

HENRIK HOLMBOE

## Dansk Radiærordbog

### Abstract

#### *A Radial Dictionary of Danish.*

A Radial Dictionary is a KWIC-concordance, where instead of using key words key letters are used. Any pair of letters can be used as a key to identify and find all words containing a specific substring of two or more letters. All words containing the substring in question will be found together in the dictionary whether the substring is word initial, medial or word final. Thus the radial dictionary can be used as a morpheme dictionary as well. For people interested in derivation and composition, which are very frequent phenomena in Danish, the dictionary gives easy access to examples and material difficult to find otherwise.

Jeg er glad for denne lejlighed til at fremlægge resultaterne af mit arbejde med Dansk Radiærordbog ved de nordiske datalingvistikdage i Reykjavik. I 1986 kunne jeg på Säby Säteri præsentere planer for udarbejdelse af radiærordbogen for et mindre nordisk forum af kolleger. Ringen slutes således med denne rapport.

Både ordbogen og ordbogens koncept er nyt for danskens vedkommende; derfor vil jeg indledningsvis redegøre for ordbogens opbygning: En radiærordbog gør det muligt at finde alle ord, der rummer en valgt substreng på to eller flere bogstaver, cf. fig 1. Denne substreng kan være en konsonantgruppe, en stavelse, et morfem, en del af et kompositum eller en derivationsendelse. Alle de ord, der rummer den valgte substreng, vil findes samlet på ét sted i ordbogen, hvadenten substrengen står initialt, medialt eller finalt. I en almindelig ordbog, dvs. en ordbog sorteret alfabetisk fra ordenes begyndelse (progressivt alfabetiseret), kan man nemt finde alle ord, der begynder med et bestemt præfix eller en bestemt stavelse. I en retrograd ordbog (finalalfabetiseret ordbog) er det tilsvarende let at finde alle ord, der slutter på et bestemt suffix eller en bestemt stavelse eller på en bestemt gruppe af bogstaver. Radiærordbogen kombinerer disse to muligheder med muligheden for også at finde ordmediale segmenter.

Denne specielle måde at opstille ordene på tilvejebringes af en algoritme, som placerer ordet lige så mange forskellige steder i ordbogen, som det rummer bogstavpar. Hvis man har en bestemt ordmængde og lader denne ordmængde gennemløbe den radiære sortering, vil det føles som en eksplosion af ordmængden, i det mindste for et sprog som dansk, som er relativt rigt på komposita

reprografisk	kraftfoder
litograf	kraftidiot
litografere	kraftig
litografering	kraftkilde
litografi	kraftløs
litografisk	kraftprøve
hektograf	kraftspring
hektografere	kraftstation
hektografering	kraftudfoldelse
fotograf	kraftudtryk
fotografere	kraftværk
fotografering	vandkraft
fotografi	spændkraft
fotografialbum	købekraft
fotografiapparat	tyngdekraft
fotografiramme	bagekraftig
fotografisk	legekraft
nedfotografere	viljekraft
telefotografi	virkekraft
telefotografisk	dømmekraft
pressefotograf	kernekraft
farvefotografi	kernekræftværk
filmfotograf	sekraft
røntgenfotografere	hestekraft
kanonfotograf	slagkraft
mikrofotografere	sprængkraft
mikrofotografering	ikrafttrædelse
mikrofotografi	ikrafttræden
amatørfotograf	ikrafttrædt
kryptograf	centrifugalkraft
kryptografi	tangentalkraft
kartograf	centripetalkraft
kartografi	maskinkraft
kartografisk	donkraft
ortografi	skaberkraft
ortografisk	lærerkraft
autograf	urkraft
autografjæger	arbejdskraft
autografsamler	modstandskraft
palæograf	anslagskraft
palæografi	indbildningskraft
palæografisk	tiltrækningskraft
giraf	manddomskraft
girafgantisk	guddomskraft
girafftrusser	livskraft
papiraffald	livskraftig
kraft	lovskraft
kraftanstrengelse	drivkraft
kraftcentral	agorafobi
kraftesløs	doktorafhandling

Figur 1.

og derivativer. Mit udgangspunkt var en ordbog på ca. 50.000 ord, og resultatet blev en radiærordbog på ca. 400.000 ord. En sådan sortering lader sig naturligvis i praksis kun gennemføre ved hjælp af en datamat.

Ordbogen anvendes på følgende måde: Først vælger man den substreng, som man vil finde eksempler på. Derefter anvender man substrengens sidste bogstavpar som opslagsnøgle. Alle ord i ordbogen er ordnet efter bogstavpar, som er trykt med fed skrift. Den nævnte nøgle vil altså altid passe til et bogstavpar i ordbogen. Med nøglen som udgangspunkt skal man nu søge retrogradt alfabetisk mod ordets begyndelse, og når man har fundet overensstemmelse med den ønskede substreng, vil hele materialet, dvs. alle eksempler, være samlet i de følgende linier, alfabetisk sorteret, først retrogradt, derefter—dvs. alt andet lige—progressivt. Denne bevægelse fra bogstavparret først til venstre, så til højre, har fremkaldt associationen om radier, og derfor betegnelserne radiærordbog og radiær sortering.

Sorteringsalgoritmen er implementeret på følgende måde, cf. fig.2: Ethvert ord, man ønsker at sortere, kopieres over i lige så mange ordfiler, som ordet rummer bogstavpar. Et ord på fem bogstaver kopieres således over i fire forskellige filer, da der jo er fire bogstavpar i en streng på fem bogstaver. Der findes en specifik fil for hvert enkelt bogstavpar; denne fil vil rumme alle eksempler på netop dette bogstavpar. Ethvert ord transformeres nu til en ny streng, der består af spejlbilledet af den del af ordet, der går forud for det bogstavpar, der er karakteristisk for filen, efterfulgt af resten af ordet i dets uændrede form. Disse nye strenge alfabetiseres nu normalt i overensstemmelse med det pågældende sprogs alfabetiseringsregler, og derefter spejles første del af strengene tilbage igen til deres oprindelige form, og hermed er selve sorteringen gennemført. For at gøre resultatet mere brugervenligt har jeg centreret bogstavparret og trykt det med fed skrift. Denne sidste operation gør det også iøjnefaldende, at radiærordbogen kan betragtes som en bogstavparkordans ordnet og opstillet efter KWIC-princippet.

Som nævnt kan man i radiærordbogen finde eksempler på enhver forekommende substreng i sproget. Hvis det, man søger, er eksempler på forskellige bogstavkombinationer, vil en radiærordbog formentlig være til hjælp i ethvert sprog. Hvis man vil bruge radiærordbogen som morfemordbog, vil den være mere eller mindre nyttig afhængigt af nogle typologiske karakteristika ved det sprog, man studerer. Hvis sproget er som dansk derved,

- at det har mange derivativer og komposita,
- at det har en kreativ og innovativ orddannelse,
- at dets stammer og derivativer er relativt stabile og
- at derivation og komposition iværksættes ved agglutination,

så kan radiærordbogen skaffe sin bruger nem adgang til materiale, som det ellers ville være vanskeligt at samle sammen.

Hér skal man formentlig også søge en del af forklaringen på, at fænomener som komposition og derivation i dansk er så sparsomt udforskede og beskrevet. Fænomenerne er efter min mening vigtige, specielt set i et oversættelsesperspektiv. Et sprog som dansk anvender jo orddannelsen til at løse en række problemer, som på f.eks. engelsk løses af syntaks og ordstilling. Med andre ord: Hvis man ønsker samme information om dansk, som man får om engelsk ved at iagttage svarene på spørgsmål vedrørende syntaks, hvilke spørgsmål skal man da stille? Mit svar vil være: Naturligvis syntaktiske spørgsmål, men tillige spørgsmål vedrørende orddannelse.

Som ovenfor anført er ordningskriteriet for radiærordbogen alene bogstavstrenge, altså et rent overfladekriterium eller—anderledes udtrykt—et kriterium, der alene tager hensyn til sprogets udtryksside. I hvilket omfang vil en sådan ordbog kunne anvendes som morfemordbog? Vil ordbogen ikke give forkerte eller i det mindste upræcise oplysninger, når man søger efter størrelser på tegnniveau og ikke kun efter udtrykssidens inventar? Naturligvis er betænkeligheder af denne type berettigede.



Figur 2.

Fordi en substreng er identisk med et morfems udtryksside, er denne substreng ikke nødvendigvis et eksempel på dette morfem; så set fra morfemanalysens synspunkt vil der være „støj“ i et opslag; algoritmen overgenererer, men ikke meget. Betragter man en række ord med et bestemt morfem, vil man meget let kunne se bort fra de irrelevante og distraherende eksempler, cf. fig. 3. Fordelen ved at sortere efter bogstavstreng er, at alle mulige eksempler er samlet. Også hvis man var interesseret ikke i morfemer, men f.eks. i konsonantgrupper, vil alle eksempler være til stede samlet, og ikke kun de eksempler, som ville slippe igennem det filter, som en hvilken som helst morfemteori vil etablere.

sammenklæpelig  
 stænklop  
 skulderklap  
 tænderklaprende  
 skyklap  
 kollaps  
 diskusprolaps  
 volapyk  
 plapre  
 efterplapper  
 overlåpning  
 overlåppe  
 papirlåp  
 slåp  
 slåp-af-  
 slåppe  
 slåppe-af-  
 slåppelse  
 slåpsvans  
 afslåppe  
 afslåppelse  
 papirslåp  
 æskulåpstav

*Figur 3.*

Til sidst vil jeg nævne, at ordbogen allerede har bevist sin praktiske anvendelighed på Hovedstadens Ordblindeskole i København. Denne skoles klienter har det specifikke sproghandicap, at de vanskeligt kan isolere og genkende ordenes bogstaver enkeltvis, og derfor opbygger de deres læsefærdighed ved at lære at genkende større grupper af bogstaver, hele ord eller dele af ord. Radiærordbogen har gjort det lettere for skolens lærere at sammensætte relevant undervisningsmateriale for klienterne.

Afdeling for Datalingvistik  
 Handelshøjskolen i Århus  
 Fuglesangs Allé 4  
 8210 Århus V