Political parties and the democratic mandate : comparing collective mandate fulfilment in the United Kingdom and the Netherlands

Louwerse, T.P.

Chapter 4

Comparing electoral and parliamentary competition

This study looks at the congruence of the spaces and structures of the electoral and parliamentary competition. This is essentially an institutional-level approach to the question of the party mandate: does the way parties speak to each other during the election campaign predict how they speak to each other in parliament? It is quite common to hear people talk about 'the position' of a party or that it has 'moved' to the left or right. Nevertheless, there is a lot of discussion about how one can estimate these issue positions. These cannot be directly observed, because they do not physically exist nor are they part of an uncontested social reality. Different methods of estimating party positions may lead to different and equally plausible outcomes. Therefore, the choice for a specific method must be made in light of the requirements of this study and the availability of data for the selected cases.

4.1 Research design

Ideally, a study of the party mandate would encompass a broad range of (developed) liberal democracies. Given the spatial method employed here and the need for original data collection this is not feasible. Comparing party positions as laid down in manifestos with what party MPs say in parliament requires a lot of text to be analysed. Even with computerized methods of content analysis, data collection and analysis of manifestos and parliamentary speech takes resources. Therefore, the number of countries in this study has to be limited.

I have selected two 'typical' cases: one majoritarian and one consensus democracy, that are nonetheless quite similar in other respects (age of the political system and the party system, economic development). This increases the chance

Table 4.1: *Comparison of the United Kingdom and The Netherlands on the main independent variable and controls*

	Regime type[a]	**Age of political system**[b]	**Average GDP per capita**[c]	**EU Member since**
United Kingdom	-1.21	1832/1928	25,767	1973
The Netherlands	1.23	1848/1919	20,548	1950

[a] Position on executive-parties dimension (Lijphart, 1999).
[b] First year: introduction of parliamentary government with suffrage based on property rights. Second year: introduction of general suffrage with equal rights for men and women.
[c] Average Gross Domestic Product per capita 1980-2008 (Purchasing Power Parity) in US$ 2005 prices. Source: World Bank (2010)

that variation on the dependent variable (mandate fulfilment) is indeed the result of the main independent variable (type of democracy). For example, differences in the levels of mandate fulfilment between two countries cannot be explained by different levels of economic development if both countries' Gross Domestic Products (GDP) are (roughly) the same. Although the current design is not a full-fledged most similar systems design in which all possible rival explanations are excluded (Lijphart, 1971, 1975), the design makes it at least plausible that a difference between two countries in terms of mandate fulfilment is a result of regime type.

The main explanatory variable is the regime type, e.g. the position of a country on Lijphart's (1999) executive-parties dimension. The two countries selected for this study are The Netherlands and The United Kingdom. These are very different on the variable regime type (fourth highest and fifth lowest on the executive-parties dimension respectively), but similar in other respects. It could be argued that in systems with little democratic tradition or systems that are unstable, mandate fulfilment would be lower. Examples of clear violations of the party mandate have been documented in Latin America (Przeworski et al., 1999). The political systems of the two selected countries have been stable over the last two centuries and the development of a fully representative system of government has taken a similar pace (see table 4.1). A second alternative explanation that was excluded due to the case selection is the level of income per capita. The Gross Domestic Product (GDP) per capita is roughly similar in both countries. Third, membership of the European Union (EU) is also included in table 4.1, because the presence of a supra-national decision-making body could have influence on the way which parties are able to fulfil their electoral mandates. Although Britain has entered the European community later, both have

been members for the largest part of the period covered by this study. Therefore, if there is a difference between majoritarian and consensus democracies in terms of mandate fulfilment, one is likely to observe it in the comparison of the United Kingdom and the Netherlands. Conversely, if I find a difference between the United Kingdom and the Netherlands in terms of mandate fulfilment, some rival explanations can be excluded and it is likely to be the result of my main explanatory variable.

4.1.1 Case selection

The difficulty of a small-n study is the relatively large influence of measurement error and a-typical cases. If only one case (an election and subsequent parliament) would be used for each country, errors in the measurement of party positions may lead to false conclusions. Furthermore, one may assume that there is some variation in the levels of mandate fulfilment between cases. If one would study only one case per country, the probability of selecting an outlier is rather large (King et al., 1994). To increase the number of cases, I selected elections and subsequent parliaments from six decades. An additional advantage of this approach is the possibility to track changes over time. To increase the number of observations in this two-country study, six cases are studied in each country. The number of observations can be further increased by treating each party's position on each policy dimension as a separate observation for some of the analyses.

For the selection of cases, five criteria were used. First, equal spread of cases over the period of interest (1950-2006). The period between 1950 and 2010 was one of democratic stability in both countries. Furthermore, both countries have witnessed gradual but remarkable changes in the way the political system functions. In both countries the relationship between voters and parties has changed, leading to changes in parties and party systems. With six cases per country, this means roughly one case per decade. Second, an equal distribution of party participation in government. Hence, each party's share of government participation must be similar in the sample as during the whole 1950-2006 period. Third, governments lasting less than three years were excluded, because very short parliaments would produce less reliable estimates of party positions. Fourth, the selected election years preferably must be similar for both countries, to correct for the influence of important international events. Fifth, to avoid that observations from one case are dependent on the observations in another case, it is preferred that no subsequent governments are included[1].

Table 4.2 displays all governments that sat for at least three years, ordered roughly per decade. Apart from the Netherlands in the 1950s and both countries in the 2000s, in each decade there is choice between two or three cases. By selecting the second option for Britain (except in the 2000s) in each decade and the

[1]Of course, the selected cases in one country will never be fully independent, because it concerns the same parties during different time periods.

Table 4.2: *Case selection*[a]

	United Kingdom		**Netherlands**	
	Years	Party[b]	Years	Parties[b]
1950s	1951-1955	Con	**1952-1956**	***PvdA*/KVP/ARP/CHU**
	1955-1959	**Con**		
1960s	1959-1964	Con	**1959-1963**	***KVP*/ARP/CHU/VVD**
	1966-1970	**Lab**	1967-1971	*KVP*/ARP/CHU/VVD
1970s	1970-1974	Con	**1972-1977**	***PvdA*/KVP/ARP/D66/PPR**
	1974-1979	**Lab**	1977-1981	*CDA*/VVD
1980s	1979-1983	Con	**1982-1986**	***CDA*/VVD**
	1983-1987	**Con**	1989-1994	*CDA*/PvdA
1990s	1987-1992	Con	**1994-1998**	***PvdA*/VVD/D66**
	1992-1997	**Con**	1998-2002	*PvdA*/VVD/D66
	1997-2001	Lab		
2000s	**2001-2005**	**Lab**	**2003-2006**	***CDA*/VVD/D66**

[a] Note: The party of the Dutch prime minister is italicized. The selected parliaments are printed in bold.
[b] For full party names, please refer to table B.2 on page 234.

Table 4.3: *Average government participation in the Netherlands 1950-2006*[a]

	Party[c]					
	CDA[b]	**PvdA**	**VVD**	**PPR**	**D66**	**LPF**
Period 1950-2006[a]	85.7	49.6	65.2	8.3	30.2	1.5
Selected cases	86.0	55.6	65.0	20.2	50.7	0.0
Difference (overrepresentation in sample)	+0.3	+6.0	-0.1	+11.9	+20.5	-1.5

[a] The figure indicates the number of days that the party participated in government as a percentage of the total number of days. The 'whole periode' figure includes all governments, not only those that sat 3 years or longer.
[b] Before 1980: Katholieke Volkspartij (KVP), Anti-Revolutionaire Partij (ARP) and Christelijk-Historische Unie (CHU)
[c] For full party names, please refer to table B.2 on page 234.

first option for the Netherlands, the criteria for case selection are fulfilled to a large degree[2]. The selection includes three Labour and three Conservative governments in the United Kingdom. For the Netherlands, table 4.3 shows that each party's participation in government in the sample is roughly equal to the participation in the whole 1950-2006 period. There is a difference between sample and population for two smaller parties, PPR and D66. For the PPR this cannot be avoided, for the only cabinet in which it ever participated is included in the analysis. D66 did not exist before 1966, therefore it could not have participated beforehand. In the 1967-2006 period, D66 participated in 42.5% of the governments, hence its overrepresentation is not dramatic. Apart from the 1950s and the 1960s, the selected parliaments sat in roughly the same years in the Netherlands and the United Kingdom.

In principle, all parties that win seats in an election are included in the analysis. There are two types of exceptions to this rule. In Britain, the analysis is limited to the three large parties, because the other parties are regional parties. The parties in Northern Ireland form a fully separate party system from the rest of the country. The regional parties in Britain, the Scottish National Party and Plaid Cymru do not compete in the whole of Britain either. For the Netherlands, I excluded some parties of which the electoral position cannot be estimated (i.e. there is no party manifesto).

[2]The 2005 and 2007 governments of the United Kingdom and the Netherlands respectively were not included, because their terms had not finished when data collection started.

4.2 Analysing parties' policy preferences

Party position is an abstract and, to a certain degree, a subjective notion. Where parties stand is conditional on the method used for studying their position. However, this does not mean that parties might very well be 'everywhere' and cannot be positioned at all. Very often different methods arrive at similar conclusions: hardly anybody will argue that the British Labour party was to the right of the Conservative parties during the 1980s, no matter which estimation procedure they use. When people talk about 'measuring' or 'estimating' party positions they do not suggest that they are measuring something which is simply out there, as if one were measuring the positions of buildings in a city or pieces on a check board. The party position compares a parties' policies or ideas with the policies or ideas of other parties, often expressed on a continuum, such as left-right or progressive-conservative.

Spaces of party competition are also abstract and constructed. The only concrete 'spaces of competition' are the buildings of parliament or maybe television studios and back rooms during campaign time. These are not the spaces of competition that I am interested in, although the most well-known dimension of party competition, Left-Right, is derived from exactly such a (parliamentary) setting (Benoit and Laver, 2006). Spaces of competition, in the abstract sense, are (visual) representations of the policy competition between political parties. The abstract nature of the space of competition makes it more difficult to estimate what it looks like. The choices in the estimation procedure are likely to affect the findings and should thus be made carefully.

4.2.1 Estimating party policy preferences

Estimating policy positions has been a major challenge for political scientists. Researchers have developed several methods of estimating party policy positions: qualitative evaluation, mass surveys, elite surveys and interviews, expert surveys, analysis of (roll call) votes and estimating positions from political texts, such as election manifestos and parliamentary speeches. Most approaches have resulted in relatively reliable and valid estimates for party positions. However, not all approaches can be used in every circumstance. For example, the analysis of parliamentary (roll call) votes can only be done if voting behaviour has been recorded.

This study seeks to compare party policy preferences during election campaigns and in parliament, both party issue saliency and party issue positions (what parties talk about and what they say). The method used to estimate these preferences must thus be very well able to distinguish between the party preferences during the election and those in parliament. This is even more of a concern, because the election and subsequent parliament occur very shortly after one another. Methods that give an estimation of a party position in, for example, the first half of the 1990s, are not specific enough to compare the preferences of Brit-

ish parties in the 1992 elections and the subsequent parliament. Another vital requirement of this study is that it should be possible to estimate party preferences from the past as this study covers the period from 1950 to 2006. The method should thus be able to arrive at an estimate long after the party preferences have been outlined.

Perhaps the least discussed, but often used, approach is a qualitative evaluation of parties' positions on issues by the researcher. This approach involves in-depth knowledge of a particular party system which allows the researchers to combine evidence from a wealth of sources to arrive at a careful ordering of parties. The need for in-depth knowledge is obviously also a major hindrance for such an approach, because it makes the study very resource-intensive. Furthermore, other researchers may arrive at rather different conclusions, so the reliability of this approach is questionable. The comparison between the party competition at election-time and the competition in parliament can benefit from the elaborateness of the analysis, but may suffer from a heavy reliance on a single (or small group of) researcher(s). This especially makes a comparison between cases more difficult.

The expert survey builds on the idea that political scientists generally have a fair idea of where parties are located on policy scales (Benoit and Laver, 2006; Huber and Inglehart, 1995; Laver, 1998; Laver and Garry, 2000; Warwick, 2005; Marks et al., 2006). However, instead of relying on a single researcher, they ask a number of experts, usually academics, to position parties on (pre-defined) policy scales. This leads to higher reliability of the analysis and provides an estimate of the uncertainty of the party positions found (Benoit and Laver, 2006). It is of course, pivotal, to find proper experts for such a purpose, especially since the studies rely on a relatively small number of respondents. A limitation of the method is that expert surveys cannot be reliably used to estimate positions in the past. Apart from that, it is often found that the expert estimates of party policy positions do hardly change over time. Instead of estimating a parties' position at a specific point in time, the expert surveys seem to capture a sort of longer-term party position (McDonald et al., 2007). This makes these surveys unsuitable for comparing party positions in the short term, e.g. before and after elections, unless you specifically ask experts to distinguish between the two. Even if one does that, the experts' estimates will probably result from a rather broad range of sources and may just reflect public opinion on parties' positions, rather than the behaviour of parties themselves (Marks et al., 2007).

The mass survey and the elite survey suffer from largely the same problems as the expert surveys in terms of the ability to collect data for the past and comparison over short periods of time. For politicians there is the additional question of the honesty of their answers (Benoit and Laver, 2006: 64). After all, politicians have an interest to show themselves to be rather congruent in terms of their policy positions between elections and manifestos. Although useful data sources for a large number of applications, expert, mass and elite surveys are not suitable for the current research design.

Roll call vote analysis has primarily been applied to the United States, but there is no reason why it cannot be applied in other countries as well (Poole, 2005). Parliamentary votes are obviously only a representation of the parliamentary competition, thus another source would be needed for the electoral competition. Although this may bias the comparison between the two spaces of competition, this should not be an insurmountable problem. Not all parliaments use roll call votes. In Great Britain many votes are indeed roll calls and the parliamentary record provides the vote of each individual MP. Roll call votes are, however, very rare in the Dutch parliament. The Second Chamber usually votes by show of hands, which is recorded by parliamentary party rather than by individual. Roll calls are only held when the expected majority is slim or when a party is internally divided, for example on ethical issues. Vote analysis of the Dutch parliament is thus practically restricted to the party level.

The lack of roll call votes is not necessarily a problem for the purposes of this study. After all, it looks at the party mandate, which means that party voting behaviour is just fine for my purposes. For Britain, one might aggregate individual votes to party votes, taking the majority of the party to represent the party line. The Dutch parliament usually votes by parliamentary party group. However, before the 1970s, roll call voting was the most important way in which the Dutch parliament voted, but this was used only sporadically. Instead, if only one party would oppose a proposal it would ask for an annotation of its opposition in the minutes. Even when considering these as parliamentary votes, the number of votes is basically too small for a proper analysis[3].

Apart from data availability considerations, the properties of the vote should also be considered. Parliamentary votes are the result of discussions in committees and the general assembly. Although they are in a sense, the final consequence of the parliamentary work, they only represent a small portion of that work. Much parliamentary work does, for example, not result in votes, for example parliamentary questions as well as many debates. In addition, the vote reflects compromises reached earlier in the parliamentary process. Many votes, especially in the Dutch parliament, concern motions and amendments, which are usually only tabled if the government does not provide a satisfactory answer to the questions or demands of a party.

The method that will be used in this study is content analysis of political text. Political texts are manifold, for example party manifestos, interviews, articles and speeches of members of parliament. Talking is part of the 'core business' of politicians. Language is the vehicle of the party mandate: the mandate is obtained through the election programme and it is fulfilled by pursuing this programme in parliament, by voting and discussing. Studying these political texts is thus one of the most direct ways in which party behaviour can be observed (Benoit and Laver, 2006).

[3]It is also very skewed, with opposition parties, especially the communist party, requesting many more annotations than any other party.

A crucial advantage of studying political text is that a clear distinction can be made between what is said during the elections, for example in the party manifesto, and what politicians say in parliament. In addition, it is possible to study parties' preferences in the past, as long as the textual sources are available. For example, there are no expert surveys or mass surveys of party policy preferences during the 1950s (at least not in both of my cases), but by studying parties' manifestos and parliamentary speeches, their preferences can be estimated.

It is sometimes argued that the weakness of manifestos and parliamentary debates is that they are strategic documents (Marks et al., 2007). For example, parties may choose to ignore some topics in their manifesto in order not to alienate voters. Furthermore, depending on the context parties operate in, they might formulate some preferences very ambivalently (Marks et al., 2007: 33). It is argued that the manifesto is therefore a poor reflection of parties' actual policy positions. In the discussion of the spatial approach to the party mandate in chapter 2, I have indicated that the strategic nature of manifestos should be taken into account in their analysis. What is absent from a manifesto is at least as important as what is included. That is exactly the reason why the spatial approach looks at party positions on issue dimensions and in political spaces rather than at individual pledges. If parties do indeed exclude important issues from their manifestos for strategic reasons, one would find incongruence between their manifesto and parliamentary positions. The strategic nature of these document is thus not a problem, in the light of the current research question, but something of interest. For this study, what matters is not what parties 'really want', but what parties really say.

4.2.2 Estimating party policy preferences from text

Party policy preferences are relevant explanations for a large number of political research questions. It is thus not surprising to find that many researchers have attempted to estimate parties' policy preferences. This has led to many different approaches. Traditionally, these methods have relied on the coding of texts by humans: people reading texts and classifying sentences or paragraphs or looking for the presence of certain policy proposals. Recently, computerized methods for estimating party policy preferences from political texts have been developed, which can greatly reduce the workload involved in estimating preferences from political texts.

Manual content analysis

There are basically three approaches of human coding of political texts: the saliency approach, the network approach and the confrontational approach. The saliency approach is based on the idea that parties do not compete by adopting confronting policy positions, but by emphasizing different issues. It has produced the very elaborate and successful Comparative Manifestos Project (CMP),

which has analysed party manifestos from many countries from 1945 onwards (Budge et al., 1987, 2001; Klingemann et al., 2006). The classification scheme reflects the assumptions of the saliency theory of party competition. This theory argues that parties do not compete by adopting conflicting views on a number of given issues, but rather by selectively emphasizing certain issues over others. Thus, instead of some parties proposing all kinds of environmentally-friendly policies and another party opposing these, 'green' parties will simply emphasize the importance of the environment while other parties will ignore the issue. Saliency theory goes even further, by arguing that on most issues there is basically only one viable policy position. For example, very few parties will argue that environmental pollution is a good thing and that climate change will further the well-being of the human race. Instead, parties that are not so much in favour of green policies will simply downplay the issue, saying that the status quo is not so bad and that the problems of climate change are exaggerated. The real competition between parties does not occur by a direct confrontation of parties, but by stressing some issues and downplaying others.

The Comparative Manifesto Project (CMP) uses a classification scheme that consists of 56 'saliency' categories. For each manifesto, the CMP dataset gives the percentage of sentences in that manifesto dedicated to one of these categories. A manifesto is coded by only one person, which means that there is no measure of inter-coder reliability. The authors of the project's books use the data to construct a Left-Right dimension of party competition[4] (Budge, 2001; Klingemann et al., 2006).

Despite the CMP's claim that differences between parties' positions are unimportant, the classification scheme consists of many 'confrontational' elements. For example, some categories present (implicitly) opposed ideas, such as 'Military: positive' and 'Military: negative'[5]. This means that the dataset could also be used in a more 'confrontational' way, if one is not willing to make the assumptions of saliency theory (Franzmann and Kaiser, 2006; Klingemann et al., 2006). However, one is limited by the categories that are included in the coding scheme (which is exactly the same for all documents in all countries across all years). An example of this limitation is that the coding scheme does not contain the issue of immigration. The application of the same coding scheme to all countries and years allows for comparative analyses of party positions, but also makes it

[4]Using a partly inductive, partly deductive method they designated 13 categories as left-wing 13 categories as right-wing. The Left-Right score is simply the percentage of manifesto (quasi-)sentences devoted to right-wing issues minus the percentage of manifesto (quasi-)sentences devoted to left-wing issues. Other authors propose different methods to calculate policy scales (Kim and Fording, 1998; Warwick, 2002; Franzmann and Kaiser, 2006; Lowe et al., 2011).

[5]In fact 20 out of 56 categories are twinned in such a way. Four other categories are also directly opposed, namely those that measure support for the 'expansion' or 'limitation' of the welfare state and education. Many other issues, especially in the economic domain, represent implicitly opposed views, such as 'Controlled Economy' and 'Free Enterprise' or 'Economic Orthodoxy' and 'Keynesian Demand Management'.

insensitive to particularities in countries and years[6].

The confrontational approach explicitly opposes the saliency theory of party competition. It maintains that party competition is centred around a number of important issues, on which parties have confronting opinions. It explicitly rejects the assumption of the saliency theory that there is basically only one viable policy position on an issue. Pellikaan et al. (2003) point out that electoral competition is not simply a case of parties 'talking past each other'. In each election, a number of important issues dominate the debate and parties often take very different positions on these issues. On minor issues, parties might indeed present proposals that no one else talks about, but the big election issues are confrontational issues.

For their analysis of the 2002 elections in the Netherlands, Pellikaan et al. (2003) selected the relevant issue dimensions based on the (unstructured) study of newspapers' campaign reporting and observers' opinions. They argue that in the Dutch election campaign of 2002 a new issue dimension had arisen, namely the 'cultural' dimension, which relates to issues on immigration and integration. This new dimension replaced religion as the second dimension of competition, next to the economic left-right dimension. For both of their issue dimensions, the researchers identify ten pivotal issues (items). They do not attempt to classify each sentence in a manifesto (like the saliency approach), but only look at support for or rejection of each of these items. For each party they simply record whether the party manifesto supports, rejects or does not mention the proposal. These scores are recoded, so that a positive score always means support for the free market, for example. The party position is simply the sum of all scores. For example, a party which supports 2 pro-market and 5 pro-state intervention proposals, receives a score of -3 (2-5)[7]. The confrontational approach only estimates parties' issue positions, it does not estimate parties' issue saliency. Obviously, the successful application of this method to a case requires a relatively great familiarity of the researcher with the situation in a country. On the other side, this makes it very easy to tailor the approach to many different countries and years: the researcher can select the relevant issue dimensions and items for each particular case.

The network approach is similar to the confrontational approach in the sense that it makes a clear distinction between party policy positions and issue saliency. On the other hand, it shares the saliency approach's method of coding each (substantive) sentence. In its original version, the network approach (also known as relational content analysis) models sentences as objects and the relations between those objects (Kleinnijenhuis and Rietberg, 1995; Kleinnijenhuis

[6]The comparative benefit of common coding scheme can also be questioned, because it is not necessarily a valid coding scheme for all countries and all years. The same can be said, probably even more strongly, about the Left-Right dimension that is constructed from this data: 'What does it really say when the Bulgarian Socialist Party in 1990 is estimated to be to the right of the UK Labour party in 2005?' (Louwerse, 2009: 108)

[7]Gemenis and Dinas (2010) propose a different calculation, in which neutral scores are ignored.

and Pennings, 2001). For example, in the sentence 'Brown wants higher taxes', Brown is the subject (actor) who has a positive attitude towards the object 'higher taxes'. Although such as method can be applied to all kinds of statements, not only policy statements, I will limit my discussion to these kinds of statements.

Kleinnijenhuis and Rietberg (1995) apply this methodology to news media; it can easily be modified to work with manifestos and parliamentary debates. Kriesi et al. (2006, 2008) apply the method for their analysis of party positions in western democracies. They use a simpler approach, leaving out the relational element of the analysis, simply coding newspaper statements that position parties one way or another. More specifically, they position each statement on one of 12 five-point issue scales, such as 'welfare', 'cultural liberalism', 'europe', 'immigration', 'army' and 'environment'. By explicitly positioning parties on issue dimensions, they make a clear distinction between the saliency of an issue for a party and its position on that issue. The issue saliency is simply the percentage of statements on one specific dimension, in other words how much parties talk about an issue. A party position reflects what message parties try to convey. The position is estimated by taking the average score of all statements on an issue. For example, a party that has 15 pro-welfare statements, 5 neutral statements and 3 anti-welfare statements, would have an average dimension score of 0.52 [8]. Thus, while the network approach by and large has the same theoretical ideas about party competition as the confrontational approach, its coding method bears similarity with both the confrontational and the saliency approach.

An obvious limitation of any manual content analysis is the amount of work involved. Hand-coding documents takes a lot of time, especially if one wants to code each document by more than one coder. Therefore, the resource intensity is very high (see table 4.4). This is particularly true for methods that rely on coding all sentences in a manifesto, namely the saliency and network approach. Because the confrontational approach limits itself to ten items per dimension, the burden of coding work is lower. This study does not only analyse party manifestos, but also parliamentary debates. This increases the coding load enormously. Therefore, the use of any form of manual coding for the present research question would require a sampling strategy, where only a part of the available (parliamentary) text would be analysed.

The validity of manual content analysis methods depends on the quality of the coding scheme. If the coding scheme accurately reflects conflicts between parties on relevant issue dimensions, its application will yield valid estimates. However, the application of the same coding scheme to many different countries and over a long period of time might lower the validity, because case-specific variation is ignored. Therefore, the saliency and network approaches show 'medium' levels of validity in table 4.4. Cross-validation of the saliency approach estimates also shows moderately high correlations with other measures of party policy, such as expert surveys (Klingemann et al., 2006; Keman, 2007). Because

[8] $\frac{(15*1)+(5*0)+(3*-1)}{(15+5+3)} = \frac{12}{23} = 0.52$

Table 4.4: *Comparison of strengths and weaknesses of several measurement approaches*[a]

	Manual coding			*Computerized coding*	
	Saliency approach	**Network approach**	**Confrontational approach**	**Wordscores**	**Wordfish**
Resource intenstity	Very High	Very High	High	Low	Low
Validity	Medium	Medium	Medium-High	Medium	Medium
Reliability	Medium	Medium	Medium	High	High
Certainty	Unknown[b]	Unknown	Unknown	Known	Known

[a] Adapted from Benoit and Laver (2006).
[b] Can be estimated using the algorithms of Mikhaylov et al. (2008) or Lowe et al. (2011).

the confrontational approach allows a researcher to select specific dimensions and items for each country and each election, it is likely to yield even more valid estimates. After all, this avoids the problems of applying a common framework to many different cases. The validity of the confrontational approach does, however, depend on the level of knowledge a researcher has about a particular case: the selection of dimensions and items is pivotal.

The reliability of the estimates is also an important characteristic. For content analysis techniques, this is primarily expressed as inter-coder reliability. It measures the degree to which different coders apply similar codes to sentences. The problem of many manual coding schemes is that inter-coder reliability is unknown. The saliency approach' Comparative Manifesto Project (CMP) only provides the score coders achieved on a 'coding test' and the Pellikaan et al. (2003) do not provide information on inter-coder reliability of the confrontational approach. Some publications employing the network approach do provide estimates, which range from a somewhat low Krippendorff's alpha of 0.67 to a more than fair 0.75 (Kleinnijenhuis et al., 2007). The more cohesive and more tailored to a specific case a coding scheme is, the higher inter-coder reliability will likely be. However, it is clear that every method that uses manual coding will suffer from inter-coder reliability problems to a certain degree .

The uncertainty of the estimates that are produced by manual content analysis is often unknown. It is, however, important to have an idea of the certainty of the estimates, especially if one is interested in changes in the positions of parties. After all, these changes might be simply the result of uncertainties in the estimation procedure. To distinguish between 'real' change and change that is the result of uncertainty, knowledge of the certainty of the estimates is necessary. Whereas one does probably have an intuitive feeling that an estimate based on coding a 10-sentence manifesto is probably less certain than an estimate that is based on a 10-page manifesto, this is often not expressed as a certainty estimate. For the Comparative Manifestos Project's estimates, some procedures have been proposed to estimate this uncertainty (Lowe et al., 2011; Mikhaylov et al., 2008). These procedures could also be modified to apply to the network approach' estimate. The confrontational approach is essentially deterministic; dimensions and items are explicated before the analysis, which means that there is no estimate of certainty that can be associated with the estimates.

Although there are variations in the underlying theories of party competition of the different types of manual content analysis as well as their coding methods, their strengths and weaknesses are generally similar. Their resource intensity is (very) high, (inter-coder) reliability is moderate, their validity is moderate to high and the certainty of the estimates is mostly unknown.

Computerized content analysis

Computerized methods of content analysis of political text have been developed as resource-extensive alternatives to manual content analysis. These techniques

usually rely on word counts[9]: they basically count how often each unique word is mentioned in each text.

One of the main selling points of computerized coding is that it drastically reduces the amount of work. Instead of having to read and code all texts, the researcher lets the computer count the unique words in the documents. These word count matrices are used to estimate party positions. Depending on the exact implementation this can reduce the time necessary to analyse a single manifesto from several hours to a few minutes (including preparation of the document). In recent years, more and more text has been made available digitally, often on the internet. Although preparing these texts for analysis does involve some work, especially when dealing with large quantities of, for example, parliamentary speech, the amount of work involved is small compared to coding a similar quantity of text manually. The resource intensity of computerized coding is thus low, especially in comparison with manual coding (table 4.4.). Because this study includes a very large amount of text (manifesto and parliamentary debates in twelve cases), this is a huge advantage.

For human procedures, validity is relatively easy to achieve, especially if the designer of the coding scheme is involved in the coding procedure. Computerized methods are more likely to suffer from invalidity problems, as the analysis is based on the analysis of word count matrices, rather than an understanding of what the text is about. While computerized techniques have become more sophisticated, validity will remain an issue. Generally, the outcomes of computerized techniques are compared to other party position benchmarks to show their validity (Budge and McDonald, 2007; Proksch and Slapin, 2009; Klemmensen et al., 2007; Slapin and Proksch, 2008). These benchmarks show that computerized coding methods do achieve moderate levels of validity.

Another advantage of using computers instead of human coders is the reliability of computers. Whereas human coders tend to disagree on the coding of sentences or items, computers will count words exactly the same time and time again. This does not mean that there is no uncertainty in the estimation, but this uncertainty does not stem from differences between the repeated applications of the coding procedure. Instead, uncertainty is related to the amount of available information: if one deals with very short texts the party position estimate is very uncertain, especially if word usage is not consistent. The uncertainty of the document scores can be estimated from variance in the parameters of the model. Information on the certainty of the estimates can, in turn, be used to assess whether party position changes are result of uncertainty of the estimate or reflect 'real' change.

Comparing the strengths and weaknesses of the various human and computerized content analysis techniques results in the conclusion that for this project, a computerized approach is most suitable. The volume of text that needs

[9]Some methods rely on other information, such as bi-grams (combinations of two words). However, both techniques presented here (Wordscores and Wordfish) rely on word count matrices.

to be analysed is simply too large to be analysed manually. At the same time, computerized methods have shown to produce valid and reliable party policy preference estimates. Therefore, this study uses a form of computerized content analysis.

This conclusion does not mean that computerized content analysis is suitable for all research questions. In some cases manual approaches are preferable. One example is when the amount of text is small. In those cases, computerized analysis will have problems to achieve robust and certain estimates, as I will discuss below. At the same time the resource intensity of manual coding is not so much a problem if there is very little text to read. Another situation in which the advantages of manual coding will probably outweigh the disadvantages, is when dealing with very specific issues or circumstances. In those cases the word-count based automatic methods might have trouble to find clear patterns in party speech, while human coders will be able to do so. The confrontational method seems particularly well-placed to deal with such situations, exactly because it leaves quite a lot of discretion for the researcher. However, in the current study the aim is to estimate parties' positions on rather broadly defined policy dimensions using a large volume of text. Here, the advantages of computerized methods are clear.

4.2.3 Computerized content analysis of political text

Wordscores

Although other computerized methods of content analysis of party manifestos have been attempted before, Wordscores presents the first method which has been applied successfully and has remained popular among scientists studying different countries and different sources. As the name suggests, it uses the word counts in documents to estimate party positions. It uses reference texts of which the position on a one-dimensional scale is known to estimate the position of virgin texts which have an unknown position. Laver et al. (2003) use the British manifestos of 1992 and 1997 as an example. Essentially, they use the manifestos of 1992 to estimate what words are typically used by a left-wing party and what words are typically used by a right-wing party. These are the word scores. Using these word scores they then estimate the position of the 1997 manifestos. If a 1997 manifesto uses a lot of right-wing words, its document score will be similar to that of a 1992 right-wing party and vice versa[10] . The method uses the information of word usage in the reference texts to estimate the policy positions of the virgin texts.

[10]The algorithm proceeds as follows: Laver, Benoit and Garry created a matrix of word counts, which contains the number of times each manifesto uses each unique word. For the reference texts, the 1992 manifestos, they enter the position of the party which is obtained from an external source, in this case an expert survey. The next step in their algorithm is to calculate a word score for each word w, which is the sum of the conditional probabilities of reading text r when reading word w times the reference position of text r, for each text r:

Wordscores requires the researcher to make a priori decisions about the nature of the issue dimensions he wants to study. By choosing a set of documents and assigning reference scores to these texts, based on an idea of an underlying dimension, the nature of the outcome dimension is determined in advance. However, if the match of reference and virgin text is poor, the estimate will have low validity. The method assumes that the word scores are similar for both the reference and virgin text, which ignores the context in which words are used. As time passes, the relevance of words and their substantive connotation can change. For example, a word like 'multicultural' was a rather broadly used term in the Netherlands in the 1990s, while it is now highly politicized. Another problem with the validity of the scores is that a lot of information in the text is not used; the algorithm only takes word counts into account. While this approach produces rather good estimates in some cases, it may fail to get the message in other cases[11].

The selection of reference texts is pivotal for the correct application of Wordscores. Laver et al. explicitly argue that the reference texts and the virgin texts

$$S_{wd} = \sum_r \left(\frac{F_{wr}}{(\sum_r F_{wr})} A_{rd}\right) \tag{4.1}$$

Where $F_w r$ is the relative frequency of word w in text r and $A_r d$ is the a priori reference text score of text r on dimension d. In other words, words that are only used by, for example, Labour gets a word score that is equal to the reference position of Labour in 1992. Words that are used to the same extent by the Liberals and the Conservatives receive a word score which is exactly the average of these parties' reference scores. The following step is to estimate the document scores of the virgin texts, which is the sum of the relative frequency times the word score, for each word:

$$S_{vd} = \sum_w (F_{wv} S_{wd}) \tag{4.2}$$

where $F_w v$ is a matrix of relative word counts for each word w for each virgin text v and $S_w d$ is the matrix of word scores from equation 4.1. The scores thus obtained need to be rescaled, because non-discriminating words will cluster together the scores of the virgin texts. Laver et al. (2003) use a transformation (the LBG transformation) that preserves the mean positions of the virgin scores, but makes their variance equal to the variance of the reference texts. An alternative transformation is proposed by Martin and Vanberg (MV), which depends only on the reference texts (Martin and Vanberg, 2007). It uses the ratio of the difference between the assigned reference scores ($A_r d$) and the raw document scores of the reference texts. Essentially it multiplies each raw scores by the number of times the raw scores of the (two most extreme) reference texts have are smaller than the assigned reference scores. For example, if the difference between the two reference text's raw scores is 0.5 and the difference between their assigned scores is 3, the multiplication factor is equal to 3 / 0.5 = 6. Benoit and Laver argue that the MV transformation could be used when the number of virgin texts is limited and the assumptions of the LBG transformation, namely that the variance of both sets of scores is equal, are not likely to be met (Benoit and Laver, 2008). Using the LBG transformation, the authors are able to calculate party manifesto positions for 1997 that are very close to expert survey estimate of those positions.

[11] Lowe (2008) points out several fundamental problems with Wordscores and suggests that it is basically an imperfect approximation of an ideal point model for words, which would only work under conditions which will never hold for word count data. His solution is to reformulate Wordscores as an ideal-point model, which can be estimated by maximum likelihood or inferred via Bayesian methods. This remains future work.

should use the same lexicon:

> This implies that we should resist the temptation to regard party manifestos as appropriate reference texts for analysing legislative speeches.
>
> (Laver et al., 2003: 315)

This does present a problem for the application of Wordscores to the current research question. After all, to test the party mandate the most logical thing to do in a Wordscores design would be to use the manifestos as reference texts for parliamentary speech in order to assess party position change. Although such an analysis would probably tell something about parties' positions, the differences between the parliamentary and manifesto lexicon will influence the scores. In fact any analysis which tries to directly compare word usage in manifestos and parliament would have to take these lexical differences into account. An alternative strategy, would be to estimate party positions in manifestos and in parliamentary speech separately. This would, in the case of Wordscores, however require two separate sets of reference texts and position: for manifestos and parliament. However, if these estimates would already be available, there would be no need to perform the Wordscores analysis at all.

Wordscores presents a large step forward in the computerized analysis of political texts. It is reliable, fast, relatively flexible and valid. However, the need for reference texts that are comparable to virgin texts does provide an obstacle for the current study, because it is problematic to use manifestos as reference texts for parliamentary debates. Even if this would work, it will mean that the parliamentary scores are calculated in terms of manifesto word usage. One of the criticisms on the pledge approach of the study of the party mandate is exactly that its analysis focuses too much on the manifestos and too little on the post-electoral behaviour of parties. Applying Wordscores to the current research question would yield a similar problem.

Wordfish

The main alternative to Wordscores also uses word count data to estimate party positions: Wordfish uses a poisson model that predicts the number of times a party mentions a certain word (Slapin and Proksch, 2008; Proksch and Slapin, 2008). The idea is that word usage is a function of a number of latent (unobserved) variables, one of which is a party's policy issue position. Other than Wordscores, the Wordfish algorithm does not use a priori information on the position of words. The method is thus inductive: it estimates parties' positions purely from the differences in their word usage. Wordfish does not require the use of reference texts, which is an advantage for the purpose of this study. On the other hand, using an inductive method does also present problems and challenges, exactly because the method picks up on differences in word usage.

The number of times a party mentions a certain word in its manifesto is an example of count data. Count data is usually assumed to follow a *poisson* distribution. Therefore Wordfish uses a poisson model to estimate the policy positions of actors from word count data. The poisson distribution has only one parameter, lambda (λ), which represents both the mean and the standard deviation. This makes it easier and quicker to estimate the model. The lambda parameter for each word j in document i is estimated by:

$$\lambda_{ij} = \exp(\alpha_i + \psi_j + \beta_j * \omega_i) \tag{4.3}$$

The equation consists of four parameters that all are all estimated by the Wordfish algorithm. Alpha (α) captures a 'fixed party effect', which essentially controls for differences in the length of party documents or speeches. When a document is very long, the word count of basically every word increases for that document, which is reflected in a higher estimate of alpha. The second parameter, psi (ψ), captures a 'fixed word effect'. Some words are used more often than other words in all texts, for example articles or words like 'we', 'it', or 'and'. This is captured by a value of psi for each word. Thus, words that are used more often in all texts, will have higher psi values. Beta (β) captures the amount of information word j conveys in distinguishing between party positions, or put differently the sensitivity of the occurrence of word j to changes in the policy position of the documents (Lowe, 2008). To put it simply, if a certain word is used many times by some parties but not by others, it discriminates well between parties and will get a highly positive or a highly negative beta value. Omega (ω) estimates the policy position of each document, the most interesting parameter for the purposes of this study.

Slapin and Proksch have made an algorithm available which they show to perform very well in finding the maximum likelihood estimator (2008)[12]. Wordfish estimates the positions of parties on a single scale. However, one can run separate analyses on different parts of the manifestos, for example on different issues. In this way the estimate of party positions can differ between issues.

Slapin and Proksch have successfully applied the Wordfish algorithm to estimate the policy positions of German party manifestos (Slapin and Proksch, 2008, 2009) and party speeches in the European Parliament (Proksch and Slapin, 2009). The method can thus be applied to both arenas. Furthermore, their analysis of the European Parliament speeches has shown that Wordfish works with different languages: the results of the German, English and French translations of the minutes were highly correlated.

The most important difference between Wordfish and Wordscores lies in the epistemology of the scales that are estimated. Wordscores uses a priori information from the reference texts and the external estimates of the positions of the

[12]The four parameters are estimated using an expectation-maximization algorithm that optimizes the log-likelihood of the model. Recently, Will Lowe's R package Austin has further improved the estimation algorithm and speed. However, the data analysis for this study had been completed earlier and has made use of Slapin and Proksch's Wordfish algorithm, version 1.3 (Proksch and Slapin, 2008).

reference text. The nature of the dimension is thus determined by the nature of the dimension on which the reference texts are scored. If the 'Taxes vs Spending' dimension from the Benoit-Laver expert survey is used, the virgin texts' positions should also be interpreted on this dimension (at least if the lexicon of both sets of text is comparable). Wordfish uses an a posteriori approach, which assumes "substantive meaning in the relative locations of key agents and uses this information to investigate the dimensional structure of the political space" (Benoit and Laver, 2006: 59). The algorithm does not use external information on the 'meaning' or 'scores' of words. It induces the party positions on a dimension basically from how different the word usage of parties is. If parties use many different words, they will have very different positions, if they use many similar words they will likely have a very similar position. Of course, a Wordfish analysis is not completely inductive. For one, researchers first have to split their text by issue category if they want to estimate party positions on more than one dimension. Secondly, as I will argue below, the application of Wordfish presents a number of practical problems. The algorithm sometimes picks up on patterns of word usage that are not of interest to the researcher. On example is that in some cases it finds that parties uses similar words in each year: they all talk about 'Bosnia' in one year, about 'Japan' in another and about 'nuclear weapons' in a third year. Thus, instead of picking up policy differences, the algorithm may find temporal differences. While I will show that it is possible to 'correct' for these influences, such a correction necessarily infringes on the 'a posteriori' character of the estimates. Nevertheless, compared to other approaches Wordfish uses relatively little a priori information.

The Wordfish algorithm can be separately applied to manifestos and parliamentary speech without the need for reference texts. The advantage of this is that it does not impose similarity between the electoral and parliamentary spaces of party competition. One does not have to assume that party competition on a specific issue, say for example the economy, is organized along the same lines during the elections and in parliament. Some argue that there is a difference between the dimensionality of the electoral and parliamentary space of competition (McDonald and Budge, 2005: 41). Of course, the meaning of the dimensions might be stable between elections and parliament, but for this study that is an empirical question rather than something which is to be easily assumed. Therefore the a posteriori approach of Wordfish fits well with the aim of this study to compare different spaces of competition. The major disadvantage of using an inductive method to estimate party position is that the algorithm might pick up on differences in word usage between documents that have little to do with substantive differences between parties. The validity of the scores on the dimension can be questioned in some cases, because they relate poorly to other measures of party policy preferences (see appendix A.3.5). Slapin and Proksch (2008) have, however, shown that their estimates of German party manifesto correlate quite strongly with other measures.

Another difference between Wordfish and Wordscores is the use of word

scores. The approach of Wordscores is to calculate the position of words on the policy dimension, using the assigned document scores of the reference texts. Wordfish, on the other hand, does not use word scores, but merely the informativeness parameter beta, which captures how informative a certain word is for the positions of parties. The advantage of having an informativeness parameter is that the model can distinguish very well between uninformative words — words that are used equally between parties — and informative words — words that are used a lot by some parties, but not by others. The disadvantage of Wordfish is that it does not use a word score parameter, because that would make it impossible to identify the model (Monroe and Maeda, 2004). The result is that the algorithm basically ignores words that are explicitly centrist (Lowe, 2008).

Given the amount of parliamentary speech in this study, computerized analysis is the preferable method from the perspective of resources. Both Wordscores and Wordfish have shown to produce reliable and, in most cases, valid estimates of party policy preferences. Furthermore, the methods can be applied in many different ways without the need for, for example, detailed and fixed coding schemes. The choice for Wordfish over Wordscores in this project is the result of two crucial arguments. First, for Wordscores there is a lack of a good set of reference texts and estimates for those texts and using manifestos as reference texts for parliamentary speech introduces all other kinds of problems. Secondly, the a posteriori approach of Wordfish seems to fit the purposes of this study very well. Exactly because my interest is in comparing the structure of the electoral and parliamentary spaces of competition, making a lot of assumptions on the similarity of those two spaces is problematic. Because Wordfish offers the opportunity to analyse the manifesto and parliamentary spaces separately, the two spaces are not 'forced' to be similar. This approach does bring along limitations for the interpretation of the spaces, because these are not directly comparable, which should be taken into account in the analysis and discussion of the results.

4.3 Data collection and the classification of text

Wordfish uses word-count matrices as its input. These matrices basically indicate how many times each document or speech uses each word. These matrices can be easily constructed from digitally available texts, but often some preprocessing is necessary, such as establishing to which party each person that took the parliamentary floor belonged to. Another step is to subdivide the text by issue. Although Wordfish can estimate parties' positions on a single dimension using the whole document, this study assumes that party competition might be different between issues: parties that agree on one issue might disagree on another issue. Therefore, I run separate analyses for different issue categories. Furthermore, part of the party mandate is that parties talk about similar issues in manifestos and in parliament (saliency). This requires the classification of manifesto paragraphs and parliamentary speech paragraphs.

4.3.1 Text collection and pre-processing

In the last few years, many political texts have become available in digital form. This greatly aids computerized analysis of these texts. In some cases, manifestos needed to be scanned, but most material was available on-line[13].

Three of the Dutch elections in the dataset, 1972, 1982 and 2003, were held only a year after the previous elections. In these cases, some parties did not write a new manifesto. Sometimes they just announced that their old manifesto was still valid, in other cases parties wrote a smaller pamphlet to 'update' their manifesto. In these cases, both the previous and current manifestos were used in the analysis.

Parliamentary debates are also available on-line for both the United Kingdom and the Netherlands. For most of the British parliamentary debates, the minutes do not record the party of an MP. This information was acquired from other sources, together with information on where the MP sat: on the front bench or back bench. The Dutch minutes have included party membership of the MPs since 1957, for the period before, this information was also acquired from other sources (Parlementair Documentatie Centrum, 2010).

4.3.2 Text classification

All texts were divided into paragraphs, which were subsequently classified in a two-step procedure. As the amount of text was too large to classify by hand, I used a computerized procedure. The first step of the procedure entails the classification of the paragraphs with the aid of a dictionary of signal words. I used the main categories of the British and Dutch Policy Agendas Project (Breeman et al., 2009). These are well-known classification schemes which are carefully designed to be exhaustive and mutually exclusive. For the Dutch case, I did add one category, namely Religion, designed to capture some of the religious remarks and issues that Christian parties often address [14]. This yields a total of 19 categories for Britain and 20 for the Netherlands.

For each of the categories, signal words were selected which indicate the discussion of that particular theme. For example, if the word 'refugee' is mentioned in a paragraph, one can be quite sure that the topic of this paragraph is migration, especially when the paragraph also contains words like 'asylum', 'foreigners'

[13]Details on the collection and pre-processing of these texts can be found in appendix A.1 on page 205.

[14]In addition, two categories were subdivided in a second stage of classification to distinguish between 'Medical-ethical' issues and other Health care issues as well as between 'Moral' issues and other 'Justice, Courts and Crime'. This second stage of classification applied a similar procedure as the one described in the main text, but only for one Policy Agendas Category. Thus, in case of Health care, a dictionary distinguished between 'General health care' and 'Medical-ethical issues' and paragraphs were assigned to one of those categories based on the occurrence of the dictionary words. This second stage classification did thus not change or influence the initial analysis. Medical-ethical and Moral issues were combined with the Religious issues in the Wordfish analysis to be able to capture a Religious-Secular dimension that is often used in analyses of Dutch politics.

and 'UNHCR'. For each issue category of interest, a moderately large number of these words (20 up to 50) were selected by reading the documents of interest and eye-balling for relevant words. Subsequently, the occurrence of each of these signal words in each paragraph was calculated. For example, if a paragraph contained four signal words of the Labour market category and only one for the Foreign affairs category, this paragraph is most likely to be about economic issues. I did correct for the length of the dictionary of each category by dividing the count for each category by the logged dictionary length. After all, if the list of signal words for a particular category is very large, one is more likely to come across one of these words. For the British dictionary, I used Laver and Garry's (2000) dictionary as a starting-point, the Dutch dictionary was partly based on the phrases used in the codebook of the Policy Agendas Project. I used unigrams (single words) as well as bi-grams (phrases of two words). Many of the dictionary words consist only of the first few letters of a word, designed to capture words with similar beginnings, but different endings, for example 'racis*' that matches 'racist, 'racism', 'racists' and all other words starting with 'racis'. For the Dutch case I used a slightly different version of the dictionary for the different years, because word usage in the first few years was quite different from later years. The same categories were used for all years in Britain and the Netherlands.

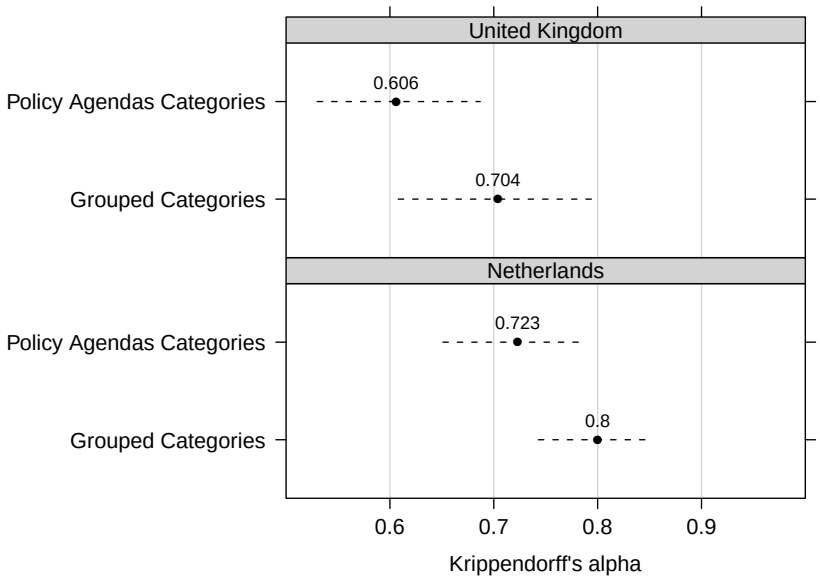
This procedure leaves an estimate of the category a particular paragraph belongs to, based on the occurrence of the signal words. In some cases, a paragraph cannot be classified using this dictionary approach: when it does not match any of the dictionary words or when dictionary words of two different categories occur equally frequently in a paragraph. In some collections of parliamentary speech, this leaves up to 40% of the paragraphs unclassified, which is too large a percentage to ignore. Therefore, paragraphs that could not be classified using the dictionary were classified with the help of an automatic classification algorithm[15].

To test the validity of the computerized coding, a sample of paragraphs from the 1992 (Britain) and 1994 (Netherlands) manifestos has also been coded manually (see appendix A.2 on page 207). Figure 4.1 presents the levels of Krippendorff's alpha, a measure of inter-coder reliability (or in this case 'computer-human reliability') which takes into account that agreement might also occur due to chance (Hayes and Krippendorff, 2007). For the 19 or 20 categories, the levels of alpha are somewhat low, especially for Britain. Nevertheless, these results are comparable to the results of other computerized techniques in similar applications (Breeman et al., 2009). I also calculated the accuracy for broader issue categories, which were used for the Wordfish estimation[16]. When looking at

[15]This algorithm uses the properties of the paragraphs that were classified using the dictionary approach to estimate a classification of the other paragraphs. The application of this Support Vector Machine (SVM) is outlined in appendix A.2 on page 207

[16]As is explained below, the Comparative Agenda Project's classification scheme with 19 or 20 categories yields too little text per category. Therefore, I grouped similar categories together to achieve

Figure 4.1: *Results of a comparison of manual and computerized classification*



Note: Dotted lines indicate 95% confidence intervals.

the computer-human reliability levels for these grouped categories, the levels of Krippendorff's alpha increase to an acceptable level of 0.7 for Britain and 0.8 for the Netherlands. The method used here does not produce a perfect agreement with the manual classifications, but for the purposes here it is acceptable. Most errors found represent somewhat ambivalent paragraphs. Wherever these are classified, they will introduce some noise into the analysis. As long as the correct classifications are a large majority, Wordfish should be able to deal with the noise in the dataset.

4.4 Using Wordfish to compare elections and parliaments

The application of any statistical method requires careful attention to the requirements of those methods and to any deficiencies they might produce. This is especially true in the application of the method is extended to cases that never have been analysed in that particular way before. In this case, Wordfish has been shown to work with both parliamentary and manifesto data, but not yet for the particular countries that are studied here. Furthermore, the comparison of parliamentary debate and manifesto requires attention.

The 19 (Britain) or 20 (Netherlands) policy agendas categories that were obtained via the classification procedure described above were combined into broader categories for the purpose of the Wordfish analysis (see table 4.5). Quite a few of these 20-odd categories received only sparse attention from parties, especially in the manifestos. The application of Wordfish to the text parties devote to these issues would result in very uncertain estimates. Slapin and Proksch's analysis of Germany consists, for example, only of three policy areas (2008). For the first three Dutch elections I limited the analysis to four issue categories, for the last three to eight. Manifestos and parliamentary speeches were longer in later years, dealing with a broader range of topics, allowing for a more detailed analysis. For Britain, five categories were used throughout. The combination of issues is based on an a priori understanding of their connection. For example, issues on macro-economy can be said to relate quite closely to business and labour market. Of course, this does introduce an element of choice into the analysis. However, the categories chosen relate closely to ideas about important policy dimensions (Pellikaan et al., 2003; Benoit and Laver, 2006). The issue categories are always the same for the manifesto and parliamentary competition in a certain case.

The manifesto and parliamentary documents were analysed separately. I ran a separate Wordfish analysis for each issue category in the manifestos and also for each issue category in parliament. To achieve (more) robust estimates of party positions in manifestos, I included the previous and next manifestos in the ana-

more robust estimates of parties' policy positions.

lysis. For example, the analysis of the 1992 manifesto positions includes the party manifestos of 1987, 1992 and 1997. Note that the 1987 and 1997 manifestos do not directly affect the positions of the parties in 1992, they only serve to increase the robustness of the word parameters. As the number of documents in the analysis increases, the word parameters will generally become more robust[17]. The 1992 manifesto positions of parties are only based on their 1992 manifestos. Including the previous and next documents is in any case necessary for the British manifestos, as the analysis would otherwise be based on only three documents (a single manifesto for each party). Because the position estimates are induced from differences in parties' word choice, having only three different parties leads to uncertain estimates. After all, idiosyncrasies are likely to influence the analysis much more when there are only three documents compared to having more[18]. For the parliamentary debates, texts were split per year: the texts produced by the MPs for a party in a single year were thus treated as a single observation. For the subsequent analyses, I used the mean of these per-year estimates (weighted by the length of the text per year) to compare parties' parliamentary positions with their manifesto positions.

To increase the robustness of the analysis, I applied two filters to the word count matrix. First, I excluded words that are used only in one document[19]. These words will distinguish perfectly between the one party (in one year) and all other parties. Wordfish does provide a correction for this perfect discrimination, but it is still recommended to remove words that are only used in one document. However, the result might be that a party with a very outspoken and unique position will be estimated to be moderate, because all of its unique words were eliminated[20]. For example, the speech of extreme-right parties on migration contains some words which are never used by other parties. Therefore, words that were used at least three times (manifestos) or five times (parliamentary speeches) were not eliminated. Secondly, very short manifestos (less than 50 word counts in the filtered dataset) or parliamentary speeches (less than 250 word counts in the filtered datasets) were excluded. The analysis of these texts is not only likely to arrive at uncertain estimates, they may also influence the positions of other parties[21].

[17]The assumption is that word usage was roughly similar in three subsequent elections.

[18]Other Wordfish analyses include more documents from a longer time frame (e.g. Slapin and Proksch, 2009). I found that both in the British and Dutch case, one would have to delete many year-specific words to ensure that Wordfish does not merely distinguish documents from different years instead of documents with different positions (see appendix A.3.2 on page 212).

[19]Document here refers to a party's manifesto in one year or the collection of speeches of one party in one parliamentary year.

[20]Especially when the party only participated in only one election.

[21]Furthermore, I corrected for the comparison of positions over time in a parliament as well as for the positioning of the (UK) government. These issues are discussed in appendix A.3 on page 210.

Table 4.5: *Overview of observations*

Country	Periods	Issue categories
United Kingdom	1955-1959 1966-1970 1970-1979 1983-1987 1992-1997 2001-2005	Economy; Environment; Foreign Affairs and Defence; Government Operations; Law and Order and Migration
Netherlands[a]	1952-1956 1959-1963 1972-1977	Economy, Health Care and Education; Foreign Affairs and Defence; Post-materialist issues; Religion, Morals and Medical-ethical
	1982-1986 1994-1998 2003-2006	Economy; Health Care and Education; Environment; Foreign Affairs and Defence; Migration; Justice, Courts and Crime; Democracy and Civil Rights; Religion, Morals and Medical-Ethical

[a] For the Netherlands different issue categories were used for the first three and last three time periods, reflecting a difference in the level of detail of manifestos and parliamentary debate.

4.5 Constructing spatial representations of party competition

Wordfish produces estimates of party positions in the electoral and parliamentary arena on a number of issue dimensions. These data are used to study the congruence of parties in both arenas on individual issue dimensions. The information can also be used in a different way, namely to construct and compare the structure of the spaces of electoral and parliamentary competition between parties.

Spatial representations of party competition are an abstract way of describing what political conflict between parties is about and how parties are positioned or (strategically) position themselves towards other parties. Spatial analyses of party competition have long been used, because they offer a concise and intuitive way of displaying political conflict. Early on, Downs (1957) used a one-dimensional representation of party positions to study the dynamics of party competition. Others have extended this type of analysis to a multidimensional space, because they found that political competition could not accurately be captured on a single dimension. After all, parties that think very differently on the economy, might very well agree with each other about European integration. The spatial approach offers a way of visualizing how inter-party conflicts are related. For example, differences between parties may be very similar on one set of issues while party conflict is very different on another set of issues. This acknowledges the fact that political conflict is not simply the aggregate of differences on specific issue dimensions. The way in which party positions on issues are related is a basic characteristic of political competition.

Many spatial analyses of political party competition use a predefined notion of the relevant dimensions of the policy space. In a particular case, researchers may argue that party competition can be best described by a single Left-Right dimension, or by looking at economic left-right and GAL-TAN[22] as a secondary dimension. This is an excellent approach if one is interested only in the position or movement of parties within a given space. However, it cannot analyse the transformation of the space itself. This is one of the aims here: to look at the congruence between the electoral and parliamentary space of competition. This involves the dimensionality of the space: which issues matter and how are issues related? In other words, the purpose here is not just to study the positions of the pieces on the chess board, but also the properties of the board itself. After all, political chess is not limited to an eight by eight square. If one assumes that the board has certain properties, these cannot be studied as variables. Therefore, the properties of the space should not be determined a priori.

The technique to construct spatial models of party competition used here is

[22]Green, Alternative, Libertarian versus Traditional, Authoritarian, Nationalist. This dimension summarizes 'new politics' issues, which are believed to relate poorly to the Left-Right divide (Marks et al., 2006: 157).

classic multidimensional scaling (MDS). MDS is a flexible technique that uses information on the differences (distances) between parties on a high number of issue dimensions to construct a low dimensional picture of parties' positions (Kruskal and Wish, 1978; Van der Brug, 1997; Borg and Groenen, 1997). The text-book example of MDS starts with a table containing the distances between cities. From this table one could easily draw a map. In the case of the map, there is a perfect fit, because the distances between the cities were actually measured from a map. Here, the fit of a two-dimensional representation is normally not perfect, because the differences between parties stem from many issue dimensions. Nevertheless, a low-dimensional spatial representation can often capture the most important differences between parties. I applied metric MDS here, because the positions have been estimated at the interval level. The distances that are used for this space stem from the different positions parties have on the issue dimensions. The distance between two parties is the Euclidean distance between their positions on all issue dimensions, which is for parties a and b equal to:

$$distance_{ab} = \sqrt{\sum_i (p_{ia} - p_{ib})^2} \tag{4.4}$$

Where p_{ia} is the position of party a on issue i. One alteration was made to the distance scores to take the salience of issues into account. Normally, the distance between parties on each issue weighs equally in an MDS analysis. However, in this study some issues are far more salient overall than other issues and therefore can said to be more important in the space of competition. After all, if parties differ very much on an issue that they talk about a lot, this is more relevant for the competition between those parties than their difference on an issue that neither of them really cares about[23]. Therefore, parties positions are multiplied by the square root of the average saliency of an issue. This preserves the relative distances between parties on each single dimension, but it lets salient issues contribute more to the total distance matrix.

The result of the MDS analysis is a low-dimensional space of party positions. Generally, one will find a one or two-dimensional solution that fits the data quite well. This space can be visualized as a plot of party positions. The horizontal and vertical dimensions of these plots have no substantive meaning: the plot can be freely rotated without losing information. Whether parties are on the 'left' or on the 'right' of the picture thus has no substantive meaning[24].

[23] Furthermore, if one were to split an issue dimension (e.g. Health Care and Education) into smaller ones, the differences on those dimensions would become more influential in the analysis. In other words: the selection of dimensions (specifically the choice to aggregate or split them) would influence the analysis very much. From this perspective, taking into account saliency in the construction of the space of competition makes also sense, because in that case the influence of two separate 'Health Care' and 'Education' dimensions would be similar to the influence of a single combined dimension.

[24] To ease interpretation I did however (if necessary) invert the dimensions so that parties that are generally regarded as left-wing are on the left.

To aid the interpretation of the space, I plotted the separate issue dimensions into the graph with a technique called 'property fitting'. The dashed lines indicate how a separate issue dimensions fits into the combined space. One can find the approximate position of a party on a separate dimension by running a line perpendicular to the dashed line, through the party estimate. The fitted lines can be found by running a regression analysis where the positions of parties on dimensions of the MDS solution explain party positions on a separate issue dimension. The slope of the line is equal to the ratio of the b's in the regression analysis[25]. This will only make sense if the explained variance (R^2) of the regression analysis is sufficiently high: if the dimensions of the MDS space cannot explain parties' positions on a specific issue dimension, one cannot draw a fitting line. Those lines were omitted if the R^2 was lower than 0.8.

For each case a separate electoral and a separate parliamentary space were constructed. After all, the Wordfish analysis resulted in estimates of party positions that cannot be compared in absolute terms: the relative distances between parties can be compared between elections and parliament, but not the absolute position of a single party in both arenas. Therefore, one cannot construct a single space in which both parties' electoral and their parliamentary positions can be represented. Although this does mean that the party positions cannot be compared directly between the two spaces, it does offer a good way to study whether the electoral and parliamentary spaces are similar. After all, the institutional perspective on the party mandate taken here goes beyond the congruence of an individual parties' positions and primarily focuses on the congruence of the spaces themselves. This involves looking at three aspects: the dimensionality of the spaces (are one, two or more dimensions necessary to display inter-party differences?), the location of issue dimensions in the spaces and the structure of party positions within the electoral and parliamentary spaces.

[25]It is given by: b_2/b_1. Note that the aspect ratio of the graph should be equal to one: 'stretching' the graph will disturb the angles of the parties' positions with the fitted issue dimension lines. Van der Brug (1997) describes a different procedure for estimating the slope of the lines.