

Martin Ejnar Hansen

## Er dårlige data bedre end ingen data? En diskussion af brugbarheden af de danske valgprogramdata

I de sidste 30 år har dataene fra Comparative Manifesto Project været en væsentlig del af den europæiske politologi. Talrige publikationer er det blevet til, men stort set ingen har været specifikt orienteret mod Danmark. Det er i og for sig godt. De danske valgprogramdata lider under markante problemer, der betyder, at al anvendelse skal ske med megen forsigtighed, og de resultater, der fremkommer, må tages med et stort forbehold.

1979 var året, hvor et stort europæisk forskningsprojekt så dagens lys, Manifesto Research Group, senere omdøbt til Comparative Manifesto Project (CMP). Nu 25 år senere er projektet udviklet til en af de største kilder til brug for komparativ analyse af politiske partier. Valgprogrammer siden 1945 for 25 lande<sup>2</sup> er blevet kodet ud fra 56 forskellige kategorier (Budge og Bara, 2001). Programmernes sætninger er opdelt i, hvad der betegnes som *quasi-sentences*. *Quasi-sentences* defineres som et argument, der er det verbale udtryk for en politisk ide eller sag, der kan kodes inden for de 56 kategorier (Volkens, 2001a: 34). Programmerne dækker en meget stor bredde af politikpositioner og er derfor "(...) a set of key central statements of party positions" (Budge et al., 1987: 18). Denne kodning har skabt et unikt datasæt, der har været anvendt i talrige publikationer. Datasættet har opnået en næsten monopolagtig status inden for komparativ analyse af partipositioner; kun ekspertsurveys kan siges at være i stand til at konkurrere. Omfangsmæssigt er kun de nationale valgundersøgelser datasæt af samme kaliber som CMP dataene, og disse dækker langt fra den samme tidsperiode.

De mange internationale publikationer til trods, er det kun få, der specifikt behandler Danmark. Kun Holmstedt og Schou (1987) anvender valgprogramdataene til at analysere de danske partipositioner.<sup>3</sup> Schou og Hearl (1992) anvender de danske regeringsprogrammer, der er kodet efter samme principper som valgprogrammerne. Det kan undre, at et så markant datasæt ikke har givet anledning til flere publikationer med udgangspunkt i Danmark. Men når de danske data analyseres mere grundigt, er det ikke så sært. Faktisk fristes man til at sige, at det er godt. De danske data lider af store problemer. Problemerne findes på tre niveauer: kodningen, udvælgelsen af valgprogrammer og sidst men ikke mindst, hvordan partierne placeres på en venstre/højreskala. Det sidste resultat er selvfølgelig stærkt påvirket af de to førstnævnte problemer.

### Valgprogrammer og deres anvendelse

En helt central beægggrund for at anvende valgprogrammer til at placere politiske partier på politikdimensioner er, at et valgprogram er den eneste autoritative, kollektive udtalelse om partiets politik (Budge, 2001: 50; Volkens, 2001a: 34).

Valgprogrammerne er som udgangspunkt vedtaget af generelt det højeste organ på den interne arena. Programmerne anvendes offensivt af partierne, hvilket gør, at de er skrevet til bredden af befolkningen og dermed meget ligefremme (Budge, 2001: 51).

Samtidig har konceptet valgprogrammer været kendt længe og er til stadighed en væsentlig del af et politisk partis mulighed for at give sine holdninger til kende. Dette gør valgprogrammet velegnet til at studere over tid og ikke mindst at sammenligne.

Når det er sagt, må man dog stille spørgsmålet, om valgprogrammer, taler eller politiske tekster generelt, er troværdige kilder. Iain McLean (2004) fremsætter to svar på dette spørgsmål. Et svagt og et stærkt. Det svage svar er, at data er, hvad de er, men at man med disse dokumenter kan generere data, der er bedre end anekdotisk viden om politik og bedre end usystematisk sampling og langt bedre end blot at ignorere dem. Det stærke svar lyder, at disse data giver rigtige, reliable og valide resultater, der er i god overensstemmelse med resultater fra andre metoder og tilsammen er et godt modstykke til eksempelvis vælgeradfærdsdata. Derudover giver valgprogrammerne i modsætning til for eksempel vælgerundersøgelser mulighed for at anvende nye teknikker på historiske data. Ydermere er CMP-datasættet det eneste datasæt med politikpositioner og ændringer over tid inkluderet (Budge, 2001: 50). Der er samtidig få, om overhovedet nogen, datamæssige alternativer til valgprogrammer, hvis man vil studere og sammenligne politiske systemer ud fra *policy* eller ideologi (Mair, 2001: 11), og hvor det datamæssige udgangspunkt er valgprogrammer. De datamæssige alternativer, der måtte være, vil ofte tage deres udgangspunkt i ordførerstatements eller generelle udtalelser fra individer, der ikke uden videre kan antages at tegne partiet som helhed (Budge, 1987: 18).

### *Indhold*

I CMP-dataene er politikmålene et udtryk for den vægtning – saliens – som partierne giver de enkelte politikker. Det betyder, at partiernes position – for eller imod – på de enkelte områder ikke fremgår direkte af datasættet. Det betyder dog ikke, at det ikke er muligt at måle afstandene mellem partierne. Det betyder blot, at selvom et parti ikke nævner en specifik politik, kan den politik kun antages at have en lav grad af saliens hos det parti, men ikke at partiet har en centerposition på området (Volkens, 2001a: 40). CMP-dataene måler ændringer i politikplacering på en plausibel og acceptabel måde (Budge, 2001: 50). På den anden side må det dog også påpeges, at CMP-dataene ikke kun er skabt til at placere politiske partier på politikdimensioner. For eksempel er dataene også blevet anvendt til at analysere, hvilke emner der er på den politiske dagsorden (se eksempelvis Saglie, 1998).

I modsætning til Stokes (1963) er der blandt CMP-publikationerne generel enighed om, at partiernes forskelle ikke findes gennem en direkte konkurrence mod hinanden, men snarere gennem et varierende fokus på en fælles position (Budge, 2001: 60). Budge mener, at en bekræftelse herpå er, at det er elektoralt godt at tage minoritetssynspunkter frem for majoritetssynspunkter (2001: 61).

CMP-dataene præsenteres ofte som værende særdeles reliable, og det vil som oftest være uproblematisk at anvende dataene til analyse, da dette har været gjort før og med stor succes. Dette vil dog være en meget forkert antagelse, hvis den danske del af data er inkluderet i anvendelsen. Problemerne med den danske del af CMP-dataene er mange, og i visse tilfælde synes de endda uoverstigelige.

### Problemerne med de danske data

Det første problem med de danske CMP-data er, at for de danske partier er ca. 30 pct. af alle *quasi-sentences* blevet kodet som ukodelige (Volkens, 2001a: 41), hvilket betyder, at gennemsnitligt er en tredjedel af indholdet af de danske valgprogrammer ikke medtaget i analysen. Holmstedt og Schou (1987) argumenterer for, at det danske politiske system er af en sådan karakter, at en kodning af de manglende 30 pct. vil give uanvendelige resultater, da det drejer sig om særdeles nationalt specifikke områder.

Argumentationen er ganske interessant, især når der i dele af politologien arbejdes med en ide om, at de skandinaviske politiske systemer er forholdsvis ens. (se for eksempel Lijphart, 1984; 1999). Som det fremgår af tabel 1, så er der langt fra den høje værdi for Danmark til den næsthøjeste værdi, som findes hos Sverige. Danmark har desuden en høj standardafvigelse på ca. 18 procentpoint.<sup>4</sup> End ikke for lande med en fundamental anden politisk kultur som Israel og Tyrkiet, hvor man må formode, at der kunne være regionsspecifikke, endog landespecifikke politikområder, som det ville være svært at få passet ind i kodningen, er der høje værdier på antallet af manglende sætninger.

Tabel 1. Pct. *quasi-sentences* der af koderne er blevet dømt *uncodeable* – udvalgte lande

Land	Pct. <i>uncodeable</i>	Standardafvigelse (i procentpoint)
Danmark	30,2	18,5
Norge <sup>a</sup>	0,5	36,0
Sverige <sup>a</sup>	12,5	42,9
Finland	2,3	3,5
Island	4,2	7,1
Tyrkiet	0,5	1,1
Israel	7,7	10,5
Japan	6,0	7,5
Irland	3,0	4,8

Kilde: Egne beregninger på baggrund af Budge et al. (2001); Volkens (2002).

Note: a) For Sverige gælder, at kun valgprogrammer fra 1985 og frem er medtaget. For Norge er kun valgprogrammer fra 1993 og frem medtaget. Standardafvigelser er beregnet med udgangspunkt i de enkelte observationer for hvert enkelt land.

Et bud på en forklaring af de manglende 30 pct. kunne være, at fejlprocenten varierer over tid, og hvis man holder sig til at anvende observationer fra de senere år, vil problemerne være minimale. Deles perioden 1945-2001 op i to halvdele, finder vi, at valgprogrammer i den første halvdel til og med 1973 mangler gennemsnitligt 37,8 pct. af alle sætninger, hvor den anden halvdel fra og med 1975 til 2001 gennemsnitligt mangler 23,5 pct. af alle sætninger. Selvom de senere data mangler færre sætninger end de tidlige, så er det stadig en procentdel, der er markant højere end nogen af de øvrige lande i CMP-datasættet. Men for at svare på om data bliver bedre i de senere år er det nødvendigt at opdele perioden yderligere. Gennem tabel 2, hvor perioden 1945-2001 er opdelt i seks mindre perioder, kan vi vurdere, om de senere data er bedre end de tidlige. Af tabellen fremgår det, at antallet af ukodede sætninger først falder efter 1985, men næsten en femtedel ukodede sætninger synes stadig ganske højt, ikke mindst sammenholdt med en høj standardafvigelse. Perioden fra 1995 er noget bedre, hvor kun ni pct. sætninger er ukodede, men standardafvigelsen har stort set samme værdi, og det tyder på visse problemer også her. Set over hele perioden er standardafvigelserne støt faldende. Det er tegn på, at der er mindre forskel i antallet af ukodede sætninger partierne i mellem. Men det ændrer ikke på, at det er først er i den sidste periode, at der er tale om et relativt lavt antal ukodede sætninger for de danske partier. Hvis målet for anvendelsen af data kun er at anvende data for valgene i 1998 og 2001, som er dem, der er inkluderet i sidste periode, så er valgprogramdataene brugbare, men hvis man ønsker at foretage sammenligninger over tid i udviklingen af partiplaceringer, dimensioner og lignende, så er de manglende 30 pct. over hele perioden ikke et udtryk for en enkelt outlier der trækker i den forkerte retning. De ukodede sætninger findes over hele perioden.

*Tabel 2.* Udviklingen i ukodede sætninger for Danmark 1945-2001

<b>Periode</b>	<b>Gennemsnitligt uncodeable i pct.</b>	<b>Standardafvigelse i procentpoint</b>
1945-1954	41,5	20,6
1955-1964	27,5	17,7
1965-1974	40,8	18,0
1975-1984	30,8	14,1
1985-1994	19,2	10,9
1995-2001	9,1	7,5
Hele perioden 1945-2001	30,2	18,5

Kilde: Egne beregninger på baggrund af Volkens (2002).

Note: Standardafvigelser er beregnet med udgangspunkt i de enkelte observationer for Danmark.

### *Betydningen af 30 pct.*

Det relevante spørgsmål at rejse her er naturligvis, om de manglende 30 pct. har nogen reel betydning, eller om det kun er et problem for statistikere og datamæssige *feinschmeckere*? Ud fra en dataorienteret synsvinkel er der ingen tvivl om, at 30 pct. manglende sætninger er et problem. Med mindre der i kodningen er anvendt en form for statistisk begrundet samplingteknik, for da kunne en mangel på 30 pct. være acceptabelt. Der er dog intet sted i nogen beskrivelse af CMP dataene, at der tales om sampling. Tabel 3 viser endvidere, at hvis der har været tale om sampling, så har den været ganske usystematisk. Meget taler for, at der ikke har været anvendt nogen form for samplingteknik på de danske valgprogramdata. Da det samtidig ikke er muligt at vide, hvilke 30 pct. der er udeladt, er det ej heller muligt at anvende forskellige statistiske værktøjer til at korrigere problemet eller dele heraf. Et endnu større problem er ganske givet, at når 30 pct. af alle sætninger ikke er blevet kodet, hvad er sikkerheden så for, at de resterende 70 pct. er kodet rigtigt?

Først og fremmest vil det være relevant at se på, hvorvidt det er hos bestemte typer partier, at der kan findes høje værdier af manglende *quasi-sentences*, eller om det er en generel trend. I tabel 3 fremgår de gennemsnitlige pct.er for hvert

*Tabel 3. Gennemsnitligt antal quasi-sentences, der er bedømt uncodeable for de danske partier 1953-2001*

<b>Parti</b>	<b>Pct.</b>	<b>Standardafvigelse (procentpoint)</b>
Socialdemokratiet	29,1	19,7
Det Radikale Venstre	27,2	15,9
Det Konservative Folkeparti	21,4	13,2
Centrumdemokraterne	34,6	13,0
Danmarks Retsforbund	32,7	21,2
Socialistisk Folkeparti	24,7	15,9
Danmarks Kommunistiske Parti	44,0	11,6
Liberalt Centrum	16,8	-
Dansk Folkeparti	19,6	21,8
Fælles Kurs	5,8	-
Kristeligt Folkeparti	16,8	6,8
De Uafhængige	25,9	12,3
Venstre	24,8	14,3
Venstresocialisterne	44,2	24,1
Fremskridtspartiet	30,5	16,6
Enhedslisten	6,8	3,8

Kilde: Egne beregninger på baggrund af Volkens (2002).

Note: For Liberalt Centrum og Fælles Kurs gælder, at der kun findes én observation. Standardafvigelser er beregnet med udgangspunkt i de enkelte observationer for hvert parti.

parti samt standardafvigelserne for disse. Her er det tydeligt, at det er et generelt problem for alle partier, at fejlmargen er høj. Ligeledes må der konstateres store udsving for de enkelte partier. Standardafvigelser på over 20 procentpoint er meget høje og er ikke med til at øge reliabiliteten af data. Kun grundig empirisk analyse af de anvendte valgprogrammer vil gøre os i stand til at få be- eller afkræftet, at de manglende 30 pct. sætninger har betydning for reliabiliteten, og om denne øges ved at få kodet flere sætninger. Men som det senere skal blive diskuteret, er det desværre ikke en farbar vej, da vi ikke kender alle valgprogrammerne, som er blevet kodet.

### *De indsamlede programmer*

Et andet problem ved CMP-dataene er selve programmerne, der ligger til grund for kodningen. For danske forhold er det specielt interessant, at der som oftest er få ændringer i partiprogrammerne. Det kan for eksempel nævnes, at Socialdemokratiet havde samme program fra 1913-1960 (Holmstedt og Schou, 1987: 180). Derfor er det ikke partiprogrammerne, der er anvendt i CMP-dataene, men derimod valgprogrammerne. Spørgsmålet her er så, om partiernes valgprogrammer er reliable datakilder, eller om der er problemer med disse. Først og fremmest er det et markant problem, at valgprogrammer på lige fod med partiprogrammer eller for den sags skyld taler fra politikerne kan anskues som strategiske udtalelser. Hvis dette accepteres, så er det klart, at de danske data må behandles med en vis forsigtighed, og den vægt, det tillægges, må gøres afhængig af dette problem. For det andet er problemet, at valgprogrammerne ikke nødvendigvis er et udtryk for partiets holdning på alle politikområder. De er snarere udtryk for partiernes holdning på de respektive områder, som de hver især mener, vil være valgets tema. For det tredje er problemet, hvilke valgprogrammer der er blevet kodet. Det kan med god ret antages, at partierne har mere end en slags valgprogram liggende på lager.

Eksempler<sup>5</sup> på de to problemer er Det Radikale Venstres valgprogram fra 1981, hvor bekæmpelse af ungdomsarbejdsløshed er hovedtemaet, hvorimod Kristeligt Folkepartis program fra samme år har ønsket om en ny regering som overskrift. I 1990 figurerer der to forskellige kilder for Socialdemokratiet, et tillæg eller annonce i *Det Fri Aktuelt* samt et andet formodentlig mere generelt program. Ved Enhedslisten første deltagelse i et valg i 1994 anvendes en valgavis for København som kilde. Det anvendte program for Venstre i 1964 er en tale holdt af den senere minister Anders Andersen, men som ikke var leder af partiet i 1964, mens det anvendte program for 1966 er "Til vælgerne i Givekredsen". For Socialistisk Folkeparti i 1987 er det et udkast til valgprogram, der anvendes. I 1990 mener Det Radikale Venstre, at det er arbejdsløshed, der er det centrale tema i valgkampen. Venstre mener derimod ved samme valg, at det er europapolitik, mens Det Konservative Folkeparti har et generelt valgmanifest. Endelig er der problemerne for eksempelvis 1984, hvor der ikke foreligger oplysninger for noget parti om, hvilket valgprogram der er blevet kodet. Den mangel er endnu mere problematisk end de øvrige, da den forhindrer en let tilgængelig replikation af kodningen med andre redskaber. Hvis resultaterne for 1984 øn-

skes replikeret, må alle valgprogrammer, valgudtalelser, endog udkast kodes, før det er muligt at sige, hvilket der har resultater i nærheden af det publicerede. Den manglende systematik i valget af kodningsmateriale gør det ikke nemmere at afhjælpe den manglende information.

De programmer, der er blevet anvendt som grundlag for kodningen, er i visse tilfælde deciderede fejlkilder. Programmer rettet mod lokale kredse, taler fra folketingsmedlemmer og udkast til valgprogrammer, er som nævnt blevet anvendt. I visse tilfælde skyldes de problematiske partiplacements, at de anvendte valgprogrammer lægger vægt på noget forskelligt. I 1990 har det kodede valgprogram for Venstre titlen "Det nye Europa, Venstre viser vejen", mens Det Radikale Venstres program ved det samme valg hed "Arbejdsløshed øverst på dagsordenen". Ved 2001 valget havde Socialdemokratiets program titlen "Mennesker først", mens regeringspartneren i Det Radikale Venstre ville have "Ansvar for resultater også i fremtiden". Venstres program havde titlen "Tid til forandring".<sup>6</sup> Et program med et tema om ungdomsarbejdsløshed kodes efter samme standard som det valgprogram, der omhandler europapolitik, og samme standard anvendes til det overordnede parti-program. Dermed opstår der markante forskelle imellem partierne, alene på baggrund af at forskelligt materiale er anvendt. Denne kritik gør det endvidere svært overhovedet at anvende de indsamlede programmer til nogen form for analyse.

Problemet med at anvende forskellige data ville i og for sig ikke være stort, blot man accepterede de begrænsninger, det gav, og at partierne ikke nødvendigvis lægger den samme vægt på hvert enkelt spørgsmål, som Robertson (1976) har påvist. Men de danske CMP-data har begået denne fejl og har ukritisk kodet forskellige valgprogrammer som udtryk for det samme, men et generelt valgmanifest kan have større dybde på de økonomiske spørgsmål end et valgmanifest, der handler om europapolitik. At man enkelte år har anvendt valgprogrammer rettet mod en specifik kreds frem for et nationalt program eller for den sags skyld udkast til valgprogrammer frem for det endelige, er faktorer, der ikke er med til at understøtte data.

### *Kritikken af CMP generelt*

En anden kilde til problemerne med CMP-dataene er af mere generel karakter; selve kodningsmetoden, som blandt også er kritiseret af Pelizzo (2003) samt Laver et al. (2003). CMP-dataene består af 56 kodningskategorier. Partiplaceringen, som vist i tabel 4, er baseret på 13 "højreorienterede" og 13 "venstreorienterede" kategorier som foreslået af Laver og Budge (1992). Det vil sige, at hvis et parti scorer særdeles højt i én af højre-kategorierne og ikke omtaler de resterende kategorier, så vil det parti blive placeret særdeles markant til højre i det politiske landskab. Faktum kan blot være, at partiet agerer rationelt og har fundet et nicheområde, hvor de øvrige partier ikke konkurrerer. Det nicheområde kan være et, der ikke figurerer i CMP-dataenes kodningskategorier. Det må også tages i betragtning, at disse kategorier blev udviklet for små 20 år siden med udgangspunkt i det britiske politiske system, at de har været anvendt som generelle kategorier for alle lande, og sidst men ikke mindst, at de vestlige sam-

fund har været igennem store omvæltninger i de 20 år.<sup>7</sup> De grundlæggende antagelser, som dataene er baseret på, kan derfor være medskyldige i, at de resultater, der fremkommer, ikke er i overensstemmelse med virkeligheden, eller rettere er langt fra nogen "normal" opfattelse af det danske politiske landskab.

Men det må dog anerkendes, at den generelle kritik der kan rejses til trods, så har CMP-dataene givet politologer en mulighed for at foretage komparative analyser, hvor data er kodet efter en fælles målestok. Problemerne med placeringen af de danske partier behøver som sådan ikke at være grundet i selve kodningsmetoden, for selv den bedste kodningsmetode må give op over for dårlige data.

### **Placeringen af de danske partier**

Før CMP-dataene kan anvendes til placering af partier, må de igennem en eller anden form for bearbejdning. Der er flere muligheder for at foretage denne bearbejdning. Fire modeller til placering af partier er sammenfattet hos Gabel og Huber (2000), hvoraf de finder, at deres egen såkaldte *vanilla method* er den bedste. Men også metoderne præsenteret af Laver og Garry (2000) samt til dels Laver og Budge (1992) viser sig at være statistisk acceptable. Metoden fremlagt af Laver og Budge (1992) er blevet den, de fleste anvender (se eksempelvis Kim og Fording, 1998; Gabel og Huber, 2000), og den vil også blive anvendt i denne artikel. Der er tale om simpel sammenlægning af alle "højreorienterede" udsagn fratrukket alle "venstreorienterede" udsagn for til sidst at fremkomme med en værdi mellem -100 og 100, der viser partiets placering på en endimensional skala. Metoden er simpel, men alligevel en metode, hvis forklaringskraft er tilstrækkelig høj til at gøre den brugbar for analysen (Gabel og Huber, 2000). Fordelen ved denne metode er dens enkelhed, der gør det muligt – i hvert fald teoretisk – at analysere partiernes position helt ned på de 56 politikområder, som CMP-dataene består af. Dette kunne være særdeles nyttigt i de tilfælde, hvor partiplaceringen, der fremkommer fra CMP-dataene, synes at være problematisk i forhold til den viden, der allerede besiddes om det danske politiske system, videnskabeligt som anekdotisk. Problemet med at kigge på de enkelte politikområder er blot, at der ofte kun er få observationer for hvert enkelt parti, og det er derfor problematisk at anvende de deltaljerede værdier til at positionere partierne i forhold til hinanden, da små ændringer på et politikområde vil give stor variation, hvis områderne analyseres hver for sig. Dette er ikke et problem i samme grad, når man arbejder med en aggregeret værdi af de forskellige politikområder. Derimod kan man anvende værdierne på de enkelte politikområder til at sige noget om, hvilke emner der er på dagsordenen, eller hvor megen vægt de enkelte partier lægger på hvert enkelt område.

I tabel 4 fremgår partiplaceringer, som er opnået ved brug af CMP-dataene ved anvendelse af metoden fremlagt af Laver og Budge (1992). De partiplaceringer, der fremkommer ved at anvende denne metode, har en meget lille sammenhæng med en generel opfattelse af det danske politiske landskab. Som det fremgår af tabellen, er der ganske store variationer for de enkelte partier mellem de forskellige valg. Generelt er resultaterne ganske overraskende.



Tabel 4. Partiplaceringer ud fra valgprogramvariabel

Valg	Partiplaceringer
1953	DKP-S-RV-V-KF-DR
1957	DKP-S-RV-DR-V-KF
1960	SF-S-RV-V-UAF-KF
1964	SF-S-RV-V-KF-UAF
1966	SF-RV-S-LC-V-KF
1968	VS-SF-RV-S-V-KF
1971	S-SF-RV-V-KF
1973	DKP-SF-RV-S-Z-CD-DR-KF-KRF-V
1975	DKP-SF-VS-S-KF-V-CD-RV-KRF-Z
1977	SF-VS-DKP-S-RV-Z-DR-V-KF-CD-KRF
1979	SF-VS-RV-S-DR-KRF-KF-Z-V-CD
1981	SF-VS-S-RV-KRF-V-Z-CD-KF
1984	SF-VS-S-RV-KRF-V-CD-KF-Z
1987	FK-S-RV-SF-CD-KF-KRF-Z-V
1988	SF-RV-S-KRF-CD-Z-KF-V
1990	S-RV-SF-CD-KF-V-KRF-Z
1994	EL-SF-S-RV-CD-KF-Z-V
1998	SF-EL-RV-S-CD-DF-KF-V-Z-KRF
2001	SF-S-EL-RV-KF-KRF-DF-V

Kilde: Egne beregninger på baggrund af Volkens (2002).

At Socialdemokratiet under Jens-Otto Krag i 1971 skulle være det mest venstreorienterede parti synes ganske overraskende. At Fremskridtspartiet med sit indtog i dansk politik i 1973 skulle have placeret sig i midten af dansk politik, er måske også et overraskende fund. Og at Venstre i 1987, 1988, 1994 og 2001 skulle være det mest højreorienterede parti, er vel sagtens noget, der også må siges at være højst overraskende. En sammenligning over tid med Venstre og Fremskridtspartiet giver det paradoksale resultat, at Venstre har haft flere positioner til højre for Fremskridtspartiet, end det omvendte er tilfældet. Ligeledes er det meget overraskende, at Det Radikale Venstre ved valgene i 1987 og 1990 skulle være mere venstreorienteret end Socialistisk Folkeparti. Især placeringen i 1990 synes at undre, da partiet i 1990 gik til valg som et regeringsparti. Enhedslistens placering som mindre venstreorienteret end SF i 1998 og end både Socialdemokratiet og Socialistisk Folkeparti i 2001 er en placering, der, hvis den er sand, sandsynligvis vil give store problemer for både Enhedslistens, Socialistisk Folkepartis og Socialdemokratiets selvopfattelse. Det samme gør sig gældende for forholdet mellem Socialistisk Folkeparti og Venstresocialisterne fra og med valget i 1975 frem til valget i 1987, hvor Socialistisk Folkeparti skulle være det mest venstreorienterede parti.

Partiplaceringerne leder hen til to påstande; enten er partierne placeret som tilfældet er, eller også er der problemer med den måde, hvorpå partiernes placering bestemmes. Førstnævnte er næppe tilfældet. Ekspertsurveys, analyser af parlamentsafstemningerne og anvendelse af samlinger af politiske tekster giver ikke et tilsvarende billede af de politiske partiers indbyrdes placering. Sidstnævnte giver den mest valide kritik, ikke mindst set i lyset af Pelizzo (2003), som viser, at der er problemer med partiernes placering i mange lande i CMP datasættet. Men for den danske case kan de tidligere viste problemer ej heller undgå at influere på placeringen af partierne, især problematikken omkring forskellige emner, som partierne lægger vægt på kunne tænkes at have en stor indflydelse. Det er ikke urimeligt at antage, at et valgprogram, der er specialiseret på et politikområde, vil give en kontraintuitiv partiplacering, når dette program sammenlignes med mere generelle af typen. Overordnet set kan man dog med rette sige, at her giver dårlige data også dårlige resultater, som ikke kan bakkes op af nogen ekstern validering.

### **Fremtidens forskning inden for området**

Den teknologiske udvikling i de seneste år har givet en udfordring til CMP-dataene generelt. Computere kan anvendes til at kode en tekst ud fra en kodebog. Princippet er det samme som for håndkodning, blot mere troværdigt. Det vil med computerkodning være muligt igen og igen at få det samme resultat, når blot kodebogen også er den samme, og som John Garry (2001) viste, så er der stort set overensstemmelse mellem håndkodning og computerkodning. Udfordringen er af en sådan karakter, at CMP-projektet må tage den seriøst og søge at imødegå den. Hvis computerkodning af valgprogrammerne giver samme resultat som håndkodning, så vil én del af kritikken kunne afvises. Ligeledes vil det med en gennemarbejdet kodebog til anvendelse for computerkodning være muligt at mindske antallet af ukodede sætninger betragteligt. Man skal dog ikke være blind for, at kritikken i stedet vil kunne rettes mod den kodebog, der bliver anvendt. Ligeledes må det tages med i overvejelserne, at CMP-dataene, minus den danske del, generelt giver relativt troværdige afspejlinger af partipositioner. En tredje vej i valgprogramforskningen er i de seneste par år blevet søgt udarbejdet af politologerne Michael Laver og Kenneth Benoit, der har skabt en ny metode til at kvantificere politiske tekster, kaldet *Wordscores* (se Laver og Benoit, 2002; Benoit og Laver, 2003; Laver et al., 2003). *Wordscores* går ud på, at der findes et sæt af politiske tekster, der ønskes analyseret. Dernæst findes en eller flere referencetekster, det vil sige tekster, som det er muligt at estimere eller antage en position på. Denne estimation eller antagelse er altafgørende for analysens videre færd. Er analysegenstanden eksempelvis valgprogrammer, foreslår Laver et al. (2003), at der anvendes ekspertsurveys af partiernes placering til at fastslå de *a priori*-positioner, der er nødvendige.

Problemerne med denne metode er først og fremmest at fastslå *a priori*-positionerne på en troværdig måde. Ekspertundersøgelser kan anvendes til at bestemme partiernes position i tidspunkt ( $t_0$ ), men kan hverken anvendes til at fastslå positionerne på tidspunkt ( $t-1$ ) eller for den sags skyld ( $t-25$ ), ej heller til at

fastslå positioner, der ligger ude i fremtiden (t10). Et fejlagtigt valg af referencetekster og *a priori*-positioner vil gøre metoden lige så sårbar over for kritikken om cirkelslutninger, som også fremføres mod eksempel analyser af afstemningsdata. Problemerne til trods, så er metoden et interessant alternativ til at få gennemanalyseret valgprogrammer fra en ny metodisk vinkel.

### *Datamæssige alternativer*

Ud over at anvende andre metoder til at kode valgprogrammerne kunne det overvejes, om der findes andre datakilder, der kunne anvendes til at placere de politiske partier i forhold til hinanden. To datamæssige alternativer, der kan antages at være dækkende for hele partiet, men som af andre årsager er mindre gode, er henholdsvis analyser af lovforslag (se for eksempel Pedersen, 1967; Pedersen, Damgaard og Olsen, 1971; Skjæveland, 2003) og samlinger af partiudtalelser, såsom *Håndbog i Dansk Politik* (Garodkin, 2001; jf. Kurrild-Klitgaard, 1997; Hansen, 2004). Analyser af partiernes afstemningsadfærd i parlamenter er dominerende i studier af den amerikanske kongres, men anvendes kun i mindre grad på europæiske parlamenter. Forskellen på CMP-dataene og afstemningsdata er, at hvor førstnævnte er inputdata, er sidstnævnte outputdata. Det vil sige, at det er tvivlsomt, om man kan anvende afstemningsdata som uafhængig variabel i analyser, da der er høj risiko for at foretage cirkelslutninger.<sup>8</sup> Dette problem er ikke lige så stort hos eksempelvis samlinger af partiudtalelser, som de kendes fra *Håndbog i Dansk Politik*. Derimod lider denne type kilde af samme problem som CMP-dataene, at den er rodfæstet i et fast antal kategorier, der giver en begrænsning på de analyser, der kan foretages. Et sidste problem er, at denne type kilde ikke findes for alle lande, og at den for den danske case kun eksisterer fra 1979 og frem. Hverken afstemningsdata eller partiudtalelser er derfor relevante alternativer til CMP-dataene, i hvert fald så længe at data skulle kunne anvendes som uafhængig variabel.

### **Sammenfatning**

Kritikken af de danske CMP-data har to hovedpunkter: Først og fremmest er der problemet med de manglende 30 pct. af de *quasi-sentences*, der er elementerne i analysen. Sekundært er det de valgprogrammer, der er anvendt til kodningen. Der er næppe nogen tvivl om, at partiplaceringer i ovenstående tabel ikke kan accepteres ukritisk. Derfor må der kigges nærmere på de elementer, som er med til at give, hvad der med god ret kan betegnes som problematiske resultater. Kritikken af CMP-dataene for Danmark er mangfoldig, en kritik der ikke tidligere har været systematisk fremført. CMP-dataene for Danmark lider af mange mangler, som diskuteret ovenfor. Dataene har kun været anvendt i mindre omfang (e.g. Holmstedt og Schou, 1987 og til dels Schou og Hearl, 1992).

Men hvorfor anvendes data så overhovedet? Fordi CMP-dataene er det eneste tilgængelige sæt af data, der analyserer valgprogrammer, hvor dårligt udvalgt disse så end måtte være. Men grænsen for, hvornår dårlige data er bedre end ingen data, er tæt på at være nået. Man kan kun håbe, at det i løbet af de kommende år vil være muligt at få foretaget en gennemgribende genindsamling af

danske partiprogrammer, valgmanifeste og lignende, at foretage en kvalitativ bedømmelse af disse, og endelig at foretage en valid, replicerbar kodning. Det er en simpel nødvendighed, hvis Danmark og de danske partier skal være repræsenteret i CMP-datasættet med troværdige data. CMP-datasættets udbredelse i den politologiske verden gør, at troværdige danske data skal være inkluderet i det datasæt, hvis vi også i fremtiden vil se komparative analyser, der anvender CMP-dataene, og hvor Danmark er inkluderet.

## Noter

1. Jeg skylder stor tak for kommentarer på denne og tidligere versioner til Robert Klemmensen, Christoffer Green-Pedersen og *Politicas* to anonyme bedømmere. En stor tak ligeledes til Mogens N. Pedersen for råd, vejledning og dybdeborende kommentarer på en tidligere, i anden sammenhæng udarbejdet, og meget tentativ udgave af nærværende artikel.
2. Se appendiks i Budge et al. (2001) for, hvilke lande der har undtagelser for denne regel.
3. Dele af dataene er anvendt af andre, se for eksempel Green-Pedersen (2004) for anvendelse af data til at illustrere forskellen mellem Socialdemokratiet og Venstre.
4. Alle standardafvigelser er beregnet på baggrund af observationerne for de enkelte manifeste, som de fremgår i Volkens (2002).
5. Eksempler på valgprogrammerne er hentet fra Volkens (2002).
6. Eksemplerne er ligeledes hentet fra Volkens (2002).
7. Volkens (2001b) anerkender, at det ville være en fordel for hele forskningsprojektet, om det var muligt at anvende en udvidet kodebog.
8. Laver (2001) er vel nok den tydeligste eksponent for dette synspunkt

## Litteratur

- Benoit, Kenneth and Michael Laver (2003). "Estimating Irish Party Positions Using Computer Wordscoring: The 2002 Elections", *Irish Political Studies*, Vol. 18, No. 1, pp. 97-107.
- Budge, Ian (1987). "The internal analysis of election programmes", pp. 15-38 in Ian Budge, David Robertson and Derek Hearl (eds.), *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*, Cambridge: Cambridge University Press.
- Budge, Ian (2001). "Validating the Manifesto Research Group approach: theoretical assumptions and empirical confirmations", pp. 50-65 in Michael Laver (ed.), *Estimating the Policy Position of Political Actors*, London: Routledge.
- Budge, Ian, David Robertson and Derek Hearl (eds.) (1987). *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*, Cambridge: Cambridge University Press.
- Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum (2001). *Mapping Policy Preferences*, Oxford: Oxford University Press.
- Budge, Ian and Judith Bara (2001). "Introduction: Content Analysis and Political Texts", pp. 1-18 in Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum, *Mapping Policy Preferences*, Oxford: Oxford University Press.
- Gabel, Matthew J. and John D. Huber (2000). "Putting Parties in Their Place: Inferring Party Left-Right Ideological Positions from Party Manifesto Data", *American Journal of Political Science*, Vol. 44, No. 1, pp. 94-103.

- Garodkin, Ib (red.) (2001). *Håndbog i Dansk Politik 2002*, København: Rosinante.
- Garry, John (2001). "The computer coding of political texts: results from Britain, Germany, Ireland and Norway", pp. 183-192 in Michael Laver (ed.), *Estimating the Policy Position of Political Actors*, London: Routledge.
- Green-Pedersen, Christoffer (2004). "Center Parties, Party Competition, and the Implosion of Party Systems. A study of centripetal tendencies in multiparty systems", *Political Studies*, Vol. 52, No. 2, pp. 324-341.
- Hansen, Martin Ejnar (2004). "Økonomi eller Velfærd? Dimensionalitet i Folketinget", *Økonomi & Politik*, 77. årgang, nr. 4, pp. 71-82.
- Holmstedt, Margareta and Tove-Lise Schou (1987). "Sweden and Denmark 1945-1982: election programmes in the Scandinavian setting", pp. 177-205 in Ian Budge, David Robertson and Derek Hearl (eds.), *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*, Cambridge: Cambridge University Press.
- Kim, Heemin and Richard C. Fording (1998). "Voter Ideology in Western Democracies 1946-1989", *European Journal of Political Research*, Vol. 33, No. 1, pp. 73-97.
- Kurrild-Klitgaard, Peter (1997). "Liberalismens Udfordring", pp. 150-173 i Jens Løgstrup Madsen, Christopher Arzrouni, Torben Grage, Peter Kurrild-Klitgaard og Finn Ziegler (red.), *Den moderne liberalisme: Rødder og perspektiver*, København: Bredtåblik.
- Laver, Michael (2001). "How should we estimate the policy positions of political actors", pp. 239-244 in Michael Laver (ed.), *Estimating the Policy Position of Political Actors*, London: Routledge.
- Laver, Michael and Kenneth Benoit (2002). "Locating TDs in Policy Spaces: Word-scoring Dáil Speeches", *Irish Political Studies*, Vol. 17, No. 1, pp. 59-73.
- Laver, Michael J. and Ian Budge (1992). "Measuring Policy Distances and Modelling Coalition Formation", pp. 15-40 in Michael J. Laver and Ian Budge, *Party Policy and Government Coalitions*, London: Macmillan.
- Laver, Michael and John Garry (2000). "Estimating policy positions from political texts", *American Journal of Political Science*, Vol. 44, No. 3, pp. 619-634.
- Laver, Michael, Kenneth Benoit and John Garry (2003). "Extracting Policy Positions from Political Texts Using Words as Data", *American Political Science Review*, Vol. 97, No. 2, pp. 311-331.
- Lijphart, Arend (1984). "Measures of Cabinet Durability – A Conceptual and Empirical Evaluation", *Comparative Political Studies*, Vol. 17, No. 2, pp. 265-279.
- Lijphart, Arend (1999). *Patterns of Democracy*, New Haven: Yale University Press
- Mair, Peter (2001). "Searching for the positions of political actors: a review of approaches and a critical evaluation of expert surveys", pp. 10-30 in Michael Laver (ed.), *Estimating the Policy Position of Political Actors*, London: Routledge.
- McLean, Iain (2004). "The Dimensionality of Party Ideologies", *Nuffield College Working Paper 2004-W8*, University of Oxford.
- Pedersen, Mogens N. (1967). "Consensus and Conflict in the Danish Folketing 1945-1965", *Scandinavian Political Studies*, Vol. 2, pp. 143-166.
- Pedersen, Mogens N., Erik Damgaard and Peter Nannestad Olsen (1971). "Party Distances in the Danish Folketing", *Scandinavian Political Studies*, Vol. 6, pp. 87-106.
- Pelizzo, Riccardo (2003). "Party Positions or Party Direction? An Analysis of Party Manifesto Data", *West European Politics*, Vol. 26, No. 2, pp. 67-89.
- Robertson, David (1976). *A Theory of Party Competition*, New York: John Wiley & Sons.
- Saglie, Jo (1998). "A Struggle for the Agenda", *Party Politics*, Vol. 4, No. 3, pp. 347-366.
- Schou, Tove-Lise and Derek J. Hearl (1992). "Party and Coalition Policy in Denmark", pp. 151-188 in Michael J. Laver and Ian Budge (eds.), *Party Policy and Government Coalitions*, Basingstoke: St. Martin's Press.

- Skjæveland, Asbjørn (2003). *Government Formation in Denmark 1953-1998*, Århus: Politica.
- Stokes, Donald E. (1963). "Spatial Models of Party Competition", *American Political Science Review*, Vol. 57, No. 2, pp. 368-377.
- Volkens, Andrea (2001a). "Manifesto research since 1979: from reliability to validity", pp. 33-49 in Michael Laver (ed.), *Estimating the Policy Position of Political Actors*, London: Routledge.
- Volkens, Andrea (2001b). "Quantifying the election programmes: Coding procedures and controls", pp. 93-110 in Ian Budge, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara and Eric Tanenbaum, *Mapping Policy Preferences*, Oxford: Oxford University Press.
- Volkens, Andrea (2002). Dataset CMP2002, Comparative Manifestos Project, Science Center Berlin, Research Unit Institutions and Social Change (Director H.-D. Klingemann) in cooperation with the Manifesto Research Group (Chairman I. Budge).