

*Henrik Holmboe*

## Dansk radiærordbog

### Abstract

The Radial Dictionary of Danish makes it possible for its user to find all Danish words containing a certain string of characters, the minimum length of the string being two characters. This principle of arrangement prove show useful especially for the description of a language like Danish with very rich composition and derivation. The typical search string will be a morph or morpheme, and the radial arrangement of this dictionary allows you to find all words containing a certain morph or morpheme no matter whether its position is word initial, word internal or word final.

Dansk radiærordbog er en ordbog, der belyser en bestemt del af sprobygningen i det danske sprog. Med et andet datamateriale som udgangspunkt ville det radiære sorteringsprincip også kunne anvendes til etablering af beskrivelser af sprogbrugen; men dette har ikke været formålet med Dansk radiærordbog i dens nuværende udformning. Radiærordbogen gør det muligt at finde ord, hvori der indgår en bestemt delstreng, hvad enten den valgte delstreng står initialt, medialt eller finalt i ordet. Navnet radiærordbog er en nydannelse, ligesom vist nok også fænomenet, og hentyder til den fremgangsmåde, man skal anvende, når man søger i ordbogen. En mere teknisk præcis betegnelse ville være "en bogstavparkordans over en lemmamængde opstillet efter KWIC-princippet", idet det valgte bogstavpar kan sidestilles med nøgleordet (keyword), og den forudgående og efterfølgende del af ordet kan sidestilles med den kontekst, i hvilken nøgleordet forekommer (Fig. 1).

Af det allerede anførte fremgår det, at den primære søgenøgle er et bogstavpar. Med udgangspunkt i bogstavparret skal man derefter lede først i retning mod strengens begyndelse, derefter mod strengens afslutning, idet ordene er ordnet tilsvarende: bogstavparret er udgangspunktet, og bogstavparrene er ordnet efter de almindelige principper, der gælder for alfabetisering af dansk. Med bogstavparret som udgangspunkt er ordene derefter sorteret retrograd alfabetisk mod ordets begyndelse og derefter som næste ordningsprioritet progressivt alfabetisk mod ordets slutning. Denne bevægelse fra "centrum" til venstre og derefter fra "centrum" til højre har fremkaldt

associationen om radier, og deraf navnet radiærordbog.

Det er klart, at radiærordbogen kan give oplysninger om bogstavpar og deres belæg. Dette er i sig selv interessant i en række sammenhænge både kommercielt og videnskabeligt, f.ex. i forbindelse med kryptering. Det kan nævnes, at der allerede findes en publiceret opgørelse af den relative frekvens af danske enkeltbogstaver (Fig. 2), og at en opgørelse over bogstavpars frekvenser i Dansk Retrogradordbog sammenlignet med frekvenserne i en tilsvarende ordmængde af løbende tekstord (dvs. sprogbrugssynspunktet) foreligger færdig omend endnu upubliceret (Fig. 3).

Hvis man ønsker at beskæftige sig med delstrengene, der er større end to bogstaver, kan sådanne søgestrengene nemt opbygges med bogstavparret som den højre afslutning som udgangspunkt. Den delstreng, der typisk vil være relevant for en række sproglige undersøgelser, er en orddel af typen morfem eller morf. Afhængigt af lingvistisk teori og skole er der uenighed om, hvordan et morfem eller et morf skal defineres. I så henseende er radiærordbogen neutral: den fordrer ikke en bestemt morfem- eller morfdefinition for at kunne give et svar. Grunden hertil er, at udgansniveauet ikke er et tegnniveau, men et niveau, der er rent udtryksplan, nemlig et bogstavpar. En egentlig morfemkonkordans ville have været mindre omfangsrig, men til gengæld have været teoriafhængig, og dermed bundet til tid og sted, på en måde, som den nuværende radiærordbog ikke er.

En bogstavparkordans over en ordmængde bliver meget omfangsrig i forhold til ordmængden på samme måde, som en liniekonkordans bliver omfangsrig i forhold til den tekst, der er dens udgangspunkt. Årsagen er, at et ord på  $x$  bogstaver skal kunne findes  $x-1$  steder i konkordansen; således består et ord på syv bogstaver af seks bogstavpar, (f.ex. består ordet BOGSTAV af BO OG GS ST TA AV). At producere en bogstavparkordans som Dansk radiærordbog på en ordmængde på ca. 50.000 ord lader sig folgelig for alle praktiske formål ikke gøre i én arbejdsdag. Materialet må opdeles, og programfasen tilrettelægges modulært.

Den fremgangsmåde, der er fulgt ved udarbejdelsen af Dansk radiærordbog er først at etablere selvstændige filer for hvert enkelt bogstavpar, dvs. 29\*29=841 filer. Et ord henføres til en fil, hvis det rummer det bogstavpar, som er specifikt for filen. Derefter sorteres de enkelte filer. Hovedprincippet i sorteringsalgoritmen er som følger: For hver enkelt af de ovenfor nævnte bogstavparfiler etableres der nye filer rummende strengene, der er en transformation af ordene i de oprindelige filer. Transformationen består af et spejlbil-

lede af den del af ordet, der går forud for den pågældende fils karakteristiske bogstavpar efterfulgt af resten af ordet, idet det bemærkes, at der for de ords vedkommende, der rummer mere end én forekomst af det karakteristiske bogstavpar, skal laves en transformation for hver forekomst af det karakteristiske bogstavpar.

De således fremkomne strenge alfabetiseres normalt. Derefter spejles den første del af strengen igen som anført ovenfor, hvorefter den ønskede sortering er tilvejebragt. Diakritiske bogstaver og ord med diakritiske bogstaver kræver som altid speciel opmærksomhed, men disse fænomener frembyder ikke større vanskeligheder i denne sortering end ved andre sorteringer af danske ord baseret på alfabetisering.

Anvendt som bogstavparkordans vil en radiærordbog som den danske formentlig have en vis interesse for alle sprog uanset deres type. Ønskes radiærordbogen anvendt som morfemkonkordans, vil dens relevans være afhængig af sprogets type, specielt for såvidt angår den teknik, sproget anvender til at sammenstille og sammenknytte de betydningsbærende enheder i en sætning. En redøgørelse herfor finder man traditionelt i morfologien, orddannelseslæren og syntaxen. Afhængigt af, hvilke grammatiske indholdsstrukturer og relationsmarkører der er grammatikaliserede i sproget, vil disse være beskrevet i morfologi eller syntax. Hvis sproget er internt flekterende, vil dette anfægte rodmorfermernes fremtrædelsesform på en måde, så deres identitet ikke er sikret, hvilket igen betyder, at en radiærordbog i rollen som morfemkonkordans, hvilket netop forudsætter rodmorfermernes lineære konstans, ikke vil være specielt relevant. Heller ikke til belysning af grammatikaliserede strukturer af eksternt flekterende eller agglutinerende type vil radiærordbogen spille nogen vigtig rolle, da den kun i ringe omfang vil kunne give oplysninger, som en almindelig ordbog eller en retrograd ordbog ikke ville kunne have givet lige så fyldestgørende og formentlig på en nemmere og lettere tilgængelig måde. Ser vi derimod på sprogenes derivations- og kompositions muligheder og specielt på de tilfælde, i hvilke disse aktualiseres ved hjælp af agglutination, kan vi konstatere, at de sprog, der udnytter disse muligheder, lader orddannelsen løse en del af de problemer, der i sprog uden disse muligheder er overladt typisk til syntaxen og herunder ordstillingen.

Det vil herefter være nærliggende at antage, at en radiærordbog vil kunne vise sig nyttig til beskrivelse af et sprog som dansk, der netop er karakteriseret ved uhyre sparsom brug af ortografisk sandhi, relativt højt kompositionsindex og højt derivationsindex fordelt

på både ret højt præfixindex og specielt højt suffixindex, og navnlig, når dansk ønskes kontrasteret med sprog, hvis indices på disse områder er lavere. Hvis man nemlig vil have samme information om f.ex. dansk, som man får om f.ex. engelsk eller fransk gennem svarene på spørgsmål vedr. syntax og ordstilling, hvilke spørgsmål skal man da stille? Svaret er: man skal bl.a. stille spørgsmål til orddannelse, og vi må konstatere, at spørgsmålene er lettere stillet end besvaret. Dansk syntax har som bekendt været genstand for en del forskning og emne for en lang række publikationer. Dette gælder ikke danskens orddannelse. Det er mit håb, at Dansk radiærordbog kan være et bidrag til at rette op på denne skævhed.

Der forefindes et eksemplar af Dansk radiærordbog på Afdeling for Datalingvistik på Handelshøjskolen i Århus, hvor evt. interesse-rede vil kunne benytte den.

reprografisk  
litograf  
litografere  
litografering  
litografi  
litografisk  
hektograf  
hektografere  
hektografering  
fotograf  
fotografere  
fotografering  
fotografi  
fotografialbum  
fotografiapparat  
fotografiramme  
fotografisk  
nedfotografere  
telefotografi  
telefotografisk  
pressefotograf  
farvefotografi  
filmfotograf  
røntgenfotografere  
kanonfotograf  
mikrofotografere  
mikrofotografering  
amatørfotografi  
kryptograf  
kryptografi  
kartograf  
kartografi  
kartografisk  
ortograf  
ortografisk  
autograf  
autografilæger  
autografsamler  
palæograf  
palæografi  
palæografisk  
giraf  
girafantisk  
giraffrusser  
papiraffald  
kraft  
kraftanstrengelse  
kraftcentral  
krafttesløs

Fig. 1.

kraftfoder  
kraftidiot  
kraftig  
kraftklide  
kraftløs  
kraftprøve  
kraftspræng  
kraftstation  
kraftudfoldelse  
kraftudtryk  
kraftværk  
vandkraft  
spændkraft  
købekraft  
tyngdekraft  
bagekraftig  
lægekraft  
viljekraft  
virkekraft  
dømmekraft  
kernekraft  
kernekræftværk  
sekraft  
hestekraft  
slagkraft  
sprængkraft  
ikrafttrædelse  
ikrafttræden  
ikrafttrædt  
centrifugalkraft  
tangentialkraft  
centripetalkraft  
maskinkraft  
donkraft  
skaberkraft  
lærerkræft  
urkræft  
arbejdskræft  
modstandskræft  
anslagskræft  
indbildningskræft  
tiltrækningskræft  
manddomskraft  
guddomskraft  
livskraft  
livskraftig  
lovskræft  
drivkraft  
agorafobi  
doktorafhandling

Fig. 2

Fra Henning Spang-Hanssen  
 Hvor hyppigt optræder de forskellige bogstaver?  
 I nyt fra Sprognævnet nr. 4, marts 1970.

Ordningenummer efter fallende hyppighed	Bogstav	Hyppighed i procent	Sammenlagt hyppighedsprocent
1	e	16,6	
2	r	8,0	Nr. 1-2: 24,6
3	n	7,7	
4	t	7,2	
5	d	6,7	Nr. 1-5: 46,2
6	i	5,8	
7	a	5,6	
8	s	5,5	
9	l	5,1	
10	o	4,5	Nr. 1-10: 72,7
11	g	4,4	
12	m	3,7	
13	k	3,2	
14	v	2,7	
15	f	2,6	
16	h	2,1	
17	u	1,6	
18	b	1,4	
19	p	1,3	
20	a	1,3	
21	ø	0,9	
22	æ	0,8	
23	y	0,6	
24	j	0,6	
25	c	0,1	
		100,0 %	

W, X, Z, Z, og Q q: hver mindre end 0,02 %.  
 Middellængden er 4,8 bogstaver (dvs. der optræder et ordmelletrum gen-  
 nemsnitligt efter hver 5 bogstaver).

Fig. 3.

Danak Retrograd Ordbog				Løbende tekst			
o/ooo	Abt. fre.	Rang	Klyngre	Rang	Abt. fre.	o/ooo	
1	26	496	ab	532	1	0	
21	829	140	ab	215	170	7	
5	203	309	ac	277	77	4	
33	1282	80	ad	72	624	35	
1	30	483	ae	443	8	0	
31	1201	87	af	32	823	46	
55	2145	44	ag	34	1291	72	
1	45	445	ah	401	19	1	
3	123	364	ai	491	3	0	
3	102	378	aj	274	78	4	
24	927	117	ak	143	263	15	
93	3246	25	al	44	984	53	
16	1422	71	am	53	810	45	
120	5044	13	an	3	3525	197	
0	14	540	ao	561	0	0	
16	628	172	ap	167	206	12	
0	0	689	aq	362	0	0	
75	2976	29	ar	18	2209	124	
40	1564	61	as	122	116	38	
64	2501	37	at	45	1710	96	
8	295	271	au	295	66	4	
23	895	126	av	60	784	44	
0	16	527	aw	563	0	0	
0	13	545	ax	405	17	1	
1	33	472	ay	503	2	0	
1	34	470	az	564	0	0	
0	1	634	aa	565	0	0	
0	0	690	ab	566	0	0	
0	0	691	aa	533	1	0	
24	939	115	ba	94	468	26	
4	150	349	bb	360	33	2	
0	1	635	bc	567	0	0	
0	16	528	bd	524	1	0	
66	2572	35	be	50	828	47	
0	9	540	bf	568	0	0	
0	2	610	bq	569	0	0	
18	16	729	br	570	0	0	
4	49	716	bs	721	131	7	
0	145	352	bt	571	52	3	
0	1	636	ba	571	0	0	
20	777	148	bl	80	594	33	
0	8	567	bm	504	2	0	
2	79	407	bn	353	37	2	
23	890	123	bo	170	196	11	
0	0	692	bp	572	0	0	
0	0	693	bq	573	0	0	
28	1078	99	br	123	306	17	
6	243	292	bs	383	23	1	
1	27	492	bt	269	82	5	
12	470	211	bu	249	103	6	