

# Glemte ord

## En undersøgelse af H.C. Ørsteds nyord og deres plads i nudansk

PETER JUEL HENRICHSEN

### ABSTRACT

Vores fejrede fysiker H.C. Ørsted var også aktiv som ordmager og skabte flere tusind nye danske ord, såvel videnskabelige som almene. Ørsteds nyord har utvivlsomt sat varige spor i det danske sprog – men hvilke, hvor mange og hvor dybe? Det undersøger vi i denne artikel. Først præsenteres Ørsteds ordstof og hans principper for orddannelse, i en historisk kontekst; derefter udvikles et kvantitativt mål for ords kognitive identitet i et levende sprogsamfund, og på dette grundlag gives et samlet billede af Ørsteds indflydelse på nudansk. Til slut samles temaer og resultater op i en diskussion om modellering af en sprogtilstand – om de formelle metoders begrænsninger, men også om deres uudnyttede muligheder i den nyeste lingvistikens tjeneste.

**EMNEORD:** H.C. Ørsted, sprogtilstand, kategorialgrammatik, morfologi, korpuslingvistik

## 1 INTRODUKTION

Ørsted skabte flere tusind nye danske ord, kraftudtryk, sproglomster, skældsord, onomatopoietika og – hvad han især blev berømt for – en stor mængde rationelle termer for naturvidenskaben. Mange af ordene lever endnu. Det er velkendt at Ørsted opfandt *ilt*, *kvælstof*, *rumfang* og *lynafleder*, men måske mindre kendt at han også stod fadder til *billedkunst*, *samfundsforhold*, *bramfri* og *ildsjæl*. Hovedparten af Ørsteds neologismer er dannet ud fra kendte morfemer og orddannelsesregler og kan let forstås i dag: *beviskraft*, *tvivlmager*, *meningsadvokat*; andre er umiddelbart uforståelige, som *æsk* og *blæel*, men kan alligevel fornemmes som meningsfulde i sammenhænge som *æskigtsmagende* og *blæelbrintesyre* fordi vi stadig kender de bagvedliggende principper for orddannelsen, som også er en del af arven efter Ørsted. På flere planer har Ørsteds ordproduktion altså haft blivende betydning for det danske

sprog, så meget står fast. Men hvilke dele er bevaret, og hvilke er glemt? Formålet med denne artikel er at give en samlet vurdering af Ørstedes nyordsprojekt og dets indflydelse på den nudanske sprogtilstand.

I Ørsted-året 2020 holdt forfatteren et foredrag om Ørstedes nyord og strejfede blandt andet spørgsmålet “Hvilke af Ørstedes ord er bevaret i nudansk?”. Det var bare ment som et oplæg til en simpel undersøgelse og burde have været glemt igen; men spørgsmålet blev hængende, ja, spaltede sig i nye, som kosten i hænderne på troldmandens lærling. Hvad vil det sige at et ord er ’bevaret’ i en sprogtilstand? Skal et flertal af sprogbrugerne kende det – eller bare fagfolk og specialister? Skal det være i aktiv brug? Skal det være produktivt og danne nye komposita? Og sværest af alt: Skal det have bevaret sin betydning i dens fulde omfang?

I denne artikel giver vi forslag til hvordan disse store spørgsmål kan underdeles i mindre, som til gengæld kan finde konkrete svar. Hertil bruger vi metoder fra kategorialgrammatik og korpuslingvistik anvendt på Ørstedes nyord (som registreret i Nielsen 1981 II). Artiklen begynder med at præsentere Ørstedes samlede nyordsproduktion og, nok så væsentligt, grundlaget for de orddannelsesprincipper som Ørsted argumenterede for. Disse historiske kilder leverer materialet til den efterfølgende analyse, der munder ud i en samlet beskrivelse af Ørstedes nyord: hvilke lemmaer der genkendes umiddelbart, hvilke der forekommer uforståelige i dag, og hvordan resten er fordelt mellem disse to yderpunkter. Til sidst genbesøger vi de store spørgsmål herover – nu i lyset af undersøgelsens resultater.

Vi bruger metoder til morfologisk analyse der er ret usædvanlige i dansk sprogforskning, men alle nødvendige begreber bliver indført efterhånden som fremstillingen kræver dem (tekniske fodnoter kan springes over uden skade for sammenhængen). En del fagtermer er, efter datalingvistisk praksis, defineret formelt snarere end empirisk; de kan altså have let afvigende betydning i forhold til i den klassiske morfologiske litteratur (fx ’morfem’, ’stamme’, ’affiks’, ’fuge’, ’ord’).

Ørstedes nyord og de tilknyttede analyser har vi samlet i en maskinlæsbar database HCOe\_nyord.01, der hermed overgives til offentligheden. Mere herom sidst i artiklen.

## 2 SPROGFORNYEREN

Ørsted var den udadvendte tænkter, 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 (cit. Nielsen 1981 I: 4). De nye ord skulle dannes af dansk morfologisk materiale, meget gerne med indlån fra tidligere sprogstadier og fra de øvrige nordiske sprog. Nyordlisten tæller mindst 3274 lemmaer (ikke 2000, som det ofte hævdes<sup>1</sup>), og variationen er enorm. Der er lydefterlignende ord (KNALDLUFT<sup>2</sup>, FORKNISTRE), billedskabende (ØRKENFØLELSE, HALESTJERNE), filosofiske (TANKEHARMONI, NATURTEMPEL), dramatiske (SANDHEDSMYRDEN, SEJRSPRAGT), politiske (MEDBORGERSIND, MAGTFULDKOMMENHED), psykologiske (NERVESTEMNING, SJELEANLÆG), poetiske (MILDTÅNDEDE, BEGEJSTRINGSLUE), muntre (ORDKRÆSEN, SLAGTERKRAFTIG) og også satirisk-spidse (TRANGHJERNET, KÆTTERSNUMSER). Ørsted kunne være en kras debattør når han fandt grund til det. Men først og fremmest har vi en stor mængde fagord, især kompositionelle dannelser, altså ord hvis betydning kan afledes systematisk af orddeles rækkefølge og semantik i analogi med allerede eksisterende ord. Ordene *adfærd*, *velfærd*, *fremfærd*, osv. fandtes allerede og tillod Ørsted at danne det nye OMFÆRD, mens SELVSTEMMIG blev afledt af *enstemmig*, *flerstemmig*, *samstemmig* osv. Ofte brugte Ørsted de velkendte affikser så kreativt at det udfordrer deres almindelige betydning, UTANKE, ALHED, FORFULDKOMNING, og somme tider nærmer ordene sig ligefrem grænsen for sprogrigtighed. Et eksempel er præfikset *mis-*, som Ørsted brugte ekstremt produktivt: MISMINDE, MISMÆLE, MISSTOR – endda MISÆGTESKAB. Ørsted veg heller ikke tilbage for at opfinde

1 Tallet er endda højere hvis man medregner de ordforslag som HCØ diskuterer, men uden at anvende uden for diskussionen selv.

2 Herfra bruger vi KAPITÆLER til Ørsteds nyord (som registreret i Nielsen 1981 II). Proprier forekommer ikke i NÅN's register og derfor heller ikke i denne undersøgelse.

nye produktive affikser som fx endelsen -LUNDIG til erstatning for det tyskklingende *-mæssig*, især anvendt i fagtermer som FORHOLDSLUNDIG, REGELLUNDIG, MIDDELLUNDIG, men også helt frit som i FORSTLUNDIG. Kort sagt, Ørsted udnyttede stort set alle de morfologiske muligheder for orddannelse som det danske sprog tilbød. Plus nogle flere.

FIGUR 1. ØRSTEDS NYORDSSAMLING I OVERBLIK

### Regler for orddannelse

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

### Kvantiteter

- 3274 lemmaer
- 324 lemmaer har homografe forekomster i RO og/eller DDO
- distribution over ordklasser
  - appellativer: 2699 lemmaer (utrum: 2018, neutrum: 678, uden singularis: 3)
  - verber: 58 lemmaer
  - adjektiver: 506 lemmaer
  - adverbier: 10 lemmaer
  - desuden enkelte flerordsforbindelser

## 3 SPROGRENSEREN

Ørsted var skeptisk over for brugen af udenlandske morfemer i danske ord, men som purist var han langt fra så rigoristisk som sine samtidige (Skautrup 1953 §74). I det hele taget var det snarere manglen på præcision, konsekvens og udtrykskraft han vendte sin skarpe pen imod. Frem for alt var Ørsted yderst kritisk over for sin samtids fysiske og kemiske nomenklatur, og hans terminologiske reformarbejde blev et forbillede for alle eftertidens videnskaber. I dette afsnit introducerer vi det centrale princip i Ørstedes orddannelsesreform, i dag kendt som *kompositionalitet*, i sin historiske kontekst.

Kemikere og alkymlister fra antikken og frem til Ørstedes egen tid

havde efterhånden, ad mange omveje, fået inddelt naturens fysiske stoffer i grupper efter deres kemiske egenskaber.

FIGUR 2. DE FYSISKE STOFFERS TRADITIONELLE KLASSIFIKATION

- grundstoffer (fx jern, kul, svovl)
- deres forbindelser (kalk, eddikesyre, salpeter)
- legeringer (bronze, stål, messing)
- mineraler (kis, blende, granit)

Mange af stoffernes navne afspejlede deres rolle i menneskelivet såsom findested, tilvirkning, sansekvalitet og daglig brug (*chilesalpeter*, *trækul*, *blåsyre*<sup>3</sup>, *vinånd*). Andre havde navne hvis oprindelse fortabte sig i fortiden. De traditionelle navne var ukritisk blevet taget ind i det moderne kemiske laboratorium, hvad der gjorde den videnskabelige samtale besværlig. Tag som eksempel ordet *salt* der – som handelsvare – er det hvide pulver man inddamper af havvand eller graver ud af en horst. Kemisk betraget er salt, i denne betydning, et molekyle bestående af et atom natrium og et atom klor, NaCl, altså i kemisk forstand en metalafledning af klorsyren. Da salt nu var så velkendt et stof, fik det imidlertid lov til at navngive hele gruppen af, ikke bare klorsyrens, men vilkårlige syrer og vilkårlige metalafledninger. De hed allesammen *salte* (og hedder det populært den dag i dag). Af tilsvarende grunde er klorsyre kendt under det misvisende navn *saltsyre*, omtrent som hvis en grammatisk kategori blev opkaldt efter ét af sine elementer så verber hed *spiseord* og substantiver *støveord*. Denne vilkårlighed var udbredt over hele den kemiske nomenklatur så de enkelte stoffers navne ingen (morfo-) logisk sammenhæng havde med hverken hinanden eller med klassifikationen i figur 2. Det betød at nye opdagelser i kemien måtte doceres i ord der i bedste fald skulle terpes som ren udenadslære, i værste fald var direkte vildledende.

Ørsted udarbejdede en navngivningsmanual som udelukkede kemiske navne der udgik fra hverdagslivet, og erstattede dem med systematiske termer. En uigennemskuelig formulering som ”salt er et salt af saltsyre”, hvor homograferne nærmest vælter over hinanden, bliver i Ørsteds nye nomenklatur (ver. 1814) til ”naterklor er et salsel

3 Paradoksalt nok er blåsyre ikke blå, men farveløs.

af klorsyre”. Klart og kompositionelt. I Ørsteds kemiske nomenklatur er alle betegnelser dannet ud fra et basisinventar af morfemer, hvor grundstofferne selvfølgelig er fødte medlemmer ([klor], [tin], [svovl], [BRINT]). Dertil kommer samlebetegnelser for grundstofferne ([metal], [gas], [ÆR]) og for atomgrupper i molekylet ([kisel], [glycin], [BLÆL]) samt betegnelser for den kemiske konstituentstruktur ([syre], [SALSEL], [ÆSK]). Basissamlingen omfatter også et antal funktionsmorfemer (præfikser som [tve], [tre], suffikser som [ig] og [ling]). Det norrøne talord *tve* forekom allerede som alment præfiks i fx *tvedelt*, *tvekamp*, og Ørsted annekterede det i betydningen: to kemiske enheder. Det eksisterende diminutivsuffix *-ling* fik den formelle betydning: kemisk forbindelse med én enhed ilt mindre end i den iltmættede forbindelse – som i eksemplerne *SALSING*, *SVOVLING*, *SYRLING* (se figur 3).

Det grundlæggende princip var at hvert stammemorfem svarede til en del af det navngivne molekyle mens funktionsmorfemerne angav deres relationer. Ved at følge Ørsteds regler kunne man nu navngive både eksisterende og hypotetiske stoffer. Ørsteds revision af den kemiske nomenklatur var således en kæmpe gevinst for den teoretiske kemi som nu kunne formulere hypoteser om endnu ukendte stoffer uden at opfinde nye morfemer.<sup>4</sup>

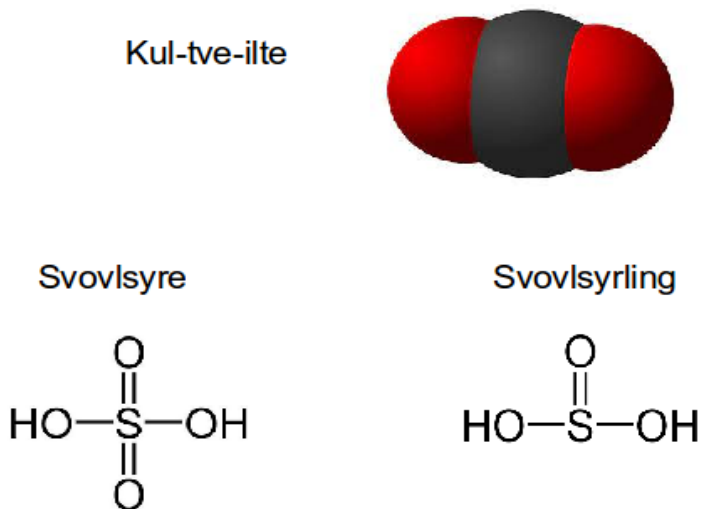
#### 4 KEMIENS ORD – ORDETS KEMI

Ørsteds principper for orddannelse i kemien går altså ud fra et minimalt antal stammer og udvider ordforrådet ved hjælp af komposita efter fastlagte regler. Der er mange spor af denne tankegang i resten af hans ordproduktion også; kemiens stringente orddannelse har tydeligvis tiltalt ham. Det virker derfor oplagt at vælge en morfologisk analysemodel til Ørsteds nyord der bygger på de samme formelle principper og begrænser mængden af grundlæggende analysekategorier til et minimum. En sådan formalisme har man i *kategorialgrammatikken*, som

---

<sup>4</sup> I dag er Ørsteds principper for kemisk nomenklatur (som også andre af hans samtidige og efterfølgere har bidraget til) blevet videreudviklet til den internationale standard IUPAC. Principperne for den moderne nomenklatur er påfaldende lig Ørsteds.

FIGUR 3. EKSEMPLER PÅ ØRSTEDS NOMENKLATUR



CO<sub>2</sub>-molekylets morfologi er direkte afspejlet i termen **KULTVEILTE**'s morfologi: Morfemet *tve* denoterer at kulatomet (afbildet i mørk grå) er divalent, dvs. binder to iltatomer (afbildet i rødt). I termen **SVOVLSYRLING** angiver diminutivsuffixet *-ling* det forhold at molekylet – i forhold til det umarkerede *svovlsyre* – har et iltatom (O) mindre.<sup>5</sup>

vi her indfører i dansk morfologi for første gang.<sup>6</sup> I det følgende gennemgår vi formalismens grundtræk ved hjælp af praktiske eksempler og henviser i øvrigt den interesserede læser til litteraturen (Wood 1993, Henrichsen 1998, Barker 2003).

Ordet *dumbhed* er, morfologisk betragtet, et substantiv sammensat af et adjektiv og suffikset [hed]. Det samme gælder Ørstedes nyord **FRANSKHED**.

A	A\N	==>	N
[dum]	[hed]		[dumbhed]

A	A\N	==>	N
[fransk]	[hed]		[FRANSKHED]

<sup>5</sup> Kilder til figuren kan findes her: <https://commons.wikimedia.org/wiki/File:Carbon-dioxide-3D-vdW.svg> (kultveilte), [https://commons.wikimedia.org/wiki/File:Acide\\_sulfurique\\_semi\\_dev.png](https://commons.wikimedia.org/wiki/File:Acide_sulfurique_semi_dev.png) (svovlsyre) og [https://commons.wikimedia.org/wiki/File:Schwefflige\\_S%C3%A4ure.svg](https://commons.wikimedia.org/wiki/File:Schwefflige_S%C3%A4ure.svg) (svovlsyrling).

<sup>6</sup> Kategorialgrammatik har kun været brugt ganske lidt i dansk sprogforskning i det hele taget; Tøgeby (1999), Henrichsen (2001), Wivel (2002) er vist en udtømmende publikationsliste.

De to ord er her vist i det såkaldte sekventformat (*sequent format*). Den del af en sekvent der står til venstre for sekventpilen, kaldes antecedenten (*antecedent*) mens delen til højre kaldes konsekventen (*consequent*). I eksemplet herover har stammen [dum] kategorien A (adjektiv) mens affikset [hed] er en sammensat funktion (A\N). Den baglæns skråstreg \ bruges til funktioner der søger et argument til venstre for sig selv, den almindelige skråstreg / når argumentet søges til højre. Funktioner af type  $x \setminus y$  og  $y/x$  siges at afbilde et  $x$  på et  $y$  (*to map  $x$  to  $y$* ).

V/V	V	==>	V	V/V	V	==>	V
[over]	[tale]		[overtale]	[over]	[ILTE]		[OVERILTE]

De traditionelle termer *præfiks* og *suffiks* er nu blevet til aliaser for henholdsvis  $y/x$  og  $x \setminus y$ , hvor  $x$  og  $y$  er kategorier hver for sig. Suffikset [hed]:A\N afbilder altså et adjektiv (A) på et substantiv (N), og præfikset [over]:V/V afbilder et verbum (V) på et verbum. Sekventanalyse går ud på at bevise at kategorien i konsekventen (fx N) kan dannes af kategorierne i antecedenten (A A\N) gennem lutter lovlige applikationer (*applications*), altså skridt hvor en funktion får det argument den beder om. Hvert morfem i et sammensat ord bidrager til ordets sekventering med én kategori. En stamme bidrager med en grundlæggende kategori (N, V eller A) mens alle andre morfemer er funktioner: præfikser, suffikser, infikser, cirkumfikser, fuger<sup>7</sup>, bøjningsmorfemer, deletegn, osv. Man kan udtrykke det sådan at hvert morfems kategori deklarerer dets bidrag til hele ordets betydning.

A/A	V/V	V	V\A	A\N	==>	N
[u]	[ud]	[tale]	[lig]	[hed]		[UUDTALELIGHED]

Prøv at finde et bevis for denne sekvent (der bare er én af flere mulige). Orddelen [udtale] skal komponeres først (V/V V ==> V), derefter [udtale][lig] (V V\A ==> A), så [u][udtalelig] (A/A A ==> A). Først nu

---

<sup>7</sup> Det er vel ikke almindeligt at klassificere fuger, deletegn m.v. som 'morfemer'; men i en kategorialgrammatisk sammenhæng er det praktisk at samle alle semantiske komponenter (betydningsbærende eller -adskillende) under én term da de bidrager på lige fod til sekventanalysen. En anden mulighed havde været at indføre en helt ny term (fx en 'morf').



kan affikset [hed]:A\N få sit argument og sekventen sit bevis. Denne derivationshistorie kan, efter behov, opfattes enten som en rapport over den morfologiske analyses stadier eller, mere ambitiøst, som en hypotese om ordets diakrone udvikling.

FIGUR 4. TRE SEKVENTERINGER AF ØRSTEDS NYORD

N	N\A/A	A	==>	A
[vægt]	[]	[fri]		[VÆGTFRI]

N	N\N/N	V	V\A	A\N	==>	N
[sjæl]	[e]	[virk]	[som]	[hed]		[SJÆLEVIRKSOMHED]

N	N\A	A\N	N\N/N	N	==>	N
[frugt]	[bar]	[hed]	[s]	[præg]		[FRUGTBARHEDSPRÆG]

Bemærk fugerne i de tre sekventer for VÆGTFRI, SJÆLEVIRKSOMHED og FRUGTBARHEDSPRÆG (figur 4). For anskuelighedens skyld har vi kategoriseret alle fuger på samme måde i denne fremstilling, såvel de materielle fuger [e] og [s] som nulfugen []. Kategorielt betragtet er en fuge, uanset sin ortografiske form<sup>8</sup>, en funktion af to argumenter og har derfor to skråstreger, en venstrevendt og en højrevendt. Fuger kan have enhver kategori  $x\backslash y$ , hvor både  $x$  og  $y$  er ét af elementerne i mængden  $\{N, V, A\}$ . Der er altså ni forskellige fugekategorier til rådighed for ord-dannelse, og alle er rigt repræsenteret i moderne dansk. Otte ud af de ni fuger forekommer også hos Ørsted, kun ikke  $V\backslash V/V$ . Den er i øvrigt også den sjældneste i nudansk, men ses fx i *løbetræne*, *tørretumble*, *sultestrejke*.

<sup>8</sup> Den såkaldte nulfuge er ganske vist usynlig i skriftbilledet (på grund af en ortografisk konvention), men er realiseret fonetisk med tryktab m.m.; den er altså ikke en *empty category* i Chomskysk forstand, dvs. en rent teoretisk enhed.

TABEL 1. FUGEKATEGORIER I ET UDVALG AF ØRSTEDS NYORD

Fugekategori	Belæg	Sekvent
N\N/N	ÅNDEFLUGT	[ånd]:N [e]:N\N/N [flugt]:N ==> [ÅNDEFLUGT]:N
V\N/N	FATTEFORM	[fatte]:V []:V\N/N [form]:N ==> [FATTEFORM]:N
A\N/N	FJERNSKRIFT	[fjern]:A []:A\N/N [skrift]:N ==> [FJERNSKRIFT]:N
N\V/V	LANDVÆNNE	[land]:N []:N\V/V [vænne]:V ==> [LANDVÆNNE]:V
V\V/V	-	-
A\V/V	PLANSLIBE	[plan]:A []:A\V/V [slibe]:V ==> [PLANSLIBE]:V
N\A/A	SAGARIG	[saga]:N []:N\A/A [rig]:A ==> [SAGARIG]:A
V\A/A	FØLGESTRENG	[følge]:V []:V\A/A [streng]:A ==> [FØLGESTRENG]:A
A\A/A	SALIGTRÆT	[salig]:A []:A\A/A [træt]:A ==> [SALIGTRÆT]:A

Med afsæt i analyserne i tabel 1 kan man fundere over om den traditionelle morfologiske term *fuge* er særlig heldigt valgt. Måske burde man erstatte den med *hængsel*. Funktionelt betragtet er et fugeelement jo ikke et passivt adskillende lag i et sammensat ord, men derimod et artikulerende led mellem samvirkende dele.

TABEL 2. MORFEMTYPER OG DERES KATEGORIALGRAMMATISKE PENDANTER

Orddel	Kategorial analyse	Eksempler leksemer og sammensætninger
stamme	grundlæggende kategori (svarende til en ordklasse)	[blå] : A - [spise] : V [ven] : N
suffiks	x\y (funktion fra kategori x til y)	[skab] : N\N [ven][skab] [lig] : V\A [spise][lig][hed] [hed] : A\N [yde][lse] [lse] : V\N
præfiks	y/x (funktion fra x til y)	[u] : A/A [u][klog] [be] : V/V [be][låne] [mod] : N/N [mod][yd][else]
fuge	x\y/y (funktion fra x og y til y)	[] : A\V/V [fn][pudse] [e] : N\N/N [barn][e][vogn] [s] : N\A/A [mand][s][høj]
affiks af højere orden	x\y\z (funktion fra x og y til z) (andre typer kan fore- komme)	[et] : N\A\A [sej][liv][et] [lang][bals][et] [hel][skind][et] <sup>9</sup>

Bemærk i tabel 2 at vi regner varianter som [lse] og [else] i kategori VN for allomorfe; tilsvarende [yd] og [yde] i kategori V, og desuden mange andre. Vi kommer nærmere ind på allomorfi i det næste afsnit.

## 5 DET DANSKE KOMPOSITUM OG DETS LÆSNINGER

Længere danske komposita vil i reglen have mere end én læsning. Et sammensat ord som *enebærbusk* kan betyde en busk med enebær eller en bærbusk af plantearten ene, afhængigt af hvilken orddel der komponeres først. Formelt betragtet har ordet netop to morfologiske analyser (læsninger) når det sekventeres sådan:

N	N\N/N	N	N\N/N	N	==>	N
[ene]	[]	[bær]	[]	[busk]		[enebærbusk]

Analysetræerne svarende til de to læsninger er vist i figur 5. Bemærk at det kategoriale regnskab går op i begge de to træer; relationen mellem datterknuder og moderknude i hver gren af et analysetræ er som antecedenten og konsekventen i en separat lille sekvent – et delbevis. I kategorialgrammatikken er bevisførelse alfa og omega. Dette at finde strukturen i en sammensat sproglig størrelse opfattes som at føre et logisk bevis fra en række præmisser (i antecedenten) til en lovlig konklusion (konsekventen), altså det som filosofien kalder et deduktivt ræsonnement. Kategorialgrammatikkens motto er ligefrem *parsing as deduction*.

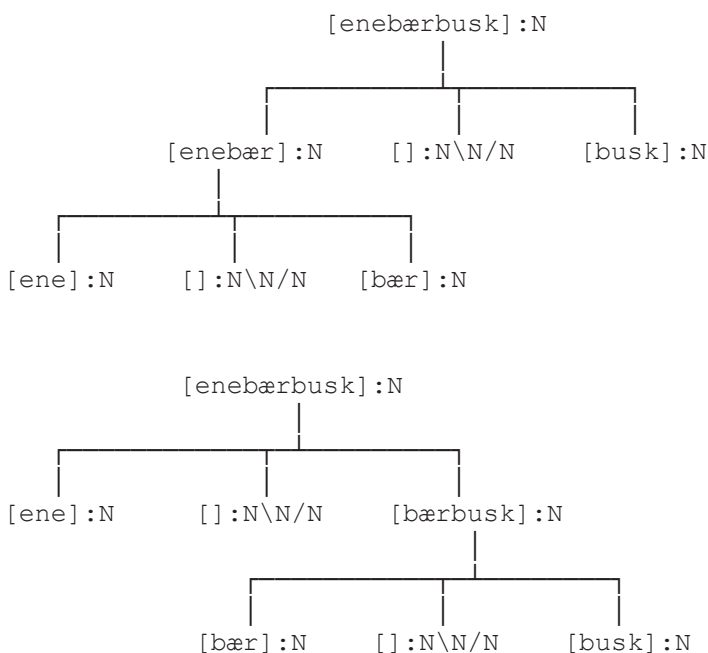
Imidlertid kendes [ene] jo ikke bare som et substantiv, men også som et nominalpræfiks i ord som *eneværelse* og *enebarn*, og det er ikke vanskeligt at forestille sig et begreb ved navn *enebær* der står i modsætning til fx *dobbeltbær*, altså en læsning med [ene] i kategori N/N. Til overflod kan *enebærbusk* også læses som en (ret sær) imperativ afledt af verbet *at buske*, der forekommer i både RO (Retskrivningsordbogen) og DDO (Den Danske Ordbog).

Den strukturelle flertydighed i ordet *enebærbusk* finder man tilsvarende i mange af Ørstedes nyord, fx VANDTRYKPRESSE, BÆREVÆGTSTANG, SILKEORMESPIND, GUDDOMSVIRKNING, ILDLUFTMÅLER, og faktisk har længere danske ord næsten altid mere end én morfologisk analyse. Skrift-

---

<sup>9</sup> Eksemplerne er fra Gram-Andersen (1992: 146-147). Bogen er helliget kompositumformen A+N+perf.particip. og har den selveksemplificerende titel 'The Purple-Eared Monster'.

FIGUR 5. ÉN ENEBÆRBUSK – TO ANALYSETRÆER



sproget har ingen systematiske midler til at pege på den intenderede læsning ud af mange mulige, og talesproget kan kun gøre det i beskedent omfang, især ved hjælp af distributionen af hovedtryk og bitryk. Trykfordelingen er dog et usikkert middel fordi det sammensatte ord, som hovedregel, kun bevarer hovedtrykket i sit *første* morfem, uafhængigt af morfemets plads i analysetræet, mens alle andre tryk svækkes til (forskellige grader af) bitryk. Denne hovedregel har undtagelser for komposita med præfikser (*selvfølgelig, u'tænkkelig, over'bærende*), uden at disse undtagelser dog er synderligt systematiske (*'selvstændig, 'uvirkelig, 'overstrømmende*), og under alle omstændigheder kan et binært træk (in casu ±hovedtryk) end ikke i princippet bruges til at udpege én specifik læsning blandt de talrige, i nogle tilfælde hundredevis, som længere komposita kan have. Kort sagt, afsenderen kan ikke meddele sin ønskede læsning af et sammensat ord med ortografiske eller fonetiske midler. I praksis er denne underspecifikation dog sjældent et problem eftersom modtageren afkoder selv lange og hypotaktiske komposita i en proces

der effektivt undertrykker langt de fleste analyser. Her er nøgleordet *mental tilgængelighed*. Jo større dele af et kompositum man kan genkende umiddelbart, des færre mulige læsninger er der tilbage og, følgelig, des mindre tvivl om ordets betydning. Vi bliver udsat for nye komposita hver dag, og hvis de er fornuftigt dannet, forstår vi dem uden besvær. Et nyord som *rengøringsparathed* (født i coronatiden) kan trække på en genkendelse af orddelene *rengøring* og *parathed*, og ingen dansk sprogbruger behøver at overveje mellemhypoteser som *rathed* eller *ringspar*.

De fleste danskere i forfatterens generation vil nok genkende ordet *enebærbusk* fra den gamle børnesang mens sprogbrugere som ikke kender sangen, er henvist til at analysere ordet ud fra dets dele. Vi finder alle seks komponenter, [ene], [bær], [busk], [enebær], [bærbusk] og [enebærbusk] i såvel RO som DDO, så man kan forvente at nogle danske sprogbrugere kender dem alle – og alle, nogle. Ordbøgerne siger dog intet om hvor præsent den enkelte sprogbruger kan formodes at have de seks leksemer og, følgelig, hvilken foretrukken læsning hun vil nå frem til. Hvis hun, som man ville forvente, opfatter [enebær] og [busk] som mere bekendte end [ene] og [bærbusk], vil hun formentlig foretrække [enebær][busk] frem for [ene][bærbusk].

### 5.1 Tolkningen af det ukendte ord

Et ukendt kompositum analyserer man, som nævnt, gennem sit forkendskab til ordets dele. Sammensætninger af almindeligt forekommende morfemer (fx i *plastikmus*) er lettere at tolke end sjældnere kombinationer (*tinokapi*). For at kunne vurdere hvor lettilgængeligt et sammensat ord er for morfologisk analyse, har vi altså brug for et kvantitativt mål for mental tilgængelighed, og ikke kun af ordet selv, men også af dets dele. Vi vil basere målet på ordfrekvenser i et korpus der er tilstrækkeligt stort til at selv de sjældneste ord den almindelige sprogbruger genkender, har forekomster (fx [ene]:N, [enebærbusk]:N). Det betyder at korpuset skal være i samme størrelsesorden som den mængde ord en typisk dansker møder i hele sit liv, eller med et groft overslag: en milliard ord.<sup>10</sup> Der er i skrivende stund kun ét alment tilgængeligt dansk korpus i denne størrelse, nemlig det nys offentliggjorte

---

10 1.000.000.000 ord svarer til 80 år med tre daglige timers (uafbrudt) sproglig eksponering.

DAGW eller Danish GigaWord Corpus (Derczynski et al. 2021), som netop omfatter godt 10 løbende ord. DAGW består af blandede tekst- og talegenrer som almindelige sprogbrugere regelmæssigt er i kontakt med, og det vil sige at (næsten) alle ordformer i DAGW forekommer i RO og/eller DDO, enten som leksemer (fx *ren*, *rengøre*, *rengøringsvanvid*) eller indirekte som regelmæssigt dannede komposita (*rengøringsparathed*). Som altid må vi dog tage forbehold for korpuslingvistens evige problem: at det frie talesprog, som er den sprogform danskerne er mest i kontakt med, er meget svært at indsamle og derfor underrepræsenteret i DAGW; men som nævnt findes der p.t. intet alternativ til DAGW i den ønskede størrelse (mere om DAGW i afsnit 7).

Vi benytter det samlede lemmainventar i RO og DDO som referencetil danskernes almindelige ordforråd og annoterer hvert lemma med dets frekvens i DAGW (i logaritmisk mål). Dermed har vi et indeks for hvor stærkt benyttede de danske lemmaer er, afledt af et bredt udvalg af danskeres sprogbrug. Bemærk at tilgængelighedsindekset, som vi vil kalde TiX, i første omgang kun er defineret for leksemer (stammer og komposita), men ikke for affikser og andre morfologiske funktionskategorier. Dette vender vi tilbage til.

TABEL 3. ENEBÆRBUSK – ANALYSERET FOR GENKENDELIGHED

Leksem	Frekvens i DAGW (ppm)	TiX-værdi (log)
{ene:N}	0,005	-19,1
{bær:N}	5,2	-12,2
{busk:N}	10,7	-11,4
{enebær:N}	0,4	-14,7
{bærbusk:N}	0,12	-16,0
{enebærbusk:N}	0,06	-16,6

I tabellen står  $\{Lem:Kat\}$  for lemmaet *Lem* i kategorien *Kat*. TiX-værdier beregnes som  $\log(n)$ , hvor  $n$  er den akkumulerede frekvens for lemmaets former. Enheden ppm står for parts-per-million. 1,0 ppm svarer til -13,8 i logaritmisk mål.

Et ord der er leksikaliseret mentalt – altså en form man genkender i sin helhed – kræver ingen morfologisk analyse, men kan hentes frem direkte fra langtidshukommelsen. I modsat fald må der flere opslag til og desuden et mentalt arbejde med at finde en foretrukken læsning

hvis der er flere mulige. Hvis vi antager at arbejdet med at analysere et sammensat ord vokser med antallet af nødvendige leksikonopslag, kan vi sortere de forskellige læsninger af *enebærbusk* efter tilgængelighed.

$\text{TiX}(\text{[enebærbusk]:N})$	=		=	-16,6
$\text{TiX}(\text{[enebær][busk]:N})$	=	$\text{TiX}(\text{[enebær]:N}) + \text{TiX}(\text{[busk]:N})$	=	$-14,7 - 11,4 = -26,1$
$\text{TiX}(\text{[ene][bærbusk]:N})$	=	$\text{TiX}(\text{[ene]:N}) + \text{TiX}(\text{[bærbusk]:N})$	=	$-19,1 - 16,0 = -35,1$
$\text{TiX}(\text{[ene][bær][busk]:N})$	=	$\text{TiX}(\text{[ene]:N}) + \text{TiX}(\text{[bær]:N}) + \text{TiX}(\text{[busk]:N})$	=	$-42,6$

Den lettest tilgængelige analyse er som ventet [enebærbusk], altså genkendelsen af ordet in toto, men den forudsætter selvsagt at man kender det i forvejen. I modsat fald foretrækkes [enebær][busk] over [ene][bærbusk], igen som ventet. Ørsteds GUDDOMSVIRKNING får, ved en tilsvarende analyse, den foretrukne nudanske læsning [guddom][s] [virkning] mens den anden læsning [gud][domsvirkning] fravælges eftersom *guddom* er omtrent 100 gange hyppigere end *domsvirkning* i DAGW.

## 5.2 De danske funktionsmorfer

Som nævnt skal tilgængelighedsindekset også omfatte funktionsmorfer. Mens leksemer kan kopieres direkte fra ordbøger og korpus, må affikser afledes af leksemerne. Der er ingen skarp grænse mellem leksemer og affikser, og mange morfer kan have begge kategoriseringer. Vi har allerede omtalt [ene]:N/N, andre eksempler er [selv]:A/A og [dygtig]:V\A som i *selvdød*, *selvbiografisk* og *flyvedygtig*, *spiredygtig*. For at et morfem skal kunne kategoriseres som affiks, må det forekomme i et større antal leksemer med den samme placering og betydning. I ord som *dumbhed*, *åndrighed*, *krænkelssparathed*, ÅNDIGHED, MODBYRDIGHED, GENSTÆNDIGHED kan suffikset [hed] i alle tilfælde parafraseres som “den egenskab at være *X*” hvor *X* er styrelsen.

A     A\N    ==>   N  
[dum]   **[hed]**           [dumhed]

V     V\N   N\A/A   A     A\N    ==>   N  
[krænk] [else] [s]     [parat] **[hed]**     [krænkelsesparathed]

A/A    N     N\A    A\N    ==>   N  
[mod]   [byrd] [ig]     **[hed]**     [MODBYRDIGHED]

Det er ikke svært at aflede suffikset [hed] af ordpar som [dumhed]-[dum], [skønhed]-[skøn], [lethed]-[let], [kådhed]-[kåd], og ordbogen har i alt 1020 lemmapar af denne type. Algorimen Fixi (figur 6) er udviklet til automatisk at udtrække affikser af en ordbog efter dette mønster.

FIGUR 6. ALGORITMEN FIXI TIL ELICITERING AF AFFIKSER (PRINCIPSKITSE)

---

Gentag for alle ord  $W$  (leksikalske indgange) i ordbogen:

Del  $W$  op i to dele, og kald dem  $W_1$  og  $W_2$

Hvis  $W_1$  findes i ordbogen, så

registrér  $W_2$  som et affiks  $y/x$ ,

hvor  $x$  er  $W$ 's kategori og  $y$  er  $W_1$ 's kategori

Hvis  $W_2$  findes i ordbogen, så

registrér  $W_1$  som et affiks  $x/y$ ,

hvor  $x$  er  $W$ 's kategori og  $y$  er  $W_2$ 's kategori

Fortsæt til alle opdelinger af  $W$  er prøvet

Fortsæt til alle indgange i ordbogen er prøvet

Udskriv alle registrerede affikser med kategori og antal belæg

---

Fixi finder alle de almindelige, produktive affikser i referenceordbogen (RO+DDO) og udskriver dem med frekvensdata afledt af DAGW. Tag som eksempel affikset *mis-*, som Fixi afleverer i tre morfologiske varianter: som verbalpræfiks [mis]:V/V (fx *mistanke*, *misforstå*), substantivpræfiks [mis]:N/N (*mistanke*, *mismod*) og adjektivpræfiks [mis]:A/A (*mismodig*, *misfarvet*). Den sidstnævnte variant er langt den mindst benyttede; A/A har kun 9 belæg (med ≈11k forekomster i DAGW), hvor



N/N har 38 belæg ( $\approx 113k$ ) og V/V har 52 ( $\approx 122k$ ). Det er altså ikke så overraskende at Ørstedes MISSTOR føles mere fremmed end fx MISSTEMNING, i betragtning af at adjektivvarianten er langt mindre produktiv end de andre. Størstedelen af Ørstedes nyord er konstrueret med affikser der stadig er produktive i nudansk, og som derfor forekommer i Fixis output; undtagelserne er de eksotiske affikser af Ørstedes egen opfindelse, fx -LUNDIG, -ÆSKIG, der selvfølgelig mangler helt.

Da Fixi-algoritmen er rent mekanisk, vil den, sammen med de anerkendte danske affikser, også finde store mængder af pseudoaffikser.

[skr]:V/N	fra belæg som [abe]:N og [skrabe]:V
[gu]:A/N	fra belæg som [sten]:N og [gusten]:A
[re]:V\W	fra belæg som [serve]:V og [servere]:V

Fixi registrerer for eksempel [skr] som et muligt præfiks i kategori V/N fordi såvel verbet 'skrabe' som substantivet 'abe' findes i ordbogen, hvad der tillader sekventen [skr]:V/N [abe]:N  $\Rightarrow$  [skrabe]:V. Pseudoaffikser som [skr] er naturligvis uønskede i output. I praksis kan man sortere dem fra ved at udnytte de frekvensdata som ordbogen har arvet fra DAGW; de ægte affikser vil nemlig, alt andet lige, forekomme langt hyppigere end de falske. Det er netop den egenskab der gør dem lærbare (ikke kun for automater, men givetvis også for sproglærende børn). Ved at sætte passende tærskelværdier i Fixi kan man opnå et output der næsten kun består af gangbare danske affikser. I tabel 4 nedenfor ses de mest frekvente affikser fundet af Fixi (se også note 11).

Tabellen illustrerer nogle generelle træk ved det danske sprog, som også 'klassiske' morfologer har bemærket. For det første er suffikser mere udbredte end præfikser, både i antal og frekvens. For det andet er præfikser næsten altid af typen  $x/x$  (kategoribevarende) mens suffikser oftest er polymorfe (kategoriskiftende). Disse tendenser møder man overalt i dansk tekst, tale og ordbøger (også Ørstedes), så de kan regnes for inhærente træk ved sproget.

Ortografiske allomorfer, såsom [lse]:V\N og [else]:V\N, læres af Fixi som uafhængige morfemer, hvad der giver ord som *nydelse* og *SOLDYRKELSE* flere alternative analyser der nok er overfladisk forskellige, men uden at være reelt flertydige.

[nyde]:V [lse]:V\N ==> [nydelse]:N  
 [nyd]:V [else]:V\N ==> [nydelse]:N

[sol]:N []:N\N\N [dyrke]:V [lse]:V\N ==> [SOLDYRKELSE]:N  
 [sol]:N []:N\N\N [dyrk]:V [else]:V\N ==> [SOLDYRKELSE]:N

Fænomenet kaldes spuriøs flertydighed (*spurious ambiguity*) – efter det latinske *spurius* der betyder 'uægte barn'. Der er mange andre eksempler på denne type af redundans, hvor stamme-affiks-kombinationer med identisk betydning og udtale alligevel registreres individuelt. En stor del af problemet skyldes de danske ortografiske vedtægter såsom reglen om dobbeltskrevne konsonanter som markør af en kort vokal

TABEL 4. DE HYPPIGSTE DANSKE AFFIKSER

Affiks	Kategori	Belæg (RO+DDO)	Eksempler (RO+DDO)	Eksempler (HCØ)
[ing]	V\N	2086	[køl ing] [lær ing]	[SITR ING]
[ning]	V\N	1509	[spis ning] [svøm ning]	[FORDAMP NING]
[hed]	A\N	1020	[dum hed] [besindig hed]	[FORMBAR HED]
[else]	V\N	889	[tilstå else] [vi else]	[TE ELSE]
[lse]	V\N	865	[skabe lse] [smøre lse]	[SELVNYDE LSE]
[er]	V\N	847	[lær er] [slagt er]	[NEJSIG ER]
[isk]	N\A	652	[dyr isk] [biolog isk]	[ELEKTROMAGNET ISK]
[af]	V/V	574	[af gøre] [af se]	[AF BRINTE]
[et]	V\A	536	[skuff et] [forband et]	[TANKEFATT ET]
[ud]	V/V	499	[ud dele] [ud gå]	[FOR U BETINGE]
[for]	V/V	457	[for dele] [for stå]	[FOR FLYGTE]
[op]	V/V	397	[op give] [genop stå]	[OP KOMLING]
[ind]	V/V	349	[ind tage] [ind se]	[IND DYPPE]
[be]	V/V	331	[be handle] [be gå]	[FORUD BE TINGE]
[for]	N/N	322	[for mand] [for syn]	[FOR SVOVL]
[hoved]	N/N	285	[hoved sag] [hoved vej]	[HOVED DRIVFJEDER]
[over]	V/V	276	[over se] [over bevise]	[OVER FORBEDRE]

De hyppigste affikser afledt med algoritmen Fixi; kolonne 3 er antallet af belæg i ordbogen (lemmapar); kolonne 4 og 5 giver eksempler på belæg fra henholdsvis de nudanske og Ørstedeske ordkilder.

(Becker-Christensen 2021 6.2). De spuriøse dobbeltformer er uskønne, og de giver Fixi et betydeligt ekstraarbejde; men i øvrigt er de harmløse da de ikke fører til nye læsninger.

Bemærk at Fixi har registreret [hoved] som et præfiks N/N, ud fra belæg som *hovedvej* og *hovedsag*. Selv om morfemet *hoved* naturligvis også er et lemma i kategori N, virker det ikke urimeligt at regne det som affiks i de nævnte belæg hvor det ikke denoterer en legemsdel, men derimod virker som en semantisk funktion der tilføjer en særlig høj status til *vej* og *sag* (med hovedet som metafor). Lignende eksempler har man i [dømme] som i *kongedømme* og FOLKEDØMME, [lig] som i *opløselig* og TUNGLØSELIG, samt [mod] som i *modgang*, MODGÆNGER. Disse morfemer forekommer også som stammer med tilsvarende betydning, men virker i belæggene som semantiske funktioner. De fleste affikser i det levende sprog – hvis ikke alle – har en oprindelse som selvstændige leksemer, og det sikre tegn til at et leksem er i færd med at overgå til en synkategorematisk funktion, er netop at det efterhånden på de fleste af sine forekomster virker ved at modificere et bredt udvalg af leksemer i de grundlæggende kategorier. Det er netop tilfældet med *hoved* og årsagen til Fixi's kategorisering.

Forventningen er altså at ægte affikser (fx [hoved]:N/N, [ning]:V\N) kan kendes på at de har mange belæg, i modsætning til pseudoaffikser ([skr]:V/N, [re]:V\V). Derfor baserer vi TiX-scoren på belæggenes hyppighed.<sup>11</sup>

### 5.3 Kendte ord og nye ord

Hermed er modellen af den nudanske sprogtilstand klar til praktisk brug; hvert nudansk leksem og hvert morfem der indgår i det, kan nu vurderes for tilgængelighed med TiX-scoring. Analysetræernes grene (svarende til orddele) kan valueres hver for sig, og de orddele som nudanskeren genkender umiddelbart (fx RUMFANG og FLADEMÅL), vil have en højere TiX-score end de der må deles op (fx RUMHOLD i [rum][hold], FLADEMIL i [flade][mil]) eller opgives helt (FLUGLIT, ÆSKIG).

---

<sup>11</sup> Af praktiske grunde udelades her de tekniske specifikationer for Fixi (algoritme, outputfiltrering), TiX-definition, TiX-moduleret ordsegmentering, sekventanalyse, lambdakalkyle m.m. Detaljer fremlægges i en kommende artikel.

Hele proceduren har fire trin, som illustreret i figur 7 med analysen af ILDLUFTMÅLER.

FIGUR 7. TILGÆNGELIGHEDSANALYSENS FIRE TRIN

Segmentering	[ild][luft][mål][er]
Sekventering	[ild]:N <sub>(1)</sub> []:(N\N)/N <sub>(2)</sub> [luft]:N <sub>(3)</sub> []:(N\N)/N <sub>(4)</sub> [mål]:V <sub>(5)</sub> [er]:V\N <sub>(6)</sub> ==> N
Bevisførelse	1. læsning: 4·(6·5)·(2·3·1) 2. læsning: 2·(4·(6·5)·3)·1
Tilgængelighed	1. læsning: TiX( [ildluft] ) + TiX( [måler] ) = -19,3-11,4 = <b>-30,7</b> 2. læsning: TiX( [ild] ) + TiX( [luftmåler] ) = -10,3-19,3 = <b>-29,6</b>

Sekventen i figur 7 er forsynet med indeksnumre, samt parentesstruktur der viser rækkefølgen af applikationerne. Som parenteserne viser, beder fugeelementerne om at få deres højreargument før venstreargumentet. Det fremgår at sekventen har to beviser (og ILDLUFTMÅLER følgelig to læsninger), men i stedet for analysetræer er de vist i figuren som såkaldte lambdaformler. Tallene i formlerne henviser til sekventens indeksnumre, og symbolet · står for et applikationstrin; fx står 6·5 for applikationen [mål]:V [er]:V\N. Tal i samme parentes (fx 2·3·1) svarer til søsterknuder i analysetræet.

Læsningen [ildluft][måler] har en TiX-værdi på -30,7, mod -29,6 for læsningen [ild][luftmåler]. Den foretrukne læsning er altså den sidstnævnte. Der er dog ikke stor forskel på de to målinger, og forudsigelsen er at nudanskeren vil være i tvivl om den intenderede læsning. Forfatteren indrømmer villigt at have været usikker og opsøgte derfor kildeteksten til ordet ILDLUFTMÅLER (Ørsted 1826). Det viser sig at være et måleinstrument som i samtiden kaldtes et *oxygenometer*, mens ILDLUFT er et tidligt synonym til ILT, Ørsteds senere så populære kemiske betegnelse.<sup>12</sup> Skal det anses for en fejl ved TiX-indekset at det udpegede den forkerte læsning af ILDLUFTMÅLER som den foretrukne? Nej, tværtimod. Formålet med tilgængelighedsanalysen er at forudse hvad nudanskeren, ikke en alvidende modtager, vil opleve ved mødet med Ørsteds nyord.

<sup>12</sup> Ørsted brugte første gang termene ILT, BRINT og KVÆLSTOF i sit hovedværk *Kraftlæren*, som fortalt i Christensen (2009, kap. 28; se især I:458).

## 6 ØRSTEDS NYORD OG DERES PLADS I NUDANSK

Vi har udsat alle Ørsteds nyord for den analytiske procedure der er udviklet herover. For hvert af Ørsteds nyord har vi identificeret den foretrukne læsning, altså den blandt alle mulige morfologiske analyser som maksimerer tilgængelighedsindekset. Denne læsnings TiX-indeks tager vi som et mål for hvor intakt ordet er i dansk, eller rettere, hvor tilgængeligt (eller endda brugbart) det vil forekomme for den nudanske sprogbruger. Nyord der findes identiske i RO+DDO, vil, alt andet lige, score højest. Ordformer som kan opløses i to hyppigt brugte orddele, scorer også relativt højt (LÆGEKRAFT, KIRKESTRID), mens TiX-værdierne hurtigt falder for ord som består af sjældnere dele (LEDETRISSE, TVÆRGÆNGE) eller som må deles i tre eller flere segmenter (MEDVANDRINGSSTOF, LEGEMLIGHEDSFORM) for at finde en nudansk læsning. Nederst i listen har vi de nyord som slet ikke kan finde en læsning uden et ukendt morfem, og som altså får prædikateret: Utilgængeligt.

TABEL 5. ØRSTEDS KURANTE NYORD

TiX	Nyord
>-12	mindretal; væremåde; billedkunst; udstråling; selvforståelse
-13 - -12	ilt; samfundsudvikling; modtræk; lyspunkt; udtryksform; luftfart; ildsjæl; hovedkraft; mellemleder; <u>fordampe</u> ; kvælstof; brint; <u>spændstig</u> ; <u>indordne</u> ; samfundsliv; faldskærm
-14 - -13	foreteelse; vekselvirkning; følgevirkning; klangbund; livserfaring; vandmasse; verdensbillede; berøringsflade; <u>bramfri</u> ; <u>højstemt</u> ; overskuelighed; samfundsforhold; tankevirkksomhed; <u>fortætte</u> ; grundtanke; <u>gennemprøve</u> ; opkomling; tidevand; autoritetstro; <u>indleve</u> ; <u>indholdsrig</u> ; grundform
-14,5 - -14	brugskunst; frostgrad; farvestof; lysbillede; spændstighed; sammendrag; <u>umagnetisk</u> ; <u>aftænke</u> ; lyskilde; luftmodstand; sanseindtryk; skæringspunkt; tankeeksperiment; verdensanskuelse; <u>meddelsom</u> ; grundstemning; ueftertænksomhed; lysstyrke; billeddannelse; billedflade; modstandskraft; magtfuldkommenhed; hverdagsprog; fordampning; lynafleder; lysforhold; opdrift; jordår; synsvidde
-15 - -14,5	faldtid; kulilte; bærekraft; væggtab; brinte; hovedakse; tankerække; vægtfylde; <u>elektromagnetisk</u> ; kulbrinte; forflytning; dannelseshistorie; luftskib; handlingsgang; virkningsgrad; elektromagnetisme; selvished; menneskeforhold; tidevandsbølge; nejsiger; flerhed; <u>glansløs</u> ; <u>snæversynet</u> ; luftforandring; arbejdsmængde; lydgang; farvesans; sammenføring; <u>foretagsom</u> ; tilstandsform

Verber er underregtet; adjektiver er underprikket; substantiver er umarkerede. Bemærk at næsten alle sammensatte ord i tabellen er genkendelige uden morfologisk analyse, dvs. reelt nudanske (se dog note 14).

I tabel 5 ses de leksemer som har højest TiX-værdi og altså er lettest tilgængelige for morfologisk analyse; størstedelen af dem forekommer stadig i den typiske danskers ordforråd – i alt fald det passive. I den modsatte ende af spektret (TiX under -40) finder man de former der ingen spor har efterladt i nudansk (idet vi ser bort fra de kemiske termer, se afsnit 3), såsom JÆVNSTRØGSGÅENDE; ANDAGTOPVÆKENDE; FORFOLDIGHEDSANGIVER; MILDTÅNDE; ÆSKIGTSMAGENDE; JORDOMBÆLTENDE; FORBIILENDE; LETFORDUNSTENDE; MELLEMLVÆSKE; VIDTOMVANDRENDE; ORIGINALITETSPRÆTENDENT; FORFULDSTÆNDIGE; BEDUGNINGSPUNKT; VANDSTRØGSPASSER; ENSBELIGGENDE; BEDSTLEDENDE; JÆVNTAFTAGENDE; KNALDLUFTSBLÆSERØR; LAKMUSVINÅND; SELVFORFOLDIGHEDSANGIVER.

Det sidste element i denne liste er samtidig det længste og morfologisk set mest komplekse af alle Ørsteds (overleverede) nyord. Ville nogen bruge ordet SELVFORFOLDIGHEDSANGIVER i dag? Næppe. Er det uforståeligt for en nudansker? Ikke nødvendigvis. Det vender vi tilbage til herunder.

Tabel 6 giver et overblik over Ørsteds nyordsproduktion og dens fordeling over TiX-intervaller. Tolket som lingvistiske observationer viser tabellen at omkring 5 % af Ørsteds nyord endnu i dag indgår i den almindelige danskers aktive ordforråd mens de næste 5 % nok er bevaret i de nudanske ordbøger (RO+DDO), men sjældent forekommende i almindelig diskurs og måske ikke kendt eller brugt af alle. De følgende 25 % forekommer ikke i ordbøgerne, men hovedparten har alligevel forekomster i DAGW og kan dermed gøre krav på at kaldes nudanske, dog marginalt, og uden at være anset for kerneordstof af ordbogsredaktionerne. De næste 25 % (med TiX-værdier i omegnen af -21) er på en måde Ørsted-ordbogens mest interessante afsnit, for her findes det forråd af glemte ord (dvs. uden forekomst i DAGW og RO+DDO) som kan recirkuleres; de forstås uden problemer, og mange af dem har et frisk udtrykspotentialt – som demonstreret i afsnit 2 herover hvis eksempler (i parenteser) er citeret fra netop dette TiX-segment. Andre anvendelser af dette “de glemte ords interval” kunne være: som materiale til sprogpil, ny lyrik, undervisningsmaterialer og – ikke mindst – til afprøvning af sprogteknologiske programmer, et tema vi vender tilbage til. Hvad angår de sidste og mindst tilgængelige 40 % af Ørsteds

TABEL 6. ØRSTEDS NYORD OG DERES TILGÆNGELIGHED I NUDANSK (KEMISKE TERMER ER UNDTAGET)

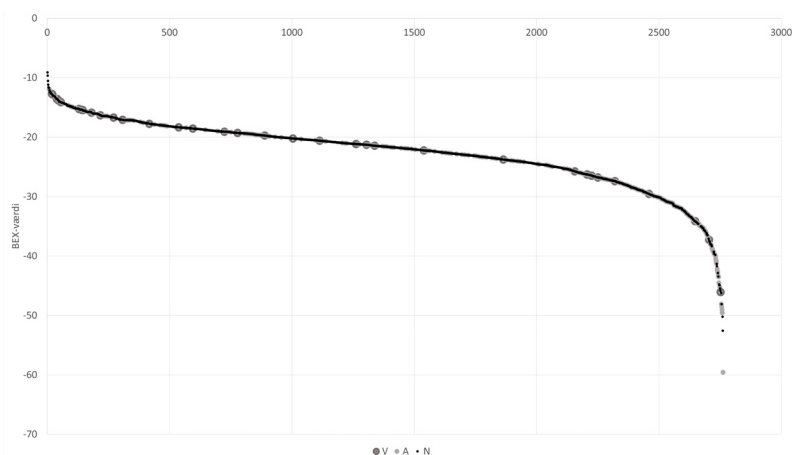
TiX	Status i nudansk (tendens)	Eksempler	Ordklasser	Andel (hele 5 %)
> -15	almindelige i nudansk, jf. tabel 5	<i>se tabel 5</i>	især N og V	5 %
-17 - -15	træffes i nudansk diskurs, evt. som enkle sammensætninger	VINDMÅLER REJSEVEN	især N og V	5 %
-20 - -17	træffes sporadisk i nudansk diskurs, men forstås umiddelbart	TALESTIL NABORUM	N, A og V	25 %
-23 - -20	træffes ikke i nudansk diskurs, men kan recirkuleres	HALESTJERNE ORDKRÆSEN	N, A og V	25 %
-26 - -23	forstås uden dybere refleksion; næppe brugbare i nudansk diskurs	PØBELTRIUMF BIFALDSGIVNING	især N og A	15 %
-30 - -26	forståelige efter nogen refleksion; ikke brugbare i nudansk diskurs	GRANSKERFRYD LUFTGLØDNING	især N og A	10 %
< -30	tungt forståelige; har formentlig kun sprogvidenskabelig interesse	KRYDSBYRDIG LIGEBESKAFFEN	især A	15 %

Tabellen er en opgørelse over lemmabestanden i databasen HCOe\_nyord.01. Lemmaerne er opdelt i 7 segmenter efter TiX-værdi. Hvert segment er vurderet for status i nudansk ud fra manuel evaluering af de enkelte lemmaer. Typiske eksempler ses i tredje kolonne.

nyord kan vi konstatere at de ingen målbar indflydelse har haft på den nudanske sprogtilstand. Ved stikprøvekontrol viser det sig at denne del af ordforrådet end ikke har indgange i Ordbog over det Danske Sprog (i modsætning til de øvrige segmenter). De lemmaer der således aldrig kom i cirkulation, er typisk lange, sammensatte adjektiver, mens Ørstedes substantiver og verber lettere slog an.

I grafen herunder (figur 8) er hele ordsamlingen afbildet med TiX-værdier på y-aksen, altså det samme datasæt som er summeret i tabel 6.

FIGUR 8. ØRSTEDS NYORD OG DERES TILGÆNGELIGHED I NUDANSK (KEMISKE TERMER ER UNDTAGET)



Hvert nyord er ét datapunkt, med TiX-værdi på y-aksen. Substantiver (N) er små sorte punkter, adjektiver (A) er mellemstore grå, verber (V) store mørkegrå. Sammenlign med tabel 6: Grafens stejle venstre del ( $TiX > -17$ ) svarer til de 5-10 % som stadig er almindelige i nudansk diskurs (relativt mange V); den brede, omtrent vandrette del ( $-30 < TiX < -17$ ) er præget af lemmaer der forekommer sporadisk eller i alt fald forstås umiddelbart (alle ordklasser); den stejle del til højre ( $TiX < -30$ ) domineres af de tungt forståelige ord (relativt mange A) som aldrig kom i cirkulation (og ikke forekommer i ODS).

## 7 DISKUSSION

Tilgængelighedsindekset TiX har vist sig effektivt til sortering af Ørstedes nyord efter deres grad af mental tilgængelighed i en nudansk kontekst. TiX-målet giver konkrete forudsigelser af om nudanskeren vil opleve et givent kompositum som velkendt, uforståeligt, eller noget derimellem. De enkelte målinger er reproducerbare, og måleprocedurens overordnede relevans er åben for evaluering ved kontrolforsøg (fx med danskere uden sprogfaglig baggrund), hvad der taler til dens fordel.<sup>13</sup> Som model af en sprogtilstand giver TiX-indeksets korpusbaserede tilgang en ret eksplicit definition af 'sprogbruger', en bestemmelse der jo er sammenfaldende med korпустeksternes intenderede modtagergruppe vægtet over de repræsenterede genrer. Ved indsamlingen af korpus DAGW blev der, som afgrænsningskriterium, kun optaget genrer som (tilnærmelsesvis) alle voksne danskere er eksponeret for og forventes at

<sup>13</sup> Desuden er TiX-målingen jo ikke begrænset til Ørstedes nyord, men fungerer for vilkårlige neologismer.



kunne forstå, fx skønlitteratur, dagblad, tidsskrifter for ikke-professionelle læsere, henvendelser fra offentlighed til borgere, chat, spontantale etc. Den præcise specifikation for DAGW, og dermed for TiX-målets immanente sprogbruger, er fremlagt i Derczynski et al. (2021); uformelt betragtet har vores undersøgelse altså sigtet mod voksne danskere i ethvert uddannelsestrin, uden atypiske sprogvanskeligheder.

På mange måder er TiX-målet altså klart defineret og praktisk anvendeligt. På minussiden må det anføres at TiX-indekset jo kun karakteriserer et lille, afgrænset hjørne af den samlede sprogtilstand. Dels siger det ingenting om dansk sprogbrug ud over ordets lille domæne, intet om syntaksforhold, faste vendinger, kontekst; og dels bygger indekset alene på frekvensdata, der jo somme tider er upålidelige som mål for mental tilgængelighed. Danske ord der har karakter af kulturrelikter, såsom *tinsoldat*, *hundehus*, *grisling*, *dukkehjem*, har for eksempel ekstremt små forekomsttal i DAGW, skønt alle kender dem. Og sidst men ikke mindst har vi slet ikke inddraget leksemernes betydning. Hvert eneste af Ørsteds nyord, selv de bedst bevarede, må nødvendigvis have udviklet sig semantisk gennem to århundreder. Ørsteds *LYSBØLGE* er så levende et ord som nogensinde, man fristes til at sige lyslevende; men betyder ordet stadig det samme i dag som i 1835 da Ørsted brugte det i sine Erindringsord til Forelæsninger over Lyset? Siden Ørsted har fysikere som Planck, Einstein og Bohr påvist at lys nok er bølger, men samtidig udbreder sig som små pakker af energi, at de rejser hurtigst af alle partikler i universet, og endda at de kan opstå ud af intet og forsvinde igen. Alle disse opdagelser, og mere til, er efter Ørsteds tid blevet føjet til *LYSBØLGE*'s semantik som konnotationer, og det samme vil gælde *LYSUDBREDELSE*, *LYSGIVER*, *LYSLEDENDE*, *LYSKILDE*, for bare at nævne nogle få. Selv de bedst bevarede lemmaer vil have ændret betydning gennem 6-8 generationers sprogbrug.<sup>14</sup> Vores undersøgelse er blind for denne type af udvikling, men til gengæld har den fundet konkrete svar på det simple spørgsmål om hvor tilgængelige (for morfologisk analyse) Ørsteds nyord er for nutidens danskere.

---

<sup>14</sup> Der er også eksempler på ord som vi genkender umiddelbart, men i en betydning uden relation til Ørsteds, såsom *LUFTFART* (betyder 'lufthastighed' hos Ørsted), *MELLEMLEDER* ('batterielektrolyt') og *ALRUM* ('univers'). I andre tilfælde er Ørsteds grundbetydning blevet afløst af en overvejende metaforisk brug, fx *LUFTFORANDRING*, *SNÆVERSNET*, *LETBEVÆGELIG*.

Man kan se vores undersøgelse af Ørstedes nyordssamling som et eksempel på datalingvistisk metode. Stående over for et komplekst problemfelt (“Hvilken indflydelse har Ørstedes nyord haft på det danske sprog?”) udsondrer man et lille, men ikke-trivielt delområde som lader sig formalisere. Det grundlæggende spørgsmål forenkler man tilsvarende (“Hvilke materielle ligheder ser vi mellem Ørstedes nyord og ordbestanden i nutidens ordbøger og korpora?”). På den måde vinder man indsigter som hjælper videre fra en første overfladisk undersøgelse til den næste, lidt dybere. I denne artikel har vi nøjedes med at bestemme et ’ord’ som en morfologisk form uden betydningsindhold. Det har givet et katalog af detaljerede ordanalyser med grammatisk kategorisering af alle morfemer og deres indbyrdes relationer. På den baggrund kan det næste skridt tages – for eksempel ved at udvikle definitionen af ’ord’ videre i retning mod det semiotiske tegn.

Dét er et fristende skridt at tage netop nu. I april 2021 startede projektet Det Centrale Ordregister (COR) der er et samarbejde mellem Dansk Sprognævn, Det Danske Sprog- og Litteraturselskab og Center for Sprogteknologi. Projektet skal blandt andet udvikle en database med semantisk annotation for de hyppigste danske ord; og ordsemantik er netop hvad Ørsted-projektet behøver for at kunne tage fat på undersøgelsen af betydningsbevarelse og betydningskift hen over århundrederne. En mulig strategi kunne være at tage udgangspunkt i de analysetræer vi har afleveret af Ørstedes lemmer. Hver gren i hvert træ bærer en lambdaformel, og en sådan formel er en indgang for en betydningsbeskrivelse. Sæt semantik til formlerne, og kalkylen komponerer ordets samlede betydning som en funktion af hvert enkelt morfems betydning. Eller sagt på en anden måde: Med semantiske data fra COR-projektet (eller tilsvarende) bliver det muligt at udlede betydningen af et ukendt ord fra betydningen af dets genkendte dele. Princippet lader sig nok bedst illustrere med et eksempel. Derfor slutter vi denne præsentation med en lambda-semantisk analyse af Ørstedes imponante nyord SELVFORFOLDIGHEDSANGIVER.

Ørsted forklarer selv ordet som en matematisk term synonym med *eksponent*, det vil sige det numeriske indeks i et potensudtryk der angiver hvor mange gange et tal skal ganges med sig selv. I ligningen  $2^3 = 8$  er 3-tallet altså selvforfoldighedsangiveren.

*Ørstedes betydningsforklaringer*

[FORFOLDIG]:A	<multipliceret, ganget>
[FORFOLDIGHED]:N	<produkt (resultatet af et gangestykke)>
[SELVFORFOLDIGHED]:N	<potens (resultatet af at gange et tal med sig selv)>
[SELVFORFOLDIGHEDSANGIVER]:N	<eksponent>

*Betydningsforklaring til de øvrige morfemer*

[selv]:N/N	<X-med-sig-selv>
[for]:A/A	<gjort-X>
[fold]:N	<antal>
[ig]:N\A	<X-mæssig>
[hed]:A\N	<det-at-være-X>
[s]:(N <sub>2</sub> \N)/N <sub>1</sub>	<X <sub>1</sub> -angående-X <sub>2</sub> >
[an]:V/V	<X-specifikt>
[giv]:V	<overføre-et-eller-andet (ting eller information)>
[er]:V\N	<væsen-der-X>

*Den semantiske analyses trin (svarende til analysetræets grene)*

[fold]:N	<antal>
[foldig]:A	<antalmæssig>
[FORFOLDIG]:A	<gjort-antalmæssig, iflg. Ørsted synonymt med 'ganget'>
[forfoldighed]:N	<det-at-være-ganget>
[selvforfoldighed]:N	<det-at-være-ganget-med-sig-selv>
[giv]:V	<overføre-information>
[giver]:N	<væsen-der-overfører-information>
[angiver]:N	<væsen-der-overfører-specifik-information>
[SELVFORFOLDIGHEDSANGIVER]:N	<væsen-der-overfører-specifik-information angående det-at-være-ganget-med-sig-selv>

Det semantiske metasprog i disse definitioner er måske ikke særlig naturligt, men med lidt god vilje kan man godt erkende resultatet af analysen som en hæderlig forklaring på termen *eksponent*.

N/N A/A N N\A A\N (N\N)/N N/N V \N ==> N  
[selv] [for] [fold] [ig] [hed] [s] [an] [giv] [er]

Hvis kategorialgrammatikken har vakt læserens nysgerrighed, er bevisførelsen for sekventen herover en god lille øvelse i *parsing as deduction*. Det korrekte analysetræ, med lambdaformler på bladene, kan ses på Sprognettets hjemmeside, [www.dsn.dk/Oersted](http://www.dsn.dk/Oersted)

## 8 UVIDENSKABELIGT EFTERORD

Vi står på tærsklen til en ny periode i sproghistorien hvor mennesker og maskiner skal til at deles om sproget (Kirchmeier et al. 2019). I USA er den typiske telefonsamtale med en bank eller et forsikringselskab et møde mellem et menneske og en maskine, og vi har allerede eksempler på det samme i Danmark. Ring til Nordea, og du bliver betjent af en virtuel assistent! Det åbner nye muligheder for lingvistikken. Ligesom fonetikken i løbet af de seneste årtier har udviklet sig fra en videnskab for specialister med interesse for sproget per se til en støttedisciplin for taleteknologien og derigennem har kunnet mangedoble sin studentermasse og nyttiggøre flere hundrede års lingvistiske observationer, står den kompositionelle semantik over for en tilsvarende mulighed. Hvordan skal en automat vide hvad et ikke-genkendt ord betyder? En oplagt mulighed for sprogteknologen er at bruge metoder som dem vi har diskuteret her, med afsæt i lingvistisk refleksion og erfaring (Henrichsen 2021). Ørstedes nyordssamling, med sin stilistiske spændvidde, sit kompositionelle udgangspunkt og sin rige blanding af kendte og glemte morfemer, vil være nærmest ideel til afprøvning af dialogsystemer der prøver at opføre sig fornuftigt. Det er bemærkelsesværdigt at den gamle hædersmand i den grad har kunnet række ud over århundrederne som inspirator.

Peter Juel Henriksen, seniorforsker  
Dansk Sprognævn  
pjh@dsn.dk

Ørstedes nyordssamling er tilgængelig som databasen HCOe\_nyord.01 med oplysninger om oprindelig bogstavering, moderne bogstavering, ordklasse, forekomst i kurante ordbøger og morfologiske analyser. Databasen kan downloades fra [dsn.dk/Oersted](http://dsn.dk/Oersted). Ordstoffet er afskrevet og korrekturlæst fra Nielsen (1981, II). En varm tak til Johanne Niclasen Jensen og Alexander Ibsen Dernild for dette arbejde. Det Kgl. Videnskabernes Selskab har i 2020 genudgivet Nielsen (1981, II), se [www.royalacademy.dk/da/Aktuelt/Oersted-ordbog](http://www.royalacademy.dk/da/Aktuelt/Oersted-ordbog). Korpus DAGW er tilgængeligt som ren tekst fra Digitaliseringsstyrelsens portal [www.sprogteknologi.dk](http://www.sprogteknologi.dk); til artiklens undersøgelse er benyttet Dansk Sprognævns PoS-annoterede version, som frigives til offentligheden primo 2022.

## LITTERATUR

- Barker, C. 2003. A gentle introduction to Type Logical Grammar, the Curry-Howard correspondence, and cut-elimination. *LSA Linguistic Society of America*, u.å. <https://semanticsarchive.net/Archive/DMzOTFiN/barker.gentle.tlg.pdf>. (tilgået 7. november 2021).
- Becker-Christensen, C. 2021. *Dansk retskrivning, bogstav – lyd – bogstav*. Odense: Syddansk Universitetsforlag.
- Christensen, D.C. 2009. *Naturens tankelæser*. København: Museum Tusulanums Forlag.
- Derczynski, L. et al. 2021. The Danish Gigaword Corpus, proceedings efter NODALIDA-21. *NoDaLida*. 2. juni 2021. <https://www.derczynski.com/papers/dagw.pdf> (tilgået 7. november 2021)
- Gram-Andersen, K. 1992. *The purple-eared monster – and its relations. A study in word formation*. Lewes: The Book Guild Ltd.
- Henriksen, P.J. 2021. H.C. Ørsted som leverandør til dansk sprogteknologi. Y. Goldshtein et al. (red.), *18. Møde om Udforskningen af Dansk Sprog, Aarhus Universitet, 8.-9. oktober 2020*, 235-250. Aarhus: Aarhus Universitet.
- Henriksen, P.J. 2001. GraSp: Grammar learning from unlabelled speech corpora. D. Roth & A. van den Bosch (red.), *Proceedings after 6th Conference on Natural Language Learning (COLING-2002 in Taipei)*, 22-28. Stroudsburg: Association for Computational Linguistics. DOI: 10.3115/1118853.1118865.

- Henrichsen, P.J. 1998. Does the sentence exist? Do we need it? K. Korta, X. Arrazola & F.J. Pelletier (red.), *Discourse, interaction and communication. Proceedings of the Fourth International Colloquium on Cognitive Science (ICCS-95)*, 1-24. Amsterdam: Springer Netherlands. DOI: 10.1007/978-94-015-8994-9.
- Kirchmeier, S. et al. 2019. *Dansk sprogteknologi i verdensklasse. Rapport fra Sprogteknologiudvalget under Dansk Sprognævn nedsat af Kulturministeriet*. København: Dansk Sprognævn.
- Nielsen, N.Å. 1981. *Sprogrenseren H. C. Ørsted*. Trøjborg: Aarhus University Press.
- Skautrup, P. 1953. *Det danske sprogs historie*. bd. III. København: Gyldendal.
- Togeby, O. 1999. Kategorial grammatik for dansk. M. Kunøe & P. Widell (red.), 7. *møde om Udforskningen af Dansk Sprog*, 349-360. Aarhus: Aarhus University Press.
- Wivel, P. 2002. Modelteoretisk semantik som meningsteori. O. Togeby & P. Widell (red.), *M/K: Mod og Kvindehjerte, Festskrift til Mette Kunøe*, 154-184. Aarhus: Modtryk.
- Wood, M.M. 1993. *Categorial Grammars*. London: Routledge.
- Ørsted, H.C. 1826. *Erindringsord af Forelæsninger over Chemiens almindelige Grundsetninger*, citeret fra N.Å. Nielsen, 1981, *Sprogrenseren H. C. Ørsted*. Bd. II 60. Trøjborg: Aarhus Universitet.
- Ørsted, H.C. 1835. *Erindringsord til Forelæsninger over Lyset*. Citeret fra N.Å. Nielsen, 1981, *Sprogrenseren H. C. Ørsted*. Bd. II. 85 Trøjborg: Aarhus Universitet.
- Ørsted, H.C. 1814. *Tentamen nomenclaturæ chemiæ omnibus linguis Scandinavico-Germanicis communis*, citeret fra N.Å. Nielsen, 1981, *Sprogrenseren H. C. Ørsted*. Bd. I, 7-40. Trøjborg: Aarhus Universitet.