at betale. Ved opsigelse af abonnementet skal du huske selv at opsiges en evt. betalingsaftale med BS.

### Flytning

Flytning skal du meddele per telefon, mail eller brev. Ved henvendelser om abonnementet skal du have dit abonnementsnummer klar.



# Nyt fra Sprognævnet

2016/1 · marts

Den der skriver <i>d</i> i <i>mand</i>	1
<i>Flygtning</i> eller <i>migrant</i> – vær opmærksom på dit ordvalg	4
Hverdagens sprogteknologi nu og i fremtiden	8
Jørgen Schack – 25 år i Dansk Sprognævn	14
Jørn Lund 70 år	15

Næste nummer udkommer i juni 2016

Ansvarshavende redaktør: Sabine Kirchmeier

Redaktionssekretær: Jørgen Nørby Jensen

ISSN 0550-7332

Layout: Falcon Grafisk Design · Dtp: Jannerup A/S

Nyt fra Sprognævnet udgives af Dansk Sprognævn. Det udkommer 4 gange om året og koster 120 kr. (inkl. moms og forsendelse) for en årgang. Man kan kun tegne abonnement hos Sprognævnet.

Usignede artikler og artikler med initialer giver udtryk for Sprognævnets mening. Artikler med navn giver ikke nødvendigvis i enhver henseende udtryk for Sprognævnets mening.

Eftertryk er tilladt når kilden angives.

## Abonnement mv.:

Telefon: 33 74 74 11 eller  
mardal@dsn.dk

Spørgetelefon: 33 74 74 74  
(Mandag-torsdag 10-12 og 13-15)

[www.dsn.dk](http://www.dsn.dk)  
[www.sproget.dk](http://www.sproget.dk)