

H.C. Ørsted som leverandør til dansk sprogteknologi

Peter Juel Henriksen
Dansk Sprognævn

1. Indledning

Hans Christian Ørsted opdagede elektromagnetismen for 201 år siden, og i jubilæumsåret sidste år blev han fejret på alle de måder den moderne medieverden kan gøre det.

Nu var elektromagnetismen, ifølge videnskabshistorikerne, formentlig blevet opdaget alligevel nogle få år senere (af Faraday, Joseph Henry eller en anden samtidig fysiker). Opdagelsens mulighed lå i tiden, eller ligefrem dens nødvendighed. Man kan derfor mene at vi gør Ørsted uret hvis vi i fejringen af hans geni fokuserer for skarpt på lykketræffet i auditoriet d. 21/4 1820; så her vil vi i stedet hylde den kreative lingvist og terminolog. Ørsted var den udadvendte tænk, han ville dele sin naturforståelse med så mange som muligt. Som Niels Bohr 100 år senere var han overbevist om at vanebundet sprog er en barriere mod ny forståelse af naturfænomener, og derfor arbejdede han systematisk med at forny det danske ordforråd inden for æstetik, naturvidenskab og filosofi så centrale begreber kunne få nye betegnelser fri for forældede konnotationer. "Næsten alle Videnskaber ere i de senere Tider blevne berigede ved nye Opdagelser, nye Tanker og Tankevendinger, hvortil Romersproget intet Udtryk har", skrev han i 1837 (Nielsen 1981 I:4). De nye ord skulle dannes af dansk morfologisk materiale, meget gerne med indlån fra tidligere sprogstadier og fra de øvrige skandinaviske sprog inklusive islandsk. Listen over Ørsteds nyord tæller mindst 3274 lemmaer (ikke kun

2000, som det ofte hævdes).¹ Nogle har slået rod og bruges stadig i dag (fx *LYSGIVER*, *HÆMVÆRK*), andre blev glemt, men er siden blevet genopfundet, og somme tider i nye betydninger (*LUFTFART*, *ALRUM*), atter andre findes nu kun som fossiler (*JÆVNSTRØG*, *MUNDTRO*). Ørstedes nyord er smukke at læse og gode at tage i munden, alt for værdifulde til at gå i glemmebogen.

2. Glemte ord som ressource for nutidens sprogfolk

Hvem har brug for ord der er morfologisk velformede, ontologisk velbegrundede, men aktuelt har ukendt betydning? I sprogteknologien benævnes sådanne ord ‘out of vocabulary’ (OOV), og de spiller en stor rolle i afestning af sprogværktøj af forskellige typer.

- Stavekontrollen skal kunne analysere ukendte komposita som er regelmæssigt dannet og bøjet,
- PoS-taggeren skal kunne genkende og isolere affikser og andre morfologiske komponenter,
- talegenkenderen skal rekonstruere et fonologisk muligt ord ud fra slørede akustiske observationer,
- syntesestemmen skal læse fornuftigt hen over et ukendt ord uden at bryde tekstens sammenhæng,
- den virtuelle assistent skal kunne udlede en sandsynlig betydning af nydannede ord og vendinger.

De fleste applikationer har en orddatabase indbygget som dækker en stor del af det danske ordforråd, men som alle leksikografer ved, kan ingen ordbog blive komplet – alene fordi sprogsamfundet ikke er homogen. Visse ordgrupper vil være stort set ukendt uden for bestemte kredse af sprogbrugere, fx fagord, navne, ungdommens ord, samt ord på vej på pension. I modsætning til en klassisk ordbog kan en sprogteknologisk applikation ofte gøre noget fornuftigt med ord den ikke genkender, især når de er dannet morfologisk regelmæssigt og bøjet analogt med velkendte ord (fx komposita som *børnefjeder*,

¹ Tallet er endda væsentligt højere hvis man medregner de ordforslag som HCØ diskuterer, men uden at anvende uden for diskussionen selv. Ørstedes nyord (registreret i Nielsen 1981) skrives med *KAPITÆLER* i denne artikel. Alle andre nyord, heriblandt forfatterens egne, skrives i *snedfont*.

avlskrager, nanochok, henvalgte). Sværere bliver det når ordene består af en blanding af kendte og ukendte morfemer, fx fremmedsproglige indlån, for så bliver selve analysens mulighed usikker. Hvilke kriterier skal ord som *venkata*, *vefa* og *valton* klassificeres efter? Er de overhovedet ord?²

Heldigvis ønsker de fleste sprogbrugere at blive forstået, så de fleste nydannede ord har en struktur der kan afkodes systematisk. Derfor bliver nye sprogværktøjer især testet af mod OOV-ord der er dannet morfologisk og fonotaktisk regelmæssigt. Et OOV-materiale til afprøvningsbrug kan man skaffe på flere forskellige måder. I den interviewundersøgelse som Dansk Sprognævn gennemførte i 2017-19 (Kirchmeier et al 2019) blev tre indsamlingsmetoder nævnt: ad hoc (dvs. uden formelt princip), kopiering fra uafhængige kilder, samt automatisk genererede nonsensord. Lad os undersøge dem enkeltvis.

Det mest benyttede princip er tilsyneladende slet intet. Mange sprogteknologiske applikationer sendes til marked kun testet på OOV-ord der, groft sagt, er rystet ud af ærmet uden opmærksomhed om sproglig dækning og systematik. Det er en risikabel metode. Man kan næppe forudse hvordan ens applikation vil klare sig på ukendt input ved selv at simulere det ukendte. Flere firmaer peger da også på dette catch, men nævner samtidig at den typiske kravspecifikation ikke går i detaljer om afprøvningsmetoder og -materialer.

Disse eksempler blev stillet til rådighed af et dansk firma:

- to nye *flumser*
- jeg *blasper* jævnlige
- fem *gryvsomme* forslag
- hun blev *dylingvis* fyret

- to nye *peberkjoler*
- jeg *overspadserer* jævnlige
- fem *makreltrætte* forslag
- hun blev *køligvis* fyret

² *Venkata*, *Vefa* og *Valton* er alle godkendte drengenavne i Danmark per oktober 2020.

De første fire OOV-ord er ukendte morfemer mens de sidste fire er ukendte sammensætninger. Bemærk hvordan de ikke-analyserbare ord i første serie virker morfologisk flertydige i sig selv, men entydige i en syntaktisk kontekst. Dette firma var opmærksomme på at generere OOV-ord efter flere forskellige orddannelsesregler for at sikre bredden i afprøvningen. De indrømmede dog samtidig at OOV-materialer lavet ad hoc er tilbøjelige give et for optimistisk testresultat fordi man, bevidst eller ubevidst, kommer til at vælge ordformer som ens applikation er godt forberedt på.

En mere principfast afprøvning satser i stedet på uafhængige ordkilder. Et andet firma anvendte fx Ikeas produktnavne ved afestning af dansk taleteknologi: *vardagen*, *boholmen*, *edsvik* osv. Her er ulempen at eksterne ordsamlinger i reglen er dannet ud fra kriterier som fx æstetisk, humoristisk eller poetisk virkning, semantisk domæne (i Ikea's tilfælde: svenske stednavne) eller anden systematicitet som er irrelevant for en afestning. Desuden vil eksterne ordkilder ofte være skævt fordelt over ordklasser. Flere danske firmaer bruger fx OOV-lister af varebetegnelser eller esoteriske fagord der næsten udelukkende består af substantiver. Et sidste problem, som Ikea-listen kan illustrere, er at eksterne ordkilder let kan medføre en uventet tendens i afestningen (performance bias). Hvis man fx ønsker at afprøve Google's talegenkender, giver OOV-ord som *lardagen*, *hobolmen* og *ydsvik* et ganske realistisk indtryk af genkenderens duelighed i mødet med det ukendte ord. Prøver man derimod med *vardagen*, *boholmen* og *edsvik* – som fonetisk set er helt analoge – bliver man ikke bare genkendt med ekstremt høj sikkerhed, men prompte ekspederet videre til www.ikea.com.

I stedet for eksterne data kan man vælge at generere OOV-materialer med automatiske metoder. Det er ikke vanskeligt at lave ukendte danske ord med en algoritme der bygger på tilfældighed og samtidig anvender de almindelige regler for orddannelse. Mange statsansatte vil nikke genkendende til den slags ord. Når man logger på Statens IT-plattform, får man af sikkerhedsgrunde et nonsensord sendt til sin mobiltelefon, som man derefter returnerer fra sin computer. Så er der givet håndslag. Disse kodeord er konstrueret så de ikke findes i eksisterende ordbøger, men dog (med lidt god vilje) kan udtales, dvs.

følger dansk fonotaks. Her er nogle eksempler på statslige kunstord³ høstet i efteråret 2020: *takpus*, *isprop*, *svedsi*, *tyktap*, *sexypy* og – man blues næsten – *hunpik*. Naturligvis kan man tackle opgaven mere seriøst og generere OOV-materialer som dækker dansk orddannelse på en klogere måde, men så længe ordene er afledt af regler og principper som sprogingeniøren selv har formuleret, er der altid risiko for at afprøvningen giver falsk optimisme. Testen kan lykkes, men applikationen dør alligevel – i mødet med den virkelige verden.

For at samle op: Det ideelle OOV-materiale skal afsløre og belyse de fejl som sprogteknologiske værktøjer er særligt tilbøjelige til at begå ved mødet med ikke-leksikaliserede ord – altså give stof til djævelens advokat. På baggrund af samtaler med danske sprogingeniører (og egne erfaringer) kan vi opstille disse minimumskriterier for den gode OOV-liste. Eksemplerne i parentes er kun til illustration.

Morfologisk diversitet

- alle åbne ordklasser
- produktive og mindre produktive orddannelsesprincipper (*hestesang*, *kongegul*, *bagkikke*)
- regelmæssige og uregelmæssige bøjninger (*frahviskede*, *fraspurgte*)
- flertydig morfemstruktur (*træ-klam/træk-lam*)

Udtalemæssig diversitet

- lydrette og ikke-lydrette former (*limfilister*, *vejrbagage*)
- flertydig fordeling af hovedtryk (*selvhændigste*, *forudnævnelige*)
- flertydig selektion af vokalkvalitet, længde og stød (*smagslag*, *korside*)

Sjældne former

- komposita med ukendte morfemer (*lummekone*, *underhjure*)
- komposita med uventet fuge (*sekundsviser*, *månedkalender*)
- komposita fra mindre produktive orddannelsesregler (*længereglidende*, *hanøl*, *halyfemsoste*)
- dvandvacomposita (*tyskindisk*, *blindstum*)
- monomorfemiske nonsensord (*slaf*, *gurba*)

³ OOV-ord fra Statens-IT bliver, i sagens natur, kun brugt én gang. Derfor er der ingen sikkerhedsrisiko ved at dele dem efterfølgende (og more sig over dem).

Udvalg

- mindst 1000 lemmaer (flere er bedre)
- mange og varierede eksempler i hver kategori herover
- kun reelt ukendte ord, dvs. lemmaer uden forekomster selv i meget store tekstkorpora

I næste afsnit undersøger vi hvordan Ørstedes nyord ville klare sig som OOV-materiale.

3. Ørstedes nyord som OOV

Ørstedes ordliste er imponerende både i omfang og variation. Vi har lydefterlignende ord (KNALDLUFT, FORKNISTRE), billedskabende (ØRKENFØLELSE, HALESTJERNE), filosofiske (TANKEHARMONI, NATURTEMPEL), dramatiske (SANDHEDSMYRDEN, SEJRSPRAGT), politiske (MEDBORGERSIND, MAGTFULDKOMMENHED), psykologiske (NERVESTEMNING, SJÆLEANLÆG), poetiske (MILDTÅNDENDE, BEGEJSTRINGSBLUE), muntre (ORDKRÆSEN, SLAGTERKRAFTIG) og også satirisk-spids (TRANGHJERNET, KÆTTERSNUSER). Ørsted kunne være en kras debattør når han fandt grund til det. Men først og fremmest har vi naturligvis en stor mængde fagord, især kompositionelle dannelser, altså ord hvis betydning kan afledes systematisk af orddelene og deres rækkefølge i analogi med allerede eksisterende ord. Ordene *adfærd*, *velfærd*, *fremfærd*, osv. fandtes allerede og tillod Ørsted at danne det nye OMFÆRD. Eksisterende ord som *enstemmig*, *flerstemmig*, *samstemmig* tillader SELVSTEMMIG. Ind imellem bruger Ørsted de velkendte affikser så kreativt at det udfordrer deres almindelige betydning, UTANKE, ALHED, FORFULDKOMNING, eller ligefrem nærmer sig grænsen for sprogrigtighed. Et godt eksempel er forstavelsen mis- der bruges ekstremt produktivt: MISSTOR, MISMINDE, MISMÆLE – endda MISÆGTESKAB.

Den del af Ørstedes nyordsarbejde der har betydet mest for eftertiden, er utvivlsomt hans kemiske og fysiske nomenklatur. I nogle tilfælde har Ørsted forklaret de formelle orddannelsesregler bag fagtermerne (det gælder især for de kemiske termer, se Nielsen 1981 I:7-40), men iøvrigt er reglerne så indlysende at de kan afkodes af ordene selv. Bemærk hvor let det er at udfylde de tomme felter i tabel 1 herunder.

NATER	NATERÆSK	NATERÆR		
	KALKÆSK	KALKÆR		
	BARYTÆSK	BARYTÆR		
		STRONTINÆR		
	JERNÆSK		JERNILTE	
		YTTERÆR	YTTERILTE	
		CIRKONÆR	CIRKONILTE	
			PLATINILTE	
KVÆLSTOF				KVÆLSTOFTVEILTE
	TINÆSK			TINTVEILTE
			KVIKSØLVILTE	KVIKSØLVTVEILTE

Tabel 1. Grundtermer og sammensætninger – et lille udsnit af Ørsteds nye kemiske nomenklatur med gennemført kompositionel orddannelse.

Tabellen illustrerer hvordan Ørsted gør op med det han kalder “en hovedfejl ved den kemiske nomenklatur i alle skandinavisk-germanske sprog”, nemlig det forhold at “navnene på de meste enkle legemer, af hvilke navnene på alle de øvrige enten måtte afledes eller sammensættes, allerede selv er sammensat af flere, og at disse er så ufuldkomment forbundet, at de mere synes at danne en definition end et ord.” (oversat fra latin af N.Å. Nielsen; ib. I:9). I modsætning til Ørsteds systematiske betegnelser var de fleste traditionelle kemiske termer beskrivende og henviste til sanseindtryk, findesteder, praktiske anvendelser eller var på anden måde uden morfologisk relation til hinanden. Ørsteds termer danner let nye sammensætninger, såsom ilte-*X*, *X*-iltelig, *X*-iltning, *X*-iltet, *X*-brintet osv., hvor *X* vælges frit (ILTESYRE, LETILTELIG, SVOVLBRINTET) og rekursivt (SELENBRINTILTESYRE, FOSFORJERNESALT, ILTESALTSYRESALT). Han er altid opmærksom på at nye ord skal dannes efter kendte regler; de må ikke “stride mod Sprogets Natur, være korte og saa vidt muligt ikke være vanskelige at udtale”, som han skriver i sin afhandling ‘Sprogforædling. Regler for Orddannelseindledningen og didhørende Bemærkninger’ (udg. posthumt; ib. I:65ff).

I sin nye rolle som OOV-materiale kan Ørsteds nyordssamling altså byde på en righoldig blanding af kendte og ukendte morfemer holdt sammen af et bredt udvalg af orddannelsesprincipper hvoraf

langt de fleste er helt traditionelle mens nogle nærmer sig grænserne for “Sprogets Natur”. Ordsamlingen går fra de korteste former til de længste, fra ÆR og ÆSK til SELVFORFOLDIGHEDSANGIVER og METAL-BEARBEJDNINGSUDSTRÆKNING, og synes på enhver måde at opfylde kriterierne for det gode OOV-materiale.

For at lette databehandlingen har vi samlet Ørstedes nyordsmateriale i en database kaldet HCOe_nyord.01 og gjort den tilgængelig for download (mere information sidst i artiklen).

Ørstedes nyord (HCOe_nyord.01)

Kvantiteter

- 3274 lemmaer
- 293 lemmaer forekommer i Retskrivningsordbogen og/eller Den Danske Ordbog (fx OPDRIFT, RUMFANG, BRAMFRI)
- distribution over åbne ordklasser:
 - appellativer: 2699 lemmaer (utrum: 2018, neutrum: 678, uden singularis: 3)
 - verber: 58 lemmaer
 - adjektiver: 506 lemmaer
 - adverbier: 10 lemmaer (MILLIONFOLD samt sammensætninger med -gængs/-vis/-strøms/-strøgs)
 - øvrige: 1 (HID- OG DID-TÆNKEN)

Varieret orddannelse

- monomorfemiske nydannelser (ILT, SVÆV, NATER)
- migration over ordklassegrænser (subst. ÆSK, verb. ÆSKNE, adj. ÆSKIG, ÆSKIGTSMAGENDE)
- regelmæssige komposita (TALLOV, RØDKÅLSVINÅND)
- regelmæssig affigering (TEELSE, UDELBAR, FORUDBETINGE)

Morfologiske udfordringer

- komposita med uventet fuge (SKIKKELSELØS, FAGSMAND, TUNGLØSELIG)
- komposita med lumske orddele (TANÆR, MINDSTILTELIG, ÆTSENDHED)
- dvandvakomposita (YDMYGFORNEM, FORSTANDIGSANSELIG)

Fonetiske udfordringer

- syllabisk flertydighed (AMMONÆSK, LERÆRAMALGAM)
- trykfordeling (GLYCINÆR, KULFORSVOVLBRINTET)
- vokalselektion (HULRETHED, FORVEJRET)

Som nævnt er mange af Ørsteds nyord stadig i brug i moderne dansk. Det eksakte antal er dog ikke let at afgøre da det afhænger kritisk af i alt fald to forhold: dels den ortografiske justering af leksemerne og dels deres semantiske forskydning. I databasen HCOe_nyord.01 er der indført en lang række ortografiske normaliseringer (de oprindelige stavemåder er dog også bevaret). Først og fremmest er dobbelt-a erstattet med bolle-å overalt, og appellativer stavet med lille begyndelsesbogstav; derefter er der indført en lang række betydningsneutrale justeringer af de enkelte morfemer: jærn→jern, hem→hæm (fx HEMVÆRK→HÆMVÆRK), vexle→veksle, svovel→svovl osv. I dette arbejde har vi kunnet følge hovedprincipperne i Becker-Chrisensten (2021), især 2.3-2.6. Men hvor langt skal man fortsætte? Ørsted skriver fx både FOSFORBRINT og FOSFORBRINTE, VARMEMÅLING og VARMEMÅLNING. Skal de anses for ortografiske eller leksikalske varianter af det samme lemma? Hvad angår den semantiske forskydning er afgørbarhedsproblemet endnu større: I hvilken grad er et nyord der genfindes i fx Den Danske Ordbog, men i en ændret betydning, stadig 'det samme'?

Vi kan dog konstatere, selv med disse forbehold, at langt over 100 af Ørsteds nyord stadig lever i bedste velgående.

4. To demonstrationer

Her viser vi to små eksempler på sprogteknologisk brug af Ørsteds nyord som OOV-materiale, dels som input til PoS-tagging og dels til automatisk lydskrivning.

En PoS-tagger er det der i AI-terminologi hedder en *classifier*, en automat som læser et input (i dette tilfælde et ord eller en række af ord) og klassificerer hvert element i forhold til et forud givent udvalg af kategorier (ordklasse og flektion). PoS-taggeren kan være trænet til kun at kende hovedordklasserne (N, V, ADJ, ADV, PRON, PREP osv.), men som regel ønsker man en mere finkornet morfologisk analyse. Vi har udfordret tre af markedets bedste danske PoS-taggere

med forskellige former af adjektivet FLYGIG (‘flygtig’) og verbet at ENSTERSÆTTE (‘isolere’).

<i>input til taggerne</i>	PoS-tagger A	PoS-tagger B	PoS-tagger C
FLYGIG	N_INDEF_SING (?)	N_INDEF_SING (?)	ADJ_INDEF_SING (+)
en FLYGIG materie	ADJ_INDEF_SING (+)	ADJ (+)	N_INDEF_SING (-)
stofferne er FLYGIGE	ADV (-)	N_INDEF_SING (-)	N_DEF_PLUR (-)
materien er ikke FLYGIG	ADJ_INDEF_SING (+)	N_INDEF_SING (-)	ADJ_INDEF_SING (+)
alkohol er FLYGIGERE end vand	ADJ_COMP (+)	N_INDEF_PLUR (-)	ADJ_COMP (+)
ENSTERSÆTTE	N_INDEF_SING (?)	N_INDEF_SING (?)	N_INDEF_SING (?)
at ENSTERSÆTTE eleverne	V_INF (+)	V_INF (+)	V_INF (+)
man ENSTERSÆTTER eleverne	N_INDEF_SING (-)	V_IMP (-)	N_INDEF_SING (-)
man ENSTERSATTE eleverne	ADV (-)	V_IMP (-)	V_INF (-)
man har ENSTERSAT eleverne	V_PARTICIP_PAST (+)	V_PARTICIP_PAST (+)	V_PARTICIP_PAST (+)

Tabel 2. Ørstedes nyord FLYGIG og ENSTERSÆTTE (med og uden kontekst) analyseret for klasse og form af tre danske PoS-taggere. Analyserne er bedømt for korrekthed: (+) = korrekt, (?) = tvivlsom, (-) = ukorrekt. De tre taggere er præsenteret anonymt da formålet her er demonstration, ikke evaluering.

Alle tre PoS-taggere klarer sig bedre end *chance* (den tilfældige klassifikation), som det fremgår, men ingen klarer sig så godt som en menneskelig morfolog med samme udgangspunkt (jf. bedømmelserne i parentes). Vigtigere i denne sammenhæng er det dog at tabeldata af denne type kan bruges som et nyttigt redskab til videreudvikling af PoS-taggere. Man ser tydeligt i tabel 2 hvor taggerne A-C har deres stærke og svage sider.

Lydskrivning er en langt vanskeligere disciplin end PoS-tagging. En automatisk lydskrivningsalgoritme er, i AI-termer, ikke bare en classifier men en *transducer* (en afbildningsalgoritme). Hvor PoS-taggeren kan træffe sin beslutning ud fra relativt få træk af et sprog, skal lydskriveren kende hvert eneste morfem. En lydskriver vil ofte,

lige som den menneskelige transskriptør, stå med en lang række mulige fonetiske afbildninger af det samme ord. Dette gælder ganske særligt når ordet forekommer i en kontekst, fx i en sætning, fordi ordenes lydlig realisering varierer systematisk med deres syntaktiske og pragmatiske funktion. Lydskrivningsalgoritmens mål er altså ikke en korrekt lydskrift i snæver forstand, men derimod en fonetisk form som – realiseret af en god syntesestemme – vil virke upåfaldende i en given sammenhæng. For at bedømme en lydskrivners kvalitet bør man derfor kende et menneskes forventning. Vi har testet Dansk Sprognævnets lydskrivningsalgoritme på nogle af Ørsteds mere drilske ord; til sammenligning har otte af Sprognævnets egne medarbejdere indtalt de samme ord, og optagelserne er derefter blevet transskriberet manuelt.

<i>HCØ's nyord</i>	Informanters udtale (i udvalg)	Sprognævnets lydskriver
FLUGLIT	[fl"u:lid] [fl"uglid] [fl"åglid]	[fl"u:lid]
SKELFUNDIG	[sg"ælfân?di] [sg"ælfândi] [sgælf"ân?di]	[sg"ælfândi]
LERÆRAMALGAM	[ler"ær?amalgAm] [l"eRær? amalgA:?m] [l"eR ær? amalgAm?]	[l"ærær?amalgA:?m]
TUNGSMELTELIGHED	[tânsm"æl?d0liheÐ?] [t"ân?dsmæld0lihe:~Ð] [t"ân?smæld-liheÐ?]	[t"ânsmæl?d0liheÐ?]
CIRKONILTE	[siRkon"ild0] [siRk"o:~nild0] [s"iRkCnild0]	[siRkCn"ild0]
KULFORSVOVLBRINTET	[k"ålfCsv"Owlbrænd0Ð] [kålf"O:svOwlbrænd0Ð] [kålfCsvOwlbr"ænd0Ð]	[kulfCsv"Owlbræn:d0Ð]

Tabel 3. Otte DSN-medarbejders udtale af udvalgte HCØ-nyord, manuelt transskriberet; de samme ord ifølge Sprognævnets automatiske lydskriver. Lydskriften er i dansk SAMPA (IPA-ækvivalent til sprogteknologiske anvendelser). Vi bruger [ˈ] for hovedtryk; [ʔ] for stød; [0] for schwa; [-] for assimileret schwa; [C] for fuldvokalen i ‘komme’; [O] for fuldvokalen i ‘borte’; [a:] for fuldvokalen i ‘bane’; [l] for stavelsessegregering.

Som det ses i tabel 3, er Ørsteds ord ikke altid trivielle at udtale selv for trænede sprogfolk, også selv om ordene i overensstemmelse med Ørsteds egne principper er fuldkommen regelmæssigt dannede, både i form og betydning. Er man ikke bekendt med morfemet *ÆR*, vil skanderingen af *LERÆRAMALGAM* være fuld af flertydigheder som skal udredes før man kan træffe en beslutning om udtalen. En af informanterne var næsten 10 sekunder om at beslutte sig for udtalen af *LER-ÆR-AMALGAM*, og ingen klarede det hurtigere end 3 sekunder. Selve forsinkelsen af responsen er i sig selv værdifuld som måledata, fordi den sætter et tal på den mentale belastning. Forsinkelsesmålinger bliver ofte brugt i kognitionsforskning, dog mest til visuelle opgaver som ansigts- og billedgenkendelse. Ørsteds nyord, især de længere komposita, giver gode muligheder for at høste forsinkelsesdata i det lingvistiske domæne, med deres komplicerede miks af kendte og ukendte dele.

5. Til ære og værdighed

Sprognævnet mødte, som før nævnt, en lang række sprogteknologer i forbindelse med vores udredningsarbejde i 2017-19. En af de tilbagevendende kommentarer var at den afsluttende projektfase – der hvor man kvalitetssikrer sit produkt før levering – er den der oftest bliver forsømt. Man når kun en hastig afprøvning med ad hoc-data som ikke giver et realistisk indtryk af applikationens robusthed over for det uventede. Det kan hurtigt ende galt når applikationen bliver udsat for de nye ordmutationer der ustandseligt opstår i et levende sprog. Vi håber og tror at Ørsteds nyordsliste her kan gøre nytte som vaccine.

Ørsted er blevet kaldt både sprogrenser og sprogfornyer. Hans tenet er at morfemer med utidsvarende konnotationer, særligt de klassiske græske og latinske stammer, skal erstattes med rationelle nydannelser; men faktisk går han langt videre og forsøger at udrense *al* fremmed indflydelse. Somme tider går han til excesser og skaber lange og komplicerede danske komposita til afløsning af hvad der virker som ret harmløse fremmedord, såsom *IKKETILBAGEKASTNING* for ‘absorption’, *PÅHINANDENFØLGEN* for ‘sekvens’ og *UIGENNEMTRÆNGELIGHEDSKRAFTIG* for ‘massiv’. Renhedsidealet gælder ikke kun indholdsmorfemer,

men også bøjninger og affikser. “Man maa forsøge at erstatte det tydske *mæssig* i Ordenes Sammensætning”, skriver han (Nielsen 1981 II:84) og foreslår i stedet *lundig*: FORHOLDSLUNDIG, HENSIGTSLUNDIG, REGELLUNDIG, HENSYNSLUNDIG, MIDDELLUNDIG, FORSTLUNDIG. I dette konkrete tilfælde må man nok konstatere at Tydsken vandt; den nyeste udgave af Retskrivningsordbogen har 43 adjektiver der ender på -mæssig og ikke ét der ender på -lundig. Ørsted opfatter altså fremmedord som forurening af det danske ordforråd. “Det er velkendt for mig, at mange er gunstigt indstillet til denne brug af udenlandsk nomenklatur, fordi de tror, at de ved at fremme sprogenes fællesskab med andre folk gavner mere, end de skader ved at gøre modersmålet urent. Men i kemien er fordelene ikke så stor, som de tror; thi i denne videnskab afledes størstedelen af kunstordene af meget få stamord, således at den, der har skaffet sig kendskab til kemien gennem bøger på et eller andet sprog, skulle kunne lære et andet sprogs vel sammensatte kemiske nomenklatur i løbet af to timer.” (ib. I:11, overs. fra latin).⁴ Historien om den danske sprogrensning på Ørsteds tid er godt fortalt i Skautrup (1953), se især §74.

Men vigtigere end Ørsteds tidstypiske purisme – som vi jo også kender i vores egen tid, hvor importord som *disclaimer* og *gamechanger* irriterer somme – er hans refleksioner over terminologiske grundprincipper. Ørsted argumenterede for at udmøntningen af nye fagord skal undgå det deskriptive og kun være repræsenterende. Hvis man knytter en terms kompositionelle logik for tæt til dens ønskede denotation, for eksempel ved at navngive tingen med en henvisning til det som aktuelt ligner dens afgørende egenskab, kan et nyt ord alt for hurtigt blive forældet, ganske særligt i en tid præget af videnskabelige nybrud. Udtrykt i maksimens form:

Nye termer for videnskaben skal betegne og ikke beskrive

I det store og hele følger Ørsteds nyordsproduktion hans egen maksime (Nielsen 1981 I:13 et pass.), dog ikke konsekvent. Blandt undtagelserne finder vi nogle af hans allermest holdbare nyord,

⁴ Samme tankegang lå bag vores grammatikeres fordanskninger. Uforståelige latinske fagord blev erstattet med uforståelige danske fagord, som børnene hører som *udsangsord*, *egenavne*, *kort telexform* osv.

ILT, BRINT, KVÆLSTOF og deres afledninger såsom ILTNINGSGRAD og AFBRITE.⁵ Forud for Ørsteds tid mente man at oxygen ('syredanner') var det konstituerende element i syrer mens hydrogen ('vanddanner') var den centrale komponent i vand, og nitrogen ('salpeterdanner') i kalisalpeter. Den dag i dag hedder grundstof nr. 8 stadig *syre* på svensk og *sauerstoff* på tysk. På Ørsteds tid var det forlængst blevet påvist at ikke oxygen men *hydrogen* er den afgørende konstituent i syrerne, og det blev klart at termene var forældede. Ørsteds nyord ILT skulle i stedet associere til 'ild' da han anså grundstoffets tilstedeværelse som forudsætningen for en forbrænding. Imidlertid er det ikke kun dette stof der nærer ild og flammer, det gør også andre grundstoffer (såsom fluor), og tilsvarende er det ikke kun BRINT der kan brænde og ikke kun KVÆLSTOF der kan kvæle en forbrænding. Ingen af navnene har den nødvendighed over sig som Ørsted gik ud fra. Ironisk nok er hans mest populære nyord altså blandt de undtagelser som bekræfter hans egen maksime.⁶

Ørsted havde sine idiosynkrasier og gjorde sine fejltrin; men er det seriøse indvendinger imod ham som tænker og inspirator? Så langt fra. Hans oeuvre er en vedvarende energikilde, ikke kun for naturvidenskaberne, men også for os sprogfolk. Ørsted var – med et af sine egne nyord – en ILDSJÆL.

Ørsteds nyord er er tilgængelig som orddatabasen HCOe_nyord.01 med oplysninger om oprindelig bogstavering, moderne bogstavering, ordklasse og bøjningsform, forekomst i kurante ordbøger, morfologisk analyse og (forslag til) udtale. Databasen kan downloades fra dsn.dk/Oersted (eller andetsteds som annonceret i dsn.dk). Ordstoffet er afskrevet og korrekturlæst fra Nielsens HCØ-ordbog (Nielsen 1981 II 3-194). En varm tak til Johanne Niclasen Jensen for dette arbejde. Det Kgl. Videnskabernes Selskab har i anledning af jubilæumsåret 2020 genudgivet Nielsens ordbog, se www.royalacademy.dk/da/Aktuelt/Orsted-ordbog

⁵ Ørsted skabte ordene til brug i sit hovedværk *Kraftlæren*, som fortalt i Christensen (2009) kap.28, se især I:458.

⁶ Rasmus Rask var kritisk over for nyordene ILT og BRINT af rent morfologiske grunde. Han påpegede at den bagvedliggende orddannelsesregel (at danne nye substantiver ved at tilføje -t) var ukendt i dansk; se Gregersen (2019).

**Hent mine ord på
dsn.dk/Oersted**



Litteratur

- Becker-Christensen, Christian (2021) *Dansk Retskrivning, Bogstav – Lyd – Bogstav*. SDU Studies in Linguistics vol. 27, Dansk Sprognævns Skrifter 51, Syddansk Universitetsforlag.
- Christensen, Dan Ch. (2009) *Naturens Tankelæser*. Bd. I og II. København: Museum Tusulanums Forlag.
- Gregersen, Frans (2019) *H.C. Ørsted og de nye ord*. Kvant, 30/4:30-34.
- Kirchmeier, Sabine; Peter Juel Henriksen; Philip Diderichsen; Nanna Bøgebjerg Hansen (2019) *Dansk Sprogteknologi i Verdensklasse*; Rapport fra Sprogteknologiudvalget under Dansk Sprognævn, ISBN 978-87-89410-77-7 (pdf kan downloades fra dsn.dk).
- Nielsen, Niels Åge (1981) *Sprogrenseren H.C. Ørsted*. Bd. I og II. Aarhus Universitet: Trøjborg.
- Skautrup, Peter (1953) *Det Danske Sprogs Historie*, bd. III; Gyldendalske Boghandel: Nordisk Forlag.