Introducing the Manifesto Data: Background

TEXTS AS DATA

Political texts are at once the most widely available source of evidence about politics and the most neglected in terms of quantitative analysis. Practically everything that gets done politically is recorded in protocols or minutes, issued as directives or laws, or reported as proceedings of committees or legislative bodies. Political causes are advocated in recorded speeches and interviews, pamphlets, posters, and, of course, party platforms and manifestos. The latter are unique in being the only authoritative party policy statement approved by an official convention or congress. Possibly because of this they stand alone in being full 'five year plans' for the development of society.

Political historians and commentators have, of course, always used texts as their major source. But they have done so selectively—giving rise to controversy about bias and unrepresentativeness in the evidence. This extends to the reading and reporting of the texts themselves. Quantitative analyses in the form of counts of key words, sentences, or expressions can be carried out according to specified procedures over the whole range of relevant documents. This provides a superior basis for tackling problems of selection, reliability and validity, or at least specifying them clearly so they can be taken into account in evaluating results.

This book, like its predecessor, reports procedures and results for one series of political documents, the programmes parties issue for elections. All the (quasi-)[1] sentences in each document are classified into a set of policy-related categories according to an established set of rules and instructions (specified in Appendix II). Each decision is recorded on a copy of the original document so that it can be reviewed and corrected if necessary. The distribution of sentences over the categories is what distinguishes one document from another. Shifts in the distribution mark policy changes over time and highlight differences in one party's document as compared with others. Parties can be compared in this way with each other, at any one election: in terms of their 'tracks' from one election to another; and either within countries or across countries.

These dynamic estimates contrast with the other main measure commonly employed to estimate party policy positions—'expert judgements' based on survey responses of political scientists in each country to questions set by the investigators about where the national parties stand. Being essentially qualitative judgements couched in quantitative form (e.g. placement of parties on a ten-point scale) they share many of the problems of historical judgement—primarily the question of what evidence the party placements are based on, and whether this is the same as that used by the other experts involved in the survey (Budge 2000). Such judgements give a little more precision to traditional 'party family' classifications

(McDonald and Kim 2007). But being based on long-standing ideological positions they do not relate to any clearly specified time point such as a particular election, and hence lack any dynamic quality (McDonald and Mendés 1999).

Documentary estimates do possess these qualities. However, their major disadvantage is the immense amount of time and resources necessary for hand-coding large numbers of documents. The estimates reported here are based on the labours of more than fifty coders over twenty-five years, not to mention the efforts of supervisors and other members of the team, with financing from the WZB. Over the lifetime of the entire MRG/CMP programme (1979–2003 and continuing) perhaps a million and a half euros have been invested in data gathering and preparation.

This makes clear what a precious resource is being opened to readers of this and the previous Oxford University Press volume! The expense and effort involved also explain why ours is the only comparative text-based data-set of its kind and why short cuts like expert judgements are often taken. The costs are prohibitive and it is only the dedication of the original MRG and the institutional basis provided by the Wissenschaftzentrum which explain the data-set's existence.

However, for the reasons given above many such text-based data-sets are essential to the development of systematic political science. As a result efforts have increasingly been devoted to computerizing content analysis. Texts can be read in as raw data and words processed into numbers using one of several programmes now available (see Chapter 6). This may well be the way content analysis will go in the future. In Chapter 6, we consider the progress that has been made and how far computerization incorporates the basic logic of the approach we have employed in hand-coding. Before we embark on more detailed discussion, however, we should summarize the development of the manifesto data as they now stand, thus concretely illustrating the strengths and weaknesses of the estimates we present in this book.

THE MANIFESTO RESEARCH GROUP AND COMPARATIVE MANIFESTO PROJECT 1979–2004

The MRG was formed in 1979 by Ian Budge and David Robertson, both at that time in the Department of Government, University of Essex. It was constituted formally as a Research Group of the European Consortium for Political Research (ECPR) which hosted many of its early meetings and obtained funding from the Volkwagenstiftung and Tercentenary Fund of the Bank of Sweden. This supported most of its work in the 1980s.

Operationally the MRG consisted of a number of political scientists who were interested in two questions: (*a*) what political issues divided post-war political parties and (*b*) were they converging or diverging in ideological or policy terms? The last question had practical implications for interpretations of post-war politics. But it also had theoretical implications for 'rational choice' interpretations of

party strategic behaviour. Downs' influential interpretation of two party competition (Downs 1957: 112–19) would lead to expectations of party convergence on the position of the median voter. The manifesto data allow us to see in practice whether this occurred.

The MRG grouped country specialists with the requisite language ability and substantive knowledge, whose interest was primarily in the politics of 'their' country, and comparativists who often had wider theoretical interests. The object was to produce a book which would analyse party divisions within separate chapters for each country—however within a common framework which permitted some comparative analysis and conclusions at the end (Budge, Robertson, and Hearl (eds.) 1987).

Robertson had already carried through a pioneering analysis of British party manifestos 1922–74 (Robertson 1976). The coding scheme based on reading these documents counted their sentences into twenty-two general policy areas. What the count of sentences over these categories recorded was the *relative emphasis* parties placed on them, not *positive* or *negative* references made to them. Political opposition was expressed by emphasizing another issue—peace, say, as opposed to military strength, freedom as opposed to planning, and so on.

It is important to emphasize two points about this coding, which formed the basis for the final MRG scheme applied to the documents covered here (Appendix II). One is that it derived from the documents themselves: relative emphases were the way in which British parties expressed themselves. Second, confrontation between the parties is there, but it takes the indirect form of emphasizing another issue rather than direct negative comment on a rival position on each issue. Budge and Farlie (1977, and especially 1983) who extended Robertson's coding to the USA (1922–76), developed a 'saliency theory' of party competition from the success of relative emphases as a coding scheme (see below and also Stokes 1966; Riker 1993; Budge et al. 2001: 78–85). However, one could use relative emphases to count words or sentences without necessarily subscribing to saliency theory.

Indeed some members of the original MRG found it hard to believe that relative emphases *were* the main way parties differentiated themselves from each other, in spite of the British and American evidence. Hence, the compromise coding scheme drawn up in 1979 created a number of pro and con categories in issue areas where direct confrontation between opposing policies seemed most likely to occur. This scheme is summarized in Table I.1 (see Appendix II for a detailed presentation).

When results from the actual coding came in about a year later it emerged that:

(*a*) national parties generally emphasized either pro or con positions but not both in any one issue area;

(*b*) the undifferentiated categories (i.e. those deriving from or developed in sympathy with the earlier relative emphases coding) performed well in differentiating parties. As a result MRG coding practices developed from then on the basis of a relative emphases approach.

Table I.1 The standard MRG/CMP coding of election manifestos 1945–98

Domain 1: External Relations	
101	Foreign Special Relationships: Positive
102	Foreign Special Relationships: Negative
103	Anti-Imperialism: Negative
104	Military: Positive
105	Military: Negative
106	Peace: Positive
107	Internationalism: Positive
108	European Community: Positive
109	Internationalism: Negative
110	European Community: Negative
Domain 2: Freedom and Democracy	
201	Freedom and Human Rights: Positive
202	Democracy: Positive
203	Constitutionalism: Positive
204	Constitutionalism: Negative
Domain 3: Political System	
301	Decentralization: Positive
302	Centralization: Positive
303	Governmental and Administrative Efficiency: Positive
304	Political Corruption
305	Political Authority: Positive
Domain 4: Economy	
401	Free Enterprise: Positive
402	Incentives: Positive
403	Market Regulation: Positive
404	Economic Planning: Positive
405	Corporatism: Positive
406	Protectionism: Positive
407	Protectionism: Negative
408	Economic Goals: Positive
409	Keynesian Demand Management: Positive
410	Productivity: Positive
411	Technology and Infrastructure: Positive
412	Controlled Economy: Positive
413	Nationalization: Positive
414	Economic Orthodoxy: Positive
415	Marxist Analysis: Positive
416	Anti-Growth Economy: Positive
Domain 5: Welfare and Quality of Life	
501	Environmental Protection: Positive
502	Culture: Positive
503	Social Justice: Positive
504	Welfare State Expansion
505	Welfare State Limitation
506	Education Expansion
507	Education Limitation
Domain 6: Fabric of Society	
601	National Way of Life: Positive
602	National Way of Life: Negative
603	Traditional Morality: Positive

604 Traditional Morality: Negative
605 Law and Order: Positive
606 Social Harmony: Positive
607 Multiculturalism: Positive
608 Multiculturalism: Negative

Domain 7: Social Groups
701 Labour Groups: Positive
702 Labour Groups: Negative
703 Agriculture: Positive
704 Middle Class and Professional Groups: Positive
705 Minority Groups: Positive
706 Non-Economic Demographic Groups: Positive

An unanticipated consequence of imposing these initial checks was, however, that a number of explicitly pro and con categories were built into the coding scheme. Production coding of manifestos from 1945 to 1982 had already incorporated them, and future coding had to include them to maintain comparability. The 'minority' category on each set of paired pro and con positions did not generally attract a high number of references. Hence, it was still possible to regard the coding scheme as essentially based on relative emphases. But it was not an entirely pure reflection of them.

This ambiguity carries several implications for users of the data:

(*a*) it makes the estimates more flexible, in that those who wish to devise paired confrontations from the data can do so (Table 6.1). Such confrontations include both the original pro and con codings and 'relative emphasis' categories that can be taken as opposed (e.g. 'freedom' vs. 'planning'). As we point out, it is not clear whether the different types of pairing reflect the same kind of opposition, however—a point developed below in relation to Table 6.1.

(*b*) inclusion of pro and con categories in addition to relative emphases, which might also be taken as indicating support or opposition in that specific instance, contributed to 'noise' and coding error. It was unclear whether a reference such as 'beating swords into ploughshares' should be coded simply as 'peace' or as 'military negative'. In contrasting the explicitly pro and con positions towards an issue, it is therefore unclear whether these account for all references to the topic in the manifesto or whether there has been some 'seepage' of references out to another issue area.

One reaction to these possibilities is to aggregate categories so that those which may have been somewhat confused with each other in coding are grouped on the same side of a scale thus improving overall validity and reliability. The prime example of this is the Left-Right continuum discussed below (Table 1.1).

Another consequence of early group discussions, with far-reaching consequences for coding, was the anxiety of country specialists about providing a sufficiently subtle coverage of 'their' national politics. This resulted in a multiplication of the number of categories from Robertson's original 22 (1976) to a final 56. Having more detail may be a strength. However, it turned out that many of the new categories attracted few references (between a third and a half of specific

scores in the data-sets are zero). They also increased ambiguity about where to locate sentences. Less serious was the allowance made for country specialists to create subcategories under general categories of the coding scheme which could be aggregated back to the general category (Budge, Robertson, and Hearl 1987: 466–7; Budge et al. 2001: 98). Most of these, contrary to the original expectations of the researchers, were also thinly populated and so have been mostly used in aggregate form. However, they are always available for those with particular interests to pick out and use and may be of special interest to CEE specialists seeking to detail events after 1990 in specific countries (see Appendix II).

These early MRG discussions feed directly into the uses that can be made of the codings and the interpretations put upon them. New users need to be alerted to this. In Chapter 6, we take up some of these points and indicate in detail how on the one hand they permit extended applications of the data and on the other impose cautions and constraints.

With these crucial decisions made, the MRG forged ahead with production coding from 1980 to 1983, eventually covering the nineteen countries whose analysis is reported in *Ideology, Strategy and Party Change* (Budge, Robertson, and Hearl (eds.) 1987). This helped to develop many of the techniques reported above such as aggregating individual categories through factor analysis to describe the policy dimensions within which parties differentiated themselves. With regard to the two original questions which had motivated the group (*a*) the main policy dimension parties emphasized, quite against MRG expectations, was the Left-Right one (see Table 1.1 for the issues which define this), (*b*) parties did converge in Left-Right terms over the thirty-year period examined. But parties moved so much and judgements were so conditioned by the first and last years chosen for comparison that the more considered conclusion was that there was no irrevocable tendency to either convergence or divergence but simply fluctuation within a defined and consistent party policy space. (The fluctuations were probably influenced by strategic considerations (cf. Adams 2001).)

By the time the original book was published it had become clear that the MRG policy estimates were capable of adaptation and extension to many other research purposes important for democracy. From 1983, they were applied to coalition formation and policymaking (Laver and Budge (eds.) 1992), leading to the conclusion that policy proximity did not really work as an explanation of why parties joined coalitions.

By this time the MRG was winding down. Many original members left having answered their initial research questions. The data and document collections clearly had a potential for general research in and beyond political science (cf. citations in the Bibliography of Manifesto-based Research). To preserve them, the collection of original documents was partly published in microfiche (Hearl 1990) and deposited at the German Social Science Archive at Cologne. Updated and periodically extended, these collections are available for interested scholars through the archive (za@za.uni-koeln.de). The data up to 1988 was also deposited at the British Social Science Data Archive at Essex. The full data-set up to 2003 is now available at both archives.

Collection and coding of documents in most of the original countries and eventually in all OECD members was passed over to the CMP based at the WZB in 1989 and directed by Andrea Volkens, using a worldwide network of coders. At the same time a separate initiative of Hans-Dieter Klingemann and the WZB, also directed by Volkens, covered collection and coding of party programmes in all the emerging democracies of CEE, as far as the Caucasus. This is the basis for the major time series reported in this book and CD-ROM for CEE countries 1990–2003 (the corresponding OECD estimates being also reported for comparisons and checks). Needless to say the CEE project, though formally distinct, was carried through using the same procedures and coding as the OECD ones. So, they are totally comparable.

With consolidation and continuance of the time series assured, a reduced MRG carried on substantive research which supplemented the manifesto material with public policy indicators such as expenditures (Klingemann et al. 1994)—conclusion, that party emphases and expenditure change generally track each other across ten democracies, as required by mandate theory. A further investigation of the operation of the median mandate (McDonald and Budge 2005) has been published by Oxford University Press. A taster of how the manifesto median estimates enable us to investigate democratic relationships is provided in Chapter 7. Other research has focused on patterns of party interaction revealed by the estimates, within a post-Downsian perspective (Downs 1957; Budge 1994; Adams 2001). A number of further research developments, which use the manifesto estimates in different ways, can be found in Bara and Weale (2006).

The last substantive publication of the MRG itself was of course the predecessor to this volume (Budge et al. 2001). It presented the full time series for OECD countries and Israel up to 1998 on a CD-ROM available for immediate use. The text illustrated uses to which the data could be put, extended them to government and (median) electoral preferences, described coding and computational procedures, and applied tests of reliability and validity. It argued for macrolevel assessment of the estimates rather than mechanical inter-coder reliability checks, given the constantly expanding nature of the data and the different languages involved. However, a generalized measure of error for Left-Right estimates was presented (Heise 1969; Budge et al. 2001: 139). This is extended here (Chapters 4 and 5).

Previous methodological assessments are equally applicable to the data we report here, so we shall not repeat them but simply consider relevant points. Interested readers should go back to the first volume for a full account. However, some aspects of the methodology have emerged as requiring more detailed consideration and we go into them in Chapters 2–6. One topic we raise is the extent to which party election programmes (manifestos, platforms) should be considered as statements of underlying party ideology or simply of current policy. These are clearly not the same thing though they may influence each other as we argue below.

We realize, of course, that most readers will be interested primarily in what our estimates tell them about real politics, particularly in CEE over the past two

decades. While methodological discussions inform us about the basis on which estimates were made they are no substitute for the estimates themselves—which we now report in Chapter 1 as 'maps' of party movement along the Left-Right dimension.

NOTE

1. A 'quasi-sentence' is defined as an argument or phrase which is the verbal expression of one idea or meaning. It is often marked off in a text by commas or (semi-)colons. Long sentences may contain more than one argument so that they need to be broken up into quasi-sentences.

This page intentionally left blank