

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, *uf. pl. ene*

tidsstemple *vb.*, -ede.

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

Fra *Du er fyret!* til *Du er sindet!*

Af Margrethe Heidemann Andersen

I sommeren 2007 mistede en del af medarbejderne i Danmarks Radio deres arbejde. I den anledning kunne man læse følgende i en artikel i Politiken 14.6.2007:

Ledelsen i DR har forsøgt at gøre dagens fyringsrunde så tålelig som mulig, men det er naturligvis aldrig rart at blive prikket.

... Det opfattes selvfølgelig meget voldsomt for den enkelte at blive 'sindet' (at få at vide, at man bliver fyret, red.) .

Det sprogligt interessante ved citatet er de ord der bliver brugt om fyringerne. Der nævnes en *fyringsrunde*, at det aldrig er rart at blive *prikket*, og at det opleves som voldsomt at blive *sindet*. Tre ord for den samme begivenhed – nemlig at en række mennesker har mistet deres arbejde. De fleste vil formentlig hurtigt kunne se at udtrykkene *at blive prikket* og *at blive sindet* er mindre direkte end at blive fyret. De er det man indenfor sprogvidenskab kaldes eufemismer, altså omskrivninger

af noget som vi opfatter som skamfuldt eller tabubelagt. Pakker vi de ubehagelige ting lidt ind sprogligt, er det nemmere at snakke om dem. Det er derfor vi har en række ord og begreber for det at gå på toiletet (*fx gå på det lille hus, gå af sides, gå ud for at pudre næsen*), for at dø (*sove stille ind, gå bort, fare til himmels*) eller for at være sindssyg (*være afvigende, være mentalt ustabil*). Og det er derfor vi ser så mange forskellige betegnelser for det at blive fyret (og i øvrigt også for at være arbejdsløs) – fordi det er ubehageligt, hårdt og til tider smertefuldt at blive fyret. Jeg vil i artiklen her komme ind på en række af de forskellige betegnelser der findes for at være blevet fyret og for at være arbejdsløs. Artiklen skal ikke opfattes som en udtømmende liste – der er ganske givet flere ord og udtryk end dem jeg nævner her, men det bliver for omfattende at skulle nævne dem alle.

At få silkesnoren

Det er ikke noget nyt fænomen at vi



laver omskrivninger for at være blevet fyret. Udtrykket *at få silkesnoren* er således relativt gammelt i dansk, og ifølge Ordbog over det danske Sprog (på ordnet.dk) blev udtrykket egentlig brugt om forhold i Orienten (Tyrkiet) eller Østen, og det henviser til at man kan ”*tilsende sin underordnede (henholdsvis: faa tilsendt) en snor af silke til tegn paa, at vedkommende bør begaa selvmord (egl. ved at kvæle sig i snoren)*”. Derudover bruges udtrykket også om at ”*befale (henholdsvis: faa befaling til) at tage sin afsked (fra et embede, en post osv.), trække sig tilbage (fra et arbejde)*”. I dag bruger vi naturligvis kun den overførte betydning, og en søgning i Infomedia i december 2014 tyder på at udtrykket især bruges om ministre, embedsmænd, topledere og andre der mister deres embede eller bestyrelsespost, eller om trænere, managere og sportsstjerner der ikke får deres kontrakt forlænget, mister deres plads på landsholdet e.l. Infomedia indeholder dog også enkelte eksempler på helt almindelige lønmodtagere der ”får silkesnoren”, som i dette eksempel fra DR.dk 24.8.2009: ”Mellem fire og seks dagplejere vil få silkesnoren, fordi der ikke længere er lange ventelister og mangel på ansatte”. Det hører dog tilsyneladende til sjældenhederne - som regel er det ikke mennesker i almindelige lønjobs der får silkesnoren, men altså snarere kontraktansatte fodboldmanagere, direktører osv.

Fra at blive fyret til at blive fritstillet

Hvis man får at vide at man har mistet sit arbejde, er man *blevet fyret*. Det ældste belæg på dette udtryk er ifølge Ordbog over det danske Sprog (på ordnet.dk) fra 1905, og det er stadigvæk meget udbredt. Det findes i alle

gængse opslagsværker over dansk sprog, og en søgning i Infomedia i december 2014 gav mere end 27.500 hits. Kunne vi så ikke bare nøjes med at bruge udtrykket *blive fyret* uden alle mulige omskrivninger? Jo, det kunne vi for så vidt godt. Problemet er bare at det (stadigvæk) er ret tabubelagt at være blevet fyret, og at det også er de færreste mennesker der bryder sig om at fortælle et andet menneske at vedkommende er blevet fyret. Derfor forsøger man at omskrive udtrykket *at blive fyret* med et andet, lidt mildere udtryk. Det fører til at vi i 1973 (se www.dsn.dk/noid) for første gang begynder at bruge udtrykket *at blive fri(t)stillet*. På www.dsn.dk/noid kan man finde flere illustrative beskrivelser af hvordan *fritstille* (som det hedder i dag) bruges. Det første eksempel stammer fra Politiken 9.2.1973 og lyder:

Ca. 50 medarbejdere bliver ’fristillet’ .. Der har været bebudet afskedigelser, men med radioledelsens nye ord fristillet skulle der ikke være tale om direkte afskedigelser. At en medarbejder bliver fristillet, betyder at den funktion, han udfører, nedlægges.

Citatet viser at det at blive fritstillet ikke nødvendigvis er helt det samme som at blive fyret. Bliver man fyret, skal man forlade sin arbejdsplads, og den funktion man udførte, bliver varetaget af en anden. Bliver man fritstillet, nedlægges funktionen. Det skulle altså – i princippet – være bedre at være fritstillet (fordi det ikke nødvendigvis er dig som person der er fundet uegnet til et bestemt arbejde) end at blive fyret. Sådan føles det dog næppe for den fritstillede person, og konsekvenserne af en fritstilling er den sam-



me som ved en fyring: arbejdsløshed og en usikker økonomi. Det gennemskuer sprogbrugerne hurtigt, og derfor kan man allerede senere i februar 1973 finde følgende belæg i avisen: ”2800 medarbejdere .. skal afskediges (eller skal vi benytte det nydannede ord: skal fristilles)” (Berlingske 20.2.1973). Særligheden mellem at blive fritstillet og at blive fyret (eller afskediget) er altså forsvundet i stort set samme øjeblik som den nye term *fritstillet* blev indført.

At blive prikket og at blive sindet

I 1992 opstår en ny betegnelse for at blive fyret, nemlig *at blive prikket* (se www.dsn.dk/noid). Udtrykket refererer egentlig til at den person der skal fyres, får et prik på skulderen og en besked om at vedkommende skal til samtale med ledelsen – altså en samtale hvori det meddeles at personen er blevet fyret. Er

det mange medarbejdere der skal fyres, kan man tale om en hel *prikkerunde*, altså en fyrringsrunde. Udtrykket *at blive prikket* bruges stadigvæk og findes også i de gængse opslagsværker. Ikke desto mindre er der i de senere år dukket endnu en ny betegnelse op, nemlig *at blive sindet*. Dette udtryk er formentlig opstået i kølvandet på ordet *sinde(t)skrivelse*, som er en hensigtserklæring (fx fra en forvaltning) om hvad man har i sinde at gøre (eller er sindet at gøre) i en bestemt sag, som i dette belæg fra Århus Stiftstidende 27.2.1999: ”Derfor ønsker ... HK at gøre brug af den udtaleret, vi har i forhold til den sindetskrivelse, vi har modtaget, og hvori der tages skridt til at afskedige ...” (se også Nyt fra Sprognævnet 2004/3). Det ældste belæg som jeg har kunnet finde på udtrykket *at blive sindet*, er det belæg fra 2007 som nævnes i indledningen til artiklen her. I et lidt yngre belæg, nemlig >

fra Politiken 28.8.2010, uddybes det hvorfor man *sinder* folk i stedet for slet og ret at fyre dem:

Mandag er der prikkerunde, og de medarbejdere, som DR vil skille sig af med, bliver 'sindet'. For teknisk set er der ikke tale om en fyring. Medarbejderne får at vide, at DR har i 'sinde' at fyre dem, og derefter har de 14 dage til at protestere, hvis de mener, at fyringen er usaglig, eller indgå frivillige aftrædelsesordninger. Derfor er det ikke store prikkedag, men store sindedag.

HR-direktøren i DR, Inger Kirk, er dog godt klar over at forskellen mellem at blive sindet og at blive fyret er minimal for den fyrede (eller sindede) medarbejder: »Får vi indsigelser, så behandler vi dem nøje, men vi mener selvfølgelig, at vi har gjort vores arbejde så godt, at afskedigelserne er saglige. For den konkrete medarbejder vil det at blive 'sindet' på mandag givet blive opfattet lige så voldsomt og indgribende som en fyring«, siger Inger Kirk».

En søgning i Infomedia i januar 2015 viser at *sindet* primært bruges i forbindelse med fyringer (eller rettere: varsel om fyringer i DR), men der er dog også eksempler på at det bruges uden for DR's regi som i dette eksempel fra alleroed.lokalavisen.dk 5.11.2014: »Teknisk set er de berørte ansatte i Allerød Kommune ikke reelt fyret, de er blevet sindet». Det tyder på at udtrykket er ved at brede sig til almensproget, og formentlig bruges det fordi det for arbejdsgiveren er vigtigt at pointere at der er forskel på at blive sindet og at blive fyret. Igen er det dog uden betydning for den medarbejder der ser ud til at miste sit

job, om han er sindet eller fyret (eller om han har fået silkesnoren, er blevet fritstillet eller er blevet prikket for den sags skyld). Realiteten er at han er blevet arbejdsløs – eller hvad vi nu skal kalde det. Sagen er nemlig den at der findes stort set lige så mange omskrivninger for det at være arbejdsløs som for det at være blevet fyret.

Arbejdsløs, arbejdsledig og jobjæger

Det ældste belæg på ordet *arbejdsløs* er ifølge supplementet til Ordbog over det danske Sprog, bind 1, fra 1834. I et lidt senere belæg fra en lov af 3. marts 1860 står der at »Den, som streifer arbejdsløs omkring . . . bliver som Løsgænger at ansee med Straf af Fængsel paa Vand og Brød» (Ordbog over det danske Sprog på ordnet.dk). På den baggrund kan det næppe undre at man finder på andre betegnelser for at være arbejdsløs, fx *arbejdsledig*, som er fra 1927. I supplementet til Ordbog over det danske Sprog finder man bl.a. dette illustrative eksempel fra Harald Hørdals roman »Man skal jo leve» fra 1934:

La os sie, der rundt regnet er et Hundretusind Arbejdsløse, - Arbejdsledige, sa han halvhøjt, saadan hedder det, det lyder endnu ikke saa brutalt og nøgternt som Arbejdsløse.

Såvel *arbejdsløs* som *arbejdsledig* er gangbare ord i dag, men alligevel ser vi igen at nyere omskrivninger finder vej til sproget. Det gælder fx ordet *jobjæger*, hvor efterleddet *-jæger* understreger den arbejdsløses ihærdige jagt på et nyt arbejde. Oprindeligt blev ordet dog brugt om en kommunalt ansat person som skulle finde arbejde til kommunens arbejds-

løse, hvilket illustreres i dette citat fra Weekendavisen 12.7.1991: ”Den kommunale jobjæger er en gevinst for de lokale skatteborgere og for de unge ledige - og han sparer samtidig staten for et millionbeløb”. Senere går ordet dog over til at betegne den arbejdsløse som i dette citat fra Prosa.dk 28.3.2012: ”Det er et tabu at være arbejdsløs i dagens Danmark, og det tabu skal brydes. Det mener Min A-kasse, som vil sætte fokus på tabuet ved at forsøge at indføre det mere positivt klingende ord ’jobjæger’ i stedet for ’arbejdsløs’”. Citatet er et godt eksempel på hvordan man forsøger at omskrive de tabubelagte ord og betegnelser med nye og mere neutrale eller positive betegnelser. Men som så ofte før er succesen begrænset fordi selve normen for hvad der er det bedste i vores samfund, stadig er den samme – nemlig at det er bedre at have et arbejde end ikke at have et arbejde. Således kan *jobjæger* have rigtig svært ved for alvor at etablere sig i vores sprog, og det findes da heller ikke i nogen af de gængse opslagsværker jeg har kigget i. Og muligvis er der allerede en ny dannelse på vej, nemlig at man er *imellem jobs*: ”Du kan sagtens være imellem jobs, selvvalgt eller pga. krisen og score damer”, som der stod i en artikel i Ekstra Bladet 21.8.2012. Udtrykket *at være imellem (to) jobs* stammer fra engelsk *to be between jobs*. Der er kun få eksempler på udtrykket i Infomedia, og det er langt fra sikkert at udtrykket etablerer sig i dansk. Formentlig vil det eksistere i nogle år for så at blive afløst af en ny betegnelse.

Udsagnene ”Jeg blev fyret i sidste måned, så jeg er arbejdsløs nu” og ”Jeg blev fritstillet/



Margrethe Heidemann Andersen (f. 1971) er seniorforsker i Dansk Sprognævn

sindet/prikket/fik silkesnoren i sidste måned, så jeg er arbejdsledig/jobjæger/imellem to jobs nu” betyder nøjagtigt det samme, i hvert fald for den almindelige sprogbruger. Men hvor første udsagn fortæller sandheden uden omsvøb, prøver de næste udsagn at tilføre det i udgangspunktet negative udsagn noget positivt. Der kan være mange gode grunde til at man hellere vil betegne sig selv som ”jobjæger” end ”arbejdsløs”, og det står selvfølgelig den enkelte person frit for at vælge den betegnelse som han eller hun har det bedst med. Man skal dog ikke være blind for at uanset hvilken betegnelse man bruger, er problemet det samme: at det i vores del af verden som oftest er et problem at være arbejdsløs, og at det er det der gør det så svært at tale om.

Talesyntese og talegenkendelse

Artikelserie:
Hvordan virker sprogteknologien
i din hverdag?

Af Philip Diderichsen og Andreas Søeborg Kirkedal

Sprogteknologien talesyntese er maskinel omsættelse af tekst til tale. Denne teknologi har eksisteret i en længere årrække og været en stor hjælp i forbindelse med oplæsning for svagtseende og andre læsesvage. Talegenkendelse har haft mere begrænset succes herhjemme, men er nu så småt på fremmarch, især på visse offentlige arbejdspladser. Talegenkendelse omsætter tale til skrift. Det bruges dels til at aflaste, dels til at effektivisere hvor der skrives meget. De to teknologier kan kombineres i det der kaldes dialogsystemer, dvs. mere eller mindre 'intelligente' digitale assistenter, som man kan stille spørgsmål til eller få til at hjælpe sig, fx på smartphones. Dialogsystemer begynder nu at dukke op i dansktalende versioner. Apples digitale assistent, Siri, er netop kommet med den seneste opdatering til iPhone og iPad; Googles tilsvarende system, Google Now, er på vej i en dansk version; og Microsofts Cortana (opkaldt efter en kunstigt intelligent karakter i Microsofts computerspilserie Halo) ligeså.

Den tid hvor talesyntese og -genkendelse kan komme til at spille en større rolle for

brugen af dansk, er således ved at tage sin begyndelse. Læs her hvordan systemerne virker.

Talesyntese

Talesyntesystemer kaldes også tekst-til-tale-systemer eller oplæsningssystemer. En tidlig pioner var videnskabsmanden Christian Gottlieb Kratzenstein, der var professor i eksperimentel fysik ved Københavns Universitet i 1700-tallet (Ohala 2011). Han forsøgte, med begrænset succes, at skabe vokallydene a, e, i, o og u ved hjælp af orgelpibelignende resonatorer i forskellige former. En senere generation af kreative fonetikere, lingvister og ingeniører ved samme universitet byggede med loddetin og elektroniske komponenter en maskine hvor man kunne finindstille en mængde overtoner i et antal tidsafsnit således at en kort lyd kunne formes til en ytring (fx, med karakteristisk studentikos humor, ytringen *Luk mig ud!*). Nu om dage foregår talesyntese selvfølgelig i computere. Men hvordan?

For at få en computer til at omforme et tekstinput til talesprog får man den i grove

Artikelserie: sprogteknologi

Dette er den anden artikel i serien om sprogteknologi. Spiller stavekontroller en rolle i udbredelsen af særskrivningsfejl? Spiller talegenkendelse en rolle for retskrivningen? Den slags spørgsmål er

på nævnets dagsorden i de kommende år. Vore læsere kan foreløbig her orientere sig om hvordan sprogteknologi egentlig virker.

træk til at køre et program i de følgende fire trin, der forklares nærmere nedenfor: 1) en tekstnormalisering, 2) en fonetisk analyse, 3) en prosodisk analyse og endelig 4) et databaseopslag, hvor den henter de lydsummer den skal bruge til at sammensætte den ønskede ytring.

Tekstnormalisering er et helt nødvendigt trin for et talesyntesystem der skal kunne 'oplæse' tekst som det ikke har 'set' før. Et godt eksempel på det er tal. Hvordan skal fx tallet 8900 udtales? Tallet skal omsættes til enkeltord for tallets komponenter – men hvad er komponenterne? Afhængigt af om der er tale om Randers' postnummer, et beløb eller måske endda en pinkode, bliver ordene helt forskellige. Og en skærmlæser for svagseende der fx siger at *Trickyve stjal otte ni nul nul kroner*, kan være svær at tage alvorligt. Et andet aspekt af tekstnormalisering er at håndtere homografer (dvs. forskellige ord der staves ens, men lyder forskelligt). De skal gøres entydige. Man kunne fx forestille sig en tekst om *bander* der *bander* meget. Her er der brug for at køre en ordklasseanalyse for at kunne give substantivet én udtale og verbet en anden. Det vil dog ikke være nok i sætninger som *Du får kost og logi hvis du kan svinge kost og spand*, hvor *kost* i begge tilfælde er et substantiv. Så må der mere avanceret, semantisk (betydningsorienteret) databehandling til.

Den fonetiske analyse er nødvendig fordi det desværre ikke er alle ords udtale der kan slås op i en ordbog, selv i kommercielle tekst-til-tale-systemer med store udtaleordbøger. Det er det evige problem med ordbøger af enhver art: Der vil altid være ord som ordbogen ikke dækker. Det kan fx være nydannede

eller nyindlånte ord i sproget. Navne er en anden stor ordgruppe der næppe kan dækkes af nogen ordbog. Her må systemet analysere de tekststrengte det ikke kender og finde en kobling mellem bogstaver og enkeltlyde. Som i megen anden sprogteknologi kan denne kobling enten findes ved hjælp af manuelt opstillede eller statistisk baserede omsætningsregler.

I den manuelle tilgang formulerer en eller flere lingvister regler for hvordan de enkelte bogstaver omsættes til forskellige lyde alt efter hvilke bogstaver der kommer før og efter.

I den statistiske tilgang tæller man de enkelte bogstav-lyd-relationer i en stor samling lydskrevne tekster hvor bogstaver og lyde allerede er koblet sammen. Fx vil sådan en optælling vise at bogstavet *i* næsten altid udtales med en e-lyd når det står før *ng* (som i *ting*), men næsten altid som et 'rigtigt' *i* før *b* (som i *skib*)¹. Både den regelbaserede tilgang og den statistiske tilgang munder således ud i et system af omsætningsprincipper som systemet herefter kan bruge til at omsætte de ukendte tekststrengte til lydskrift.

Sammensatte ord er også en del af udfordringen med ord der ikke står i ordbogen, og problematikken hænger sammen med det næste trin i processen: den prosodiske analyse. Det er nemlig den prosodiske analyse der afgør bl.a. om sammensatte ord får det rigtige tryk. Mange sammensatte ord vil være nydannede og dermed ukendte for et tekst-til-tale-system. Man ville kunne køre en kompositumanalyse og splitte sammensatte ord op i to; men det ville ikke >

¹ Websiden <http://bogstavlyd.ku.dk/> er en søgemaskine hvor man selv kan udforske den slags bogstav-lyd-relationer.

slå til. Sammensætninger udtales (i de fleste tilfælde!) netop ikke som to enkeltord efter hinanden. Orddelene får nemlig stærkere og svagere tryk afhængigt af deres placering i sammensætningen: I *den dumme bøde* og *den dummebøde* er trykket på *bøde* helt forskelligt – stærkt i eksemplet hvor der tales om en bestemt bøde, svagere i eksemplet hvor der tales om en bestemt dummebøde (hvor der er bitryk på *-bøde*).

Ordenes tryk er en del af en sætning *prosodi*. Prosodien omfatter sætningens tonale og rytmiske aspekter i det hele taget. Det er den del af fonetikken der beskæftiger sig med betydningsskabende lydænomener der rammer flere lyde (som når man ændrer trykket og dermed ændrer ordet *billigst* til ordet *bilist*) – i modsætning til de betydningsændringer der opstår når man fx udskifter et *b* med et *p* (og dermed fx ændrer ordet *bil* til ordet *pil*). Prosodien er vældigt vigtig for talesyntese, for en ellers korrekt streng af lyde kan blive helt uforståelig hvis fx trykkene i sætningen er sat forkert. Prosodien er et område indenfor dansk grundforskning. Danske fonetikere er endnu ved at undersøge præcis hvordan tryk m.m. ligger i dansk talesprog, og der er derfor en grænse for hvor perfekt prosodien endnu kan blive i dansk talesyntese. Men et tekst-til-tale-system vil udføre så god en prosodisk analyse det nu kan, og markere tryk, tone og forkortede og forlængede lyde i lydskriften fra trinnet før. Herefter har systemet en komplet fonetisk transskription af teksten og er klar til at sætte lyd på.

Til sidst foretages et databaseopslag, hvor talesyntesen finder og kombinerer de rigtige lydstumper til en færdig fonetisk streng (dvs. sekvens af sproglyde). Der er to overordnede

måder at gøre det på: syntese med difoner og syntese med større lydstumper.

Difoner er, som navnet antyder, lydenheder bygget af to lyde, således at de består af sidste del af den første lyd og første del af den anden lyd. På den måde er det centrale i lydenheden overgangen mellem to lyde. Der findes selvfølgelig flere overgange mellem to lyde end der findes enkeltlyde. Derfor vil en liste over difoner være større end en liste over enkeltlydene i et sprog, men så får man til gengæld også den vigtige overgang mellem lydene med. Systemet bygger altså den fonetiske streng ved at tage starten af den første lyd, sætte passende difoner på og til sidst sætte slutningen af den sidste lyd på. Herefter er difonsyntesen nødt til at modificere strengen ved at sætte tryk, ændre tonegangen og forlænge og forkorte visse lyde for at sætningen kan få den rigtige prosodi.

Syntese med større lydstumper følger det samme princip, men opererer med en større database fordi man så vidt muligt forsøger at få prosodien med i de lydstumper man opererer med. Man søger således at putte så lange lydstumper som muligt i sin database – gerne flere ord lange. Lydstumperne laves ved at indlæse noget tekst designet til at have de rigtige lydsekvenser og så klippe dem ud af optagelsen. Hvilke lange lydstumper der kommer med i databasen, vil i det væsentlige afhænge af deres hyppighed i talesproget.

Dansk talesyntese udbydes bl.a. af det polske firma Ivona Software (hør en demonstration på <http://www.ivona.com/us/about-us/voice-portfolio>), det danske firma Dictus ApS (http://www.dictus.dk/Lydfiler/opl%C3%A6sning_Sara.wav), Nationalbiblioteket for mennesker med læsevanskeligheder,

Nota (<http://www.nota.nu/nota-lancerer-snar-ny-talesyntese>) og Digitaliseringsstyrelsen (<http://www.adgangforalle.dk>).

Som nævnt er digitale assistenter til smartphones nu også på vej i danske versioner, og her er talesyntese en vigtig komponent. Det samme er talegenkendelse.

Talegenkendelse

Talegenkendelse er den proces hvor man omsætter sproglyde til tekst. Trekvarter århundrede efter Alexander Graham Bell opfandt telefonen, i 1952, opstod det første talegenkendelsessystem i Bell Laboratories. Det kunne genkende tal udtalt af en enkelt taler. I 1962 udviklede IBM et system (Shoebbox) der kunne genkende 16 engelske ord. Det første talegenkendelsessystem til almindelige brugere, Dragon NaturallySpeaking, blev lanceret i 1990'erne. NaturallySpeaking har bl.a. drevet talegenkendelsen i Apples Siri.

For at omsætte sproglyde til tekst skal man 1) optage og behandle lydsignalet, 2) afkode hvilke lyde der er blevet udtalt, 3) omsætte sproglydene til tekststreng og 4) vælge den mest sandsynlige tekststreng.

Først klippes lydsignalet i små lydbidder a ca. 20 millisekunder. Hver lydbid skal klassificeres som en fon (sproglyd). Foner er groft sagt bogstaverne i et lydskriftalfabet. (Her er et par eksempler på lydskrift fra Den Danske Ordbog: Ordet *vært* bliver i lydskrift til ['væʁd], og *hver* bliver til ['vɛ'ɐ]). Basisenheden i talegenkendelse er foner, ikke ord, fordi et sprog i princippet har uendeligt mange ord (man kan altid lave nye sammensatte ord osv.). Inventaret af foner er fast (der er ca. 40 i et sprog), og med dem kan man udtrykke alle ord på et sprog.

Uheldigvis kan man ikke entydigt klassificere en lydbid som en fon, men kun med en vis sandsynlighed. Fon for fon er forskellen i udtalen af ordene *værter* og *hvert år* fx meget lille. I stedet for at vælge den mest sandsynlige fon til hver lydbid og konvertere én sekvens af lydbidder til én sekvens af foner konstrueres der derfor mange forskellige fonsekvenser med forskellige sandsynligheder.

Fonsekvenserne konverteres til ordsekvenser vha. en udtaleordbog, der kobler sekvenser af foner til ord. Resultatet er igen et stort antal ordsekvenser med forskellige sandsyn-

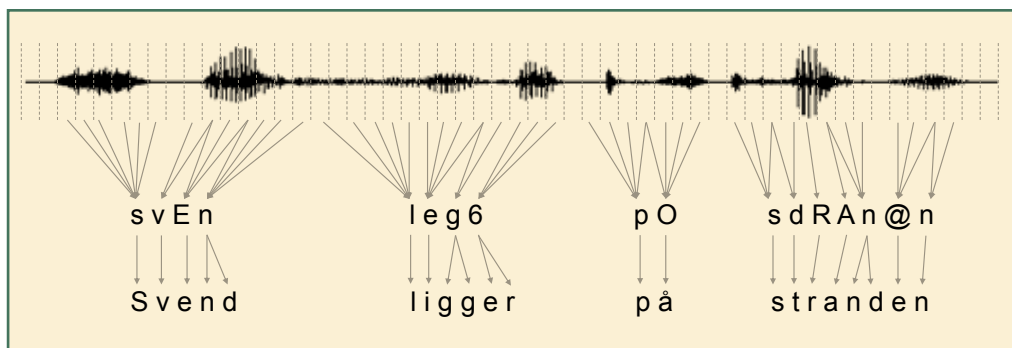


Fig. 1: Vejen fra lydsignal over foner til almindelig tekst. E, 6, O, R, A og @ angiver specifikke lyd-værdier og fungerer altså som lydskriftegn. Bemærk at der i denne figur ikke er markeret prosodiske træk (fx tryk og stød).



ligheder. Det er især pga. to udfordringer. For det første er der ingen lydlig markør af ordgrænser i talesprog. Derfor vil fonsekvensen for *hvert år* producere diverse andre muligheder både med og uden mellemrum: *vær dåre*, *hver tår*, *værter* m.fl. For det andet svarer én fonsekvens ofte til forskellige homofoner (dvs. forskellige ord der lyder ens, men staves forskelligt, fx *hver*, *vejr*, *vær* og *værd*). I sidste ende er sandsynligheden for en ordsekvens en kombination af sandsynlighederne for fonerne og sandsynlighederne for de mere eller mindre enslydende ord og ordkombinationer de ville kunne danne.

Til slut beregnes den mest sandsynlige tekststreng. Fx er *hvert år man lever* mere sandsynlig end *hver tår man lever* eller *vejr tår man lever*.

Både sandsynligheden for enkeltord ud fra fonsekvenser og sandsynligheden for sekvenser af ord skal altså beregnes, og det er ikke ligetil. Der skal laves en liste med alle sproglydes sandsynligheder i sproget. Denne liste kaldes en *akustisk model*. Der skal laves en tilsvarende liste over gængse ords sandsynligheder. Det kaldes en *sprogmodel*.

Man beregner en akustisk model med en stor samling tekst med tilhørende lydoptagelser. Inden man beregner sandsynligheder, bruger man udtaleordbogen til at oversætte teksten til en fonetisk transskription. Ordet *hvert* bliver fx oversat til fire foner – en v-fon (det stumme *h* forsvinder), en æ-fon, en r-fon og en d-fon (*t'et* lyder som et *d* sidst i et ord). Dermed bliver det muligt at beregne sandsynligheden for at en given lydbid repræsenteres af en bestemt fon. En udfordring er at en vokal plejer at være længere end en konsonant, så man skal udregne en synkro-

nisering af den fonetiske transskription og lydbidderne.

En anden udfordring kaldes *koartikulation*. Det er den effekt en sproglyd har på andre sproglyde. Fx er der en koartikulationseffekt på udtalen af *k* i *kom* og *kim* fordi bl.a. læberne er positioneret forskelligt pga. vokalen (runde ved *o*'et og brede ved *i*'et). På samme måde er vokalen i *kom* 'farvet' af *k* og *m* og derfor anderledes end *o* i *sold*. Derfor opdeles man i praksis foner i start-, midter- og slutfon. Start- og slutfoner opdeles yderligere alt efter hvilken fon der er i nærheden. Disse underinddelinger kalder man *trifoner*, og det har vist sig at være mere robust og nøjagtigt at bruge trifoner end at bruge foner.

En sprogmodel beregnes ud fra et korpus af ren tekst ved at lave en liste af sandsynligheder for små stumper løbende tekst på nogle få ord hver. En sætnings sandsynlighed kan så beregnes som produktet af sandsynlighederne for de stumper den består af, som beskrevet i artiklen om stavekontrol i sidste nummer af Nyt fra Sprognævnet (<http://www.dsn.dk/nyt/nyt-fra-sproгнаevnet/nfs-2015-1>). Her er det vigtigt at den tekst man bruger, minder om det man gerne vil genkende. Hvis man vil genkende lægediktater, skal man træne sin sprogmodel på lægediktater og ikke på fx H.C. Andersens eventyr.

I Danmark er Dragon fra Nuance Communications det mest udbredte talegenkendelsessystem og sælges af KMD, IBM, Dictus og Max Manus m.fl. Andre systemer udbydes af eFaktum, Mirsk Digital og SpeechOp. Teknologien har i en årrække været i brug til diktering af Folketingstidendes referater af Folketingets forhandlinger og andet. Et andet område hvor man i Danmark arbejder på at

indføre talegenkendelse, er lægediktering på hospitaler. Man vil også gerne udbrede teknologien i kommuner hvor der skrives mange rapporter. Indførelsen af talegenkendelse er som regel begrundet i en forventning om effektivisering baseret på at man kan indtale tekst hurtigere end man kan skrive (se fx Zapata & Søeborg Kirkedal 2015).

Når Dansk Sprognævn i disse år interesserer sig for taleteknologi, er det især fordi det kan tænkes at de tekster der produceres vha. talegenkendelse, har særlige retskrivningsmæssige problemer – mest oplagt i forbindelse med homofoner, der jo pr. definition lyder ens og ikke altid vil kunne korrigeres af talegenkenderens indbyggede stavekontrol. Skal der fx stå *køre* eller *kører*, *lære* eller *lærer* osv.? Omvendt kan det også tænkes at talegenkendelse kan være en hjælp til korrekt stavning, da en talegenkender jo fx ikke kan lave egentlige slåfejl. Der er endnu ikke megen dokumentation for denne slags påvirkninger.

Det kan virke overraskende når en maskine er i stand til at bruge talesproget, fra den gamle syntesemaskines korte, men i sin kontekst altid meningsfulde ”Luk mig ud!” til vore dages telefoners evne til at svare igen. (Spørger man fx Siri: ”Hvor bor du?”, får man tilbage med samme mønt: ”Lige her.” ..). Nu hvor dansk har vundet indpas hos de store udbydere af sprogteknologi i lommeformat, forsvinder dansk talegenkendelse og -syntese næppe igen lige med det samme (omend det kan være nok så bekymrende at udviklingen af teknologierne forsvinder ud af landet). Det vil vise sig hvor meget de endnu vil kunne



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



Andreas Søeborg Kirkedal (f. 1985), Mirsk og Copenhagen Business School.

udvikles og forbedres - og om de på et tidspunkt vil begynde at virke tilbage på vores egen brug af sproget.

Litteratur

- Ohala, John J. (2011). *Christian Gottlieb Kratzenstein: pioneer in speech synthesis*. The 17th International Congress of Phonetic Sciences (ICPhS XVII), Hong Kong, Kina.
- Zapata, Julián & Andreas Søeborg Kirkedal (2015). *Assessing the Performance of Automatic Speech Recognition Systems When Used by Native and Non-Native Speakers of Three Major Languages in Dictation Workflows*. Proceedings of the 20th Nordic Conference of Computational Linguistics, Vilnius, Litauen.

Sprognævnets hjemmeside

En gruppe i Sprognævnet arbejder med at opdatere nævnets hjemmeside: www.dsn.dk.

Bruger du siden? Så vil vi gerne høre:

- Hvad du bruger www.dsn.dk til?
- Hvor du synes hjemmesiden kan forbedres og hvordan?

Send dine kommentarer og forslag senest 1. juli 2015 på adm@dsn.dk. Skriv ”www.dsn.dk” i emnefeltet.

Vi sender en lille ting til de første 3 der svarer ☺



SPROGPRISSEN.DK

Fortjener din tekst en pris?

Sprogprisen.dk uddeles i år torsdag den 1. oktober 2015, og du kan indstille en tekst til prisen.

Vi uddeler sprogroser i følgende kategorier:

- Årets bedste brev eller e-mail til borgere, kunder eller medlemmer
- Årets bedste hjemmesidetekst
- Årets bedste tekst i sociale medier

Du kan indstille tekster i en eller flere kategorier – enten en tekst som du selv er ansvarlig for, eller en tekst som du har læst og haft glæde af.

Bag prisen står: Dansk Kommunikationsfor-
ening, Dansk Sprognævn, Det Danske Sprog- og
Litteraturselskab, Forbundet Kommunikation og
Sprog, kurtstrand.dk, Modersmål-Selskabet, RetorikMagasinet, sproget.dk og Syntaksis.

Du kan indstille teksten via [Facebook.com/sprogprisen](https://www.facebook.com/sprogprisen) eller via hjemmesiden sprogprisen.dk. Fristen er 1. september 2015.

LÆS MERE PÅ HJEMMESIDEN: [SPROGPRISSEN.DK](http://sprogprisen.dk).

Nye medier. Sprogbrug, data og metode

I 4. bind af Sprognævnets konferencereserie kan man læse om sproget i de nye medier og om hvordan man har arbejdet og vil arbejde med det i Danmark og i de andre nordiske lande.

Bindet har følgende indhold:

- Forord ved Marianne Rathje

Korpusindsamling:

- Sabine Kirchmeier-Andersen: Hvordan kan man undersøge sproget i de nye medier?
- Rickard Domeji: Att utforska språket i gamla och nya medier – så gör vi i Sverige
- Thorbjørg Breivik: Korpus fra nye medier – muligheter og begrensninger Sprog og interaktion i de nye medier

Sprog og interaktion i de nye medier:

- Jenny Myrendal: Från meningspotential till situerad betydelse: ordbetydelseförhandling i forumdiskussioner på webben
- Tina Thode Hougaard: Sproget til billederne i de sociale medier
- Astrid Westergaard: Facebook og dansen af praksisfællesskaber
- Julia Kimell: ”Jo häjjo, ha du vari å fiska någi åå då?” – ålänningars dialekt och identitetskapande på Facebook, i e-post och i sms-meddelanden.



Nye medier. Sprogbrug, data og metode. NoSSinK-symposium, København 11. – 12. april 2013. Redigeret af Marianne Rathje. Sprognævnets konferencereserie 4. Dansk Sprognævn 2015, 132 sider.

Bogen kan købes i boghandlen eller bestilles hos Dansk Sprognævn (tlf. 33 74 74 11 eller mardal@dsn.dk.) til en vejledende pris på 150 kr.



Orddelen *-hed*

SPØRGSMÅL: En af mine indvandreerelever har spurgt mig hvad *-hed* betyder i ord som *danskhed, dumhed, kærlighed*. Jeg vil lade spørgsmålet gå videre til jer.

SVAR: Orddelen *-hed*, som i *enighed, dumhed, sandhed, stilhed*, har engang været et selvstændigt ord, som har betydet 'måde' eller 'beskaffenhed'. Det er meget tidligt lånt ind i de nordiske sprog fra middelnedertysk *heit*, og det blev til *-het* i svensk, *-heit* i islandsk og *-hed* i dansk og norsk. I løbet af middelalderen blev *-hed* den mest brugte orddel som man kunne sætte efter et adjektiv, og dermed betyde måden eller beskaffenheden af det som adjektivet angiver'. *God + hed* bliver så til 'måden at være god på', *stor + hed* = 'beskaffenheden af at være stor'.

I dag bruger man *-hed* til at danne abstrakte substantiver af adjektiver, og endelsen betegner en tilstand, egenskab eller kvalitet ved det som adjektivet udsiger, fx *hurtighed, dygtighed, holdbarhed, velgørenhed, renlighed, fandanivoldskhed, belæstthed*. Endelsen er almindelig og kan sættes på stort set alle adjektiver, så kun et fåtal af de mange ord der kan dannes, bliver taget med i en ordbog. En ordbog indeholder ikke hele det danske ordforråd, kun de vigtigste ord. Vores ord-

dannelsesregler giver os mulighed for at danne ord efter regler på sammen måde som vi kan danne sætninger efter grammatikkens regler. En grammatik har ikke alle sætninger med ligesom en ordbog heller ikke har alle ord med, og i en ordbog udelukker man de ord som er regelret dannede. Et gammelt citat fra Peder Hjort (1793-1871, dansk kritiker og skolemand) kan illustrere dette:

Der er noget i vor Danskhed, som endogså kan have såre godt af noget Samqvem med Tydskheden. Man kunde kalde det Jydskheden. Men om dette Navn skulde krænke, kunde der også tales om Kjøbenhavnskheden og Indbildskheden og Overfladiskheden og flere andre *-heder*.

Når kort tillægsform bruges som tillægsord, kan *-hed* også benyttes: *beklemthed, forstemthed, nedringethed, blankslidthed, indadvendthed*. Også ved fremmedord er *-hed* helt anvendelig: *defekthed, bestialskhed, adækvathed, fairhed, pauverhed, raffinerethed*.

I ord som betegner antal, er *-hed* også brugbar: *enhed, trehed, flerhed, femhed, mangehed* – og *intethed*.

Orddelen *-hed* finder vi også i ord som betegner en person eller en ting, fx *berømthed*,

personlighed, skønhed, uskyldighed, nyhed og ofte i titler *højhed, velærværdighed, højvelbårenhed*.

Når tillægsordet ender på *-e*, falder *-e*'et bort: *ringhed* (af *ringe*), *stilhed* (af *stille*), *ægt-hed* (af *ægte*).

PJ

AÅJ	Anita Ågerup Jervelund
ESJ	Eva Skafte Jensen
IEM	Ida Elisabeth Mørch
JNJ	Jørgen Nørby Jensen
JS	Jørgen Schack
MHA	Margrethe Heidemann Andersen
PD	Philip Diderichsen
PJ	Pia Jarvad
SKA	Sabine Kirchmeier-Andersen

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 opsiget dit abonnement midt på året, vil du stadig modtage og betale for resten af årets udgivelser.

Varighed og opsigelse

Abonnementet løber indtil du opsiget 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 opsiget 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

2015/2 · juni

Fra Du er fyret! til Du er sindet! 1

Talesyntese og talegenkendelse 6

Sprognævnets hjemmeside 12

Fortjener din tekst en pris? 12

NYE BØGER

Nye medier. Sprogbrug, data og metode 13

SPØRGSMÅL OG SVAR

Orddelen *-hed* 14

Næste nummer udkommer i september 2015

Ansvarshavende redaktør: Sabine Kirchmeier-Andersen

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

www.sproget.dk