5-2003

Extracting policy positions from political texts using words as data

Michael LAVER

Kenneth BENOIT
Singapore Management University, kbenoit@smu.edu.sg

John GARRY

Extracting Policy Positions from Political Texts Using Words as Data

MICHAEL LAVER and KENNETH BENOIT *Trinity College, University of Dublin*
JOHN GARRY *University of Reading*

We present a new way of extracting policy positions from political texts that treats texts not as discourses to be understood and interpreted but rather, as data in the form of words. We compare this approach to previous methods of text analysis and use it to replicate published estimates of the policy positions of political parties in Britain and Ireland, on both economic and social policy dimensions. We "export" the method to a non-English-language environment, analyzing the policy positions of German parties, including the PDS as it entered the former West German party system. Finally, we extend its application beyond the analysis of party manifestos, to the estimation of political positions from legislative speeches. Our "language-blind" word scoring technique successfully replicates published policy estimates without the substantial costs of time and labor that these require. Furthermore, unlike in any previous method for extracting policy positions from political texts, we provide uncertainty measures for our estimates, allowing analysts to make informed judgments of the extent to which differences between two estimated policy positions can be viewed as significant or merely as products of measurement error.

Analyses of many forms of political competition, from a wide range of theoretical perspectives, require systematic information on the policy positions of the key political actors. This information can be derived from a number of sources, including mass, elite, and expert surveys either of the actors themselves or of others who observe them, as well as analyses of behavior in strategic settings, such as legislative roll-call voting. (For reviews of alternative sources of data on party positions, see Laver and Garry 2000 and Laver and Schofield 1998). All of these methods present serious methodological and practical problems. Methodological problems with roll-call analysis and expert surveys concern the direction of causality—"data" on policy positions collected using these techniques are arguably more a product of the political processes under investigation than causally prior to them. Meanwhile, even avid devotees of survey techniques cannot rewind history to conduct new surveys in the past. This vastly restricts the range of cases for which survey methods can be used to estimate the policy positions of key political actors.

An alternative way to locate the policy positions of political actors is to analyze the texts they generate. Political texts are the concrete by-product of strategic political activity and have a widely recognized potential to reveal important information about the policy positions of their authors. Moreover, they can be analyzed, reanalyzed, and reanalyzed again without becoming jaded or uncooperative. Once a text and an analysis technique are placed in the public domain, furthermore, others can replicate, modify, and improve the estimates involved or can produce completely new analyses using the same tools. Above all, in a world where vast volumes of text are easily, cheaply, and almost instantly available, the systematic analysis of political text has the potential to be immensely liberating for the researcher. Anyone who cares to do so can analyze political texts for a wide range of purposes, using historical texts as well as analyzing material generated earlier in the same day. The texts analyzed can relate to collectivities such as governments or political parties or to individuals such as activists, commentators, candidates, judges, legislators, or cabinet ministers. The data generated from these texts can be used in empirical elaborations of any of the huge number of models that deal with the policies or motivations of political actors. The big obstacle to this process of liberation, however, is that current techniques of systematic text analysis are very resource intensive, typically involving large amounts of highly skilled labor.

One current approach to text analysis is the "hand-coding" of texts using traditional—and highly labor-intensive—techniques of content analysis. For example, an important text-based data resource for political science was generated by the Comparative Manifestos Project (CMP)[1] (Budge, Robertson, and Hearl 1987; Budge et al. 2001; Klingemann, Hofferbert, and Budge 1994; Laver and Budge 1992). This project has been in operation since 1979 and, by the turn of the millennium, had used trained human coders to code 2,347 party manifestos issued by 632 different parties in 52 countries over the postwar era (Volkens 2001, 35). These data have been used by many authors writing on a wide range of subjects in the world's most prestigious journals.[2] Given the immense sunk costs of

Michael Laver's work on this paper was carried out while he was a Government of Ireland Senior Research Fellow in Political Science, Trinity College, University of Dublin, Dublin, Ireland (mlaver@tcd.ie).

Kenneth Benoit's work on this paper was completed while he was a Government of Ireland Research Fellow in Political Science, Trinity College, University of Dublin, Dublin, Ireland (kbenoit@tcd.ie).

John Garry is Lecturer in the Politics Department, University of Reading, White Knights Reading, Berkshire RG6 6AH, UK (j.a.garry@reading. ac.uk).

We thank Raj Chari, Gary King, Michael McDonald, Gail McElroy, and three anonymous reviewers for comments on drafts of this paper.

[1] Formerly the Manifesto Research Group (MRG).

[2] For a sample of such publications, see Adams 2001; Baron 1991, 1993; Blais, Blake, and Dion 1993; Gabel and Huber 2000; Kim and Fording 1998; Schofield and Parks 2000; and Warwick 1994, 2001, 2002.

generating this mammoth data set by hand over a period of more than 20 years, it is easy to see why no other research team has been willing to go behind the very distinctive theoretical assumptions that structure the CMP coding scheme or to take on the task of checking or replicating any of the data.

A second approach to text analysis replaces the hand-coding of texts with computerized coding schemes. Traditional computer-coded content analysis, however, is simply a direct attempt to reproduce the hand-coding of texts, using computer algorithms to match texts to coding dictionaries. With proper dictionaries linking specific words or phrases to predetermined policy positions, traditional techniques for the computer-coding of texts can produce estimates of policy positions that have a high cross-validity when measured against hand-coded content analyses of the same texts, as well as against completely independent data sources (Bara 2001; de Vries, Giannetti, and Mansergh 2001; Kleinnijenhuis and Pennings 2001; Laver and Garry 2000). Paradoxically, however, this approach does not dispense with the need for heavy human input, given the extensive effort needed to develop and test coding dictionaries that are sensitive to the strategic context—both substantive and temporal—of the texts analyzed. Since the generation of a well-crafted coding dictionary appropriate for a particular application is so costly in time and effort, the temptation is to go for large general-purpose dictionaries, which can be quite insensitive to context. Furthermore, heavy human involvement in the generation of coding dictionaries imports some of the methodological disadvantages of traditional techniques based on potentially biased human coders.

Our technique breaks radically from "traditional" techniques of textual content analysis by treating texts not as discourses to be read, understood, and interpreted for meaning—either by a human coder or by a computer program applying a dictionary—but as collections of word data containing information about the position of the texts' authors on predefined policy dimensions. Given a set of texts about which something is known, our technique extracts data from these in the form of word frequencies and uses this information to estimate the policy positions of texts about which nothing is known. Because it treats words unequivocally as data, our technique not only allows us to estimate policy positions from political texts written in any language but also, uniquely among the methods currently available, allows us to calculate confidence intervals around these point estimates. This in turn allows us to make judgments about whether estimated differences between texts have substantive significance or are merely the result of measurement error. Our method of using words as data also removes the necessity for heavy human intervention and can be implemented quickly and easily using simple computer software that we have made publicly available.

Having described the technique we propose, we set out to cross-validate the policy estimates it generates against existing published results. To do this we reanalyze the text data set used by Laver and Garry (2000) in their dictionary-based computer-coded content analysis of the manifestos of British and Irish political parties at the times of the 1992 and 1997 elections in each country. We do this to compare our results with published estimates of the policy positions of the authors of these texts generated by dictionary-based computer-coding, hand-coded content analyses, and completely independent expert surveys. Having gained some reassurance from this cross-validation, we go on to apply the technique to additional texts not written in English. Indeed estimating policy positions from documents written in languages unknown to the analyst is a core objective of our approach, which uses computers to minimize human intervention by analyzing text as data, while making no human judgement call about word meanings. Finally, we go on to extend the application of our technique beyond the analysis of party manifestos, to the estimation of legislator positions from parliamentary speeches. If our method can be demonstrated to work well in these various contexts, then we would regard it as an important methodological advance for studies requiring estimates of the policy positions of political actors.

A MODEL FOR LOCATING POLITICAL TEXTS ON *A PRIORI* POLICY DIMENSIONS

A Priori or Inductive Analyses of Policy Positions?

Two contrasting approaches can be used to estimate the policy positions of political actors. The first sets out to estimate positions on policy dimensions that are defined *a priori*. A familiar example of this approach can be found in expert surveys, which offer policy scales with predetermined meanings to country experts who are asked to locate parties on them (Castles and Mair 1984; Laver and Hunt 1989). Most national election and social surveys also ask respondents to locate both themselves and political parties on predefined scales. Within the realm of text analysis, this approach codes the texts under investigation in a way that allows the estimation of their positions on *a priori* policy dimensions. A recent example of this way of doing things can be seen in the dictionary-based computer-coding technique applied by Laver and Garry (2000), which applies a predefined dictionary to each word in a political text, yielding estimated positions on predefined policy dimensions.

An alternative approach is fundamentally *inductive*. Using content analysis, for example, observed patterns in texts can be used to generate a matrix of similarities and dissimilarities between the texts under investigation. This matrix is then used in some form of dimensional analysis to provide a spatial representation of the texts. The analyst then provides substantive meanings for the underlying policy dimensions of this derived space, and these *a posteriori* dimensions form the basis of subsequent interpretations of policy positions. This is the approach used by the CMP in its hand-coded content analysis of postwar European party manifestos (Budge, Robertson, and Hearl 1987), in which data analysis is designed to allow inferences to be made about the dimensionality of policy spaces and the substantive meaning of policy dimensions. A forthright recent use of this approach for a single left–right dimension can be found in Gabel and Huber 2000. Warwick (2002) reports a multidimensional inductive analysis of both content analysis and expert survey data.

It should be noted that a *purely* inductive spatial analysis of the policy positions of political texts is impossible. The analyst has no way of interpreting the derived spaces without imposing at least some *a priori* assumptions about their dimensionality and the substantive meaning of the underlying policy dimensions, whether doing this explicitly or implicitly. In this sense, all spatial analyses boil down to the estimation of policy positions on *a priori* policy dimensions. The crucial distinction between the two approaches concerns the point at which the analyst makes the substantive assumptions that allow policy spaces to be interpreted in terms of the real world of politics. What we have called the *a priori* approach makes these assumptions at the outset since the analyst does not regard either the dimensionality of the policy space or the substantive meaning of key policy dimensions as the essential research questions. Using prior knowledge or assumptions about these reduces the problem to an epistemologically straightforward matter of estimating unknown positions on known scales. What we have called the inductive approach does not make prior assumptions about the dimensionality of the space and the meaning of its underlying policy dimensions. This leaves too many degrees of freedom to bring closure to the analysis without making *a posteriori* assumptions that enable the estimated space and its dimensions to be interpreted.

The ultimate methodological price to be paid for the benefits of *a posteriori* interpretation is the lack of any objective criterion for deciding between rival spatial interpretations, in situations in which the precise choice of interpretation can be critical to the purpose at hand. The price for taking the *a priori* route, on the other hand, is the need to accept take-it-or-leave-it propositions about the number and substantive meaning of the policy dimensions under investigation. Using the *a priori* method we introduce here, however, this price can be drastically reduced. This is because, once texts have been processed, it is very easy to reestimate their positions on a new *a priori* dimension in which the analyst might be interested. For this reason we concentrate here on estimating positions on *a priori* policy dimensions. The approach we propose can be adapted for inductive analysis with *a posteriori* interpretation, however, and we intend to return to this in future work.

The Essence of Our *A Priori* Approach

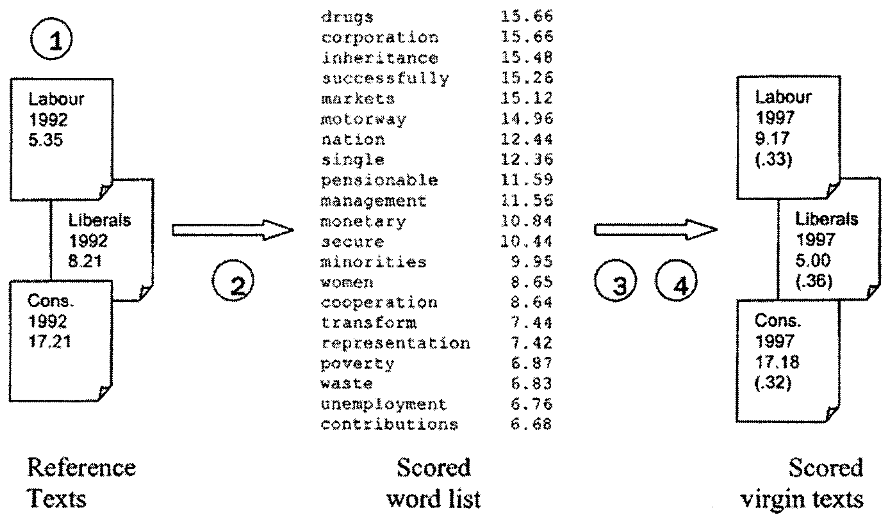
Our approach can be summarized in nontechnical terms as a way of estimating policy positions by comparing two sets of political texts. On one hand is a set of texts whose policy positions on well-defined *a priori* dimensions are "known" to the analyst, in the sense that these can be either estimated with confidence from independent sources or assumed uncontroversially. We call these "reference" texts. On the other hand is a set of texts whose policy positions we do not know but want to find out. We call these "virgin" texts. All we do know about the virgin texts is the words we find in them, which we compare to the words we have observed in reference texts with "known" policy positions.

More specifically, we use the relative frequencies we observe for each of the different words in each of the reference texts to calculate the probability that we are reading a particular reference text, given that we are reading a particular word. For a particular *a priori* policy dimension, this allows us to generate a numerical "score" for each word. This score is the expected policy position of any text, given only that we are reading the single word in question. Scoring words in this way replaces the predefined deterministic coding dictionary of traditional computer-coding techniques. It gives words policy scores, not having determined or even considered their meanings in advance but, instead, by treating words purely as data associated with a set of reference texts whose policy positions can be confidently estimated or assumed. In this sense the set of real-world reference texts replaces the "artificial" coding dictionary used by traditional computer-coding techniques.

The value of the set of word scores we generate in this way is not that they tell us anything new about the reference texts with which we are already familiar—indeed they are no more than a particular type of summary of the word data in these texts. Our main research interest is in the virgin texts about which we have no information at all other than the words they contain. We use the word scores we generate from the reference texts to estimate the positions of virgin texts on the policy dimensions in which we are interested. Essentially, each word scored in a virgin text gives us a small amount of information about which of the reference texts the virgin text most closely resembles. This produces a conditional expectation of the virgin text's policy position, and each scored word in a virgin text adds to this information. Our procedure can thus be thought of as a type of Bayesian reading of the virgin texts, with our estimate of the policy position of any given virgin text being updated each time we read a word that is also found in one of the reference texts. The more scored words we read, the more confident we become in our estimate.

Figure 1 illustrates our procedure, highlighting the key steps involved. The illustration is taken from the data analysis we report below. The reference texts are the 1992 manifestos of the British Labour, Liberal Democrat (LD), and Conservative parties. The research task is to estimate the unknown policy positions revealed by the 1997 manifestos of the same parties, which are thus treated as virgin texts. When performed by computer, this procedure is entirely automatic, following two key decisions by the analyst: the choice of a particular set of reference texts and the identification

FIGURE 1. The Wordscore procedure, using the British 1992–1997 manifesto scoring as an illustration



Step 1: Obtain reference texts with a priori known positions (`setref`)
Step 2: Generate word scores from reference texts (`wordscore`)
Step 3: Score each virgin text using word scores (`textscore`)
Step 4: (optional) Transform virgin text scores to original metric

Note: Scores for 1997 virgin texts are transformed estimated scores; parenthetical values are standard errors. The scored word list is a sample of the 5,299 total words scored from the three reference texts.

of an estimated or assumed position for each reference text on each policy dimension of interest.

Selection of Reference Texts

The selection of an appropriate set of reference texts is clearly a crucial aspect of the research design of the type of *a priori* analysis we propose. If inappropriate reference texts are selected, for example, if cookery books are used as reference texts to generate word scores that are then applied to speeches in a legislature, then the estimated positions of these speeches will be invalid. Selecting reference texts thus involves crucial substantive and qualitative decisions by the researcher, equivalent to the decisions made in the design or choice of either a substantive coding scheme for hand-coded content analysis or a coding dictionary for traditional computer-coding. While there are no mechanical procedures for choosing the reference texts for any analysis, we suggest here a number of guidelines as well as one hard-and-fast rule.

The hard-and-fast rule when selecting reference texts is that we must have access to confident estimates of, or assumptions about, their positions on the policy dimensions under investigation. Sometimes such estimates will be easy to come by. In the data analyses that follow, for example, we seek to compare our own estimates of party policy positions with previously published estimates. Thus we replicate other published content analyses of party manifestos, using "reference" party manifestos from one election to estimate the positions of "virgin" party manifestos in the next election. Our reference scores are taken from published expert surveys of the policy positions of the reference text authors, although this is only one of a number of easily available sources that we could have used with reasonable confidence. While a number of flaws can certainly be identified with expert surveys—some of which we have already mentioned—our purpose here is to compare the word scoring results with a well-known and widely used benchmark. In using these particular reference texts, we are in effect assuming that party manifestos in country c at election t are valid points of reference for the analysis of party manifestos at election $t+1$ in the same country. Now this assumption is unlikely to be 100% correct, since the meaning and usage of words in party manifestos change over time, even over the time period between two elections in one country. But we argue not only that it is likely to be substantially correct, in the sense that word usage does not change very much over this period, but also that there is no better context for interpreting the policy positions of a set of party manifestos at election $t+1$ than the equivalent set of party manifestos at election t. Note, furthermore, that any attempt to estimate the policy position of any political text, using any technique whatsoever, must relate this to some external context if the result is to be interpreted in a meaningful way, so that some equivalent assumption must always be made. As two people facing each other quickly discover, any attempt to describe one point as being to the "left" or the "right" of some other point must always have recourse to some external point of reference.

There may be times, however, when it is not easy to obtain simultaneously an authoritative set of reference texts and good estimates of the policy positions of these on all *a priori* dimensions in which the analyst is interested. In such instances it is possible to assume specific values for reference texts representing quintessential expressions of a view or policy whose position is known with a high degree of *a priori* confidence. Later in this paper, we apply our technique to legislative speeches made during a no-confidence debate, *assuming* that the speech of the leader of the government is quintessentially progovernment and that the speech of the leader of the opposition is quintessentially antigovernment.

In other words, what we require for our set of reference texts is a set of estimates of, or assumptions about, policy positions that we are prepared to stand over and use as appropriate points of reference when analyzing the virgin texts in which we are ultimately interested. Explicit decisions of substantive importance have to be made about these, but these are equivalent to the implicit decisions that must always be made when using other techniques for estimating policy positions. We do essentially the same thing when we choose a particular hand-coding scheme or a computer-coding dictionary, for example, both of which can always be deconstructed to reveal an enormous amount of (often hidden) substantive content. The need to choose external points of reference is a universal feature of any attempt to estimate the policy positions of political actors. In our application, the external points of reference are the reference texts.

We offer three further general guidelines in the selection of reference texts. The first is that the reference texts should use the same lexicon, in the same context, as the virgin texts being analyzed. For example, our investigations have (unsurprisingly) revealed very different English-language lexicons for formal written political texts, such as party manifestos, and formal spoken texts, such as speeches in a legislature. This implies that we should resist the temptation to regard party manifestos as appropriate reference texts for analyzing legislative speeches. In what follows, we use party manifestos as reference texts for analyzing other party manifestos and legislative speeches as reference texts for other legislative speeches. The point is that our technique works best when we have a number of "virgin" texts about which we know nothing and want to relate these to a small number of lexically equivalent (or very similar) "reference" texts about which we know, or are prepared to assume, something.

The second guideline is that policy positions of the reference texts should "span" the dimensions in which we are interested. Trivially, if all reference texts have the same policy position on some dimension under investigation, then their content contains no information that can be used to distinguish between other texts on the same policy dimension. An ideal selection of reference texts will contain texts that occupy extreme positions, as well as positions at the center, of the dimensions under investigation. This allows differences in the content of the reference texts to form the basis of inferences about differences in the content of virgin texts.

The third general guideline is that the set of reference texts should contain as many different words as possible. The content of the virgin texts is analyzed in the context of the word universe of the reference texts. The more comprehensive this word universe, and thus the less often we find words in virgin texts that do not appear in any reference text, the better. The party manifestos that we analyze below are relatively long documents. The British manifestos, for example, are between 10,000 and 30,000 words in length, each using between about 2,000 and 4,000 unique words. Most words observed in the virgin texts can be found in the word universe of the reference texts, while those that cannot tend to be used only very occasionally.[3] If the texts in which we are interested are much shorter than this—for example, legislative speeches are typically shorter than party manifestos—then this will tend to restrict the word universe of the reference texts and may reduce our ability to make confident inferences about the policy positions of virgin texts. As we show below when analyzing legislative speeches, the uncertainty of our estimates does increase when texts are short, although it is worth noting that, when other methods of content analysis use short texts, they typically report no estimate at all of the associated increase in uncertainty.[4] The problem of short texts is thus a problem with any form of quantitative content analysis and is not in any way restricted to the technique we propose here. And if the texts in which we are genuinely interested are short, then they are short and we just have to make the best of the situation in which we find ourselves. But the principle remains that it is always better to select longer suitable texts when these are available.

Generating Word Scores from Reference Texts

We begin with set R of reference texts, each having a policy position on dimension d that can be estimated or assumed with confidence. We can think of the estimated or assumed position of reference text r on dimension d as being its *a priori* position on this dimension, A_{rd}. We observe the relative frequency, as a proportion of the total number of words in the text, of each different word w used in reference text r.[5] Let this be F_{wr}. Once

[3] We are more specific about this when discussing particular results below.

[4] We note that in the widely used content analysis data set of the CMP, many of the texts analyzed are very short. Using the CD-ROM distributed with Budge et al. 2001, we find that about one-third of all texts in the data set comprise fewer than 100 quasi-sentences. Generously estimating each quasi-sentence to be about 20 words, this implies that one-third of the CMP texts are about 2,000 words or fewer, while well over half of all texts analyzed are probably fewer than 4,000 words each.

[5] In the analyses reported here, we use the relative frequencies of every single different word in each reference text, even very common words such as prepositions and indefinite articles. We do this for two reasons. First, to do otherwise would require knowledge of the language in which the text under analysis was written, violating our principle of treating words as data and undermining our fundamental objective of being able to analyze texts written in languages we do not understand. Second, where such common words are systematically

we have observed F_{wr} for each of the reference texts, we have a matrix of relative word frequencies that allows us to calculate an interesting matrix of conditional probabilities. Each element in the latter matrix tells us the probability that we are reading reference text r, given that we are reading word w. This quantity is the key to our *a priori* approach. Given a set of reference texts, the probability that an occurrence of word w implies that we are reading text r is

$$P_{wr} = \frac{F_{wr}}{\sum_r F_{wr}}. \quad \textbf{(1)}$$

As an example consider two reference texts, A and B. We observe that the word "choice" is used 10 times per 10,000 words in Text A and 30 times per 10,000 words in Text B. If we know simply that we are reading the word "choice" in one of the two reference texts, then there is a 0.25 probability that we are reading Text A and a 0.75 probability that we are reading Text B.

We can then use this matrix P_{wr} to produce a score for each word w on dimension d. This is the expected position on dimension d of any text we are reading, given only that we are reading word w, and is defined as

$$S_{wd} = \sum_r (P_{wr} \cdot A_{rd}). \quad \textbf{(2)}$$

In other words, S_{wd} is an average of the *a priori* reference text scores A_{rd}, weighted by the probabilities P_{wr}. Everything on the right-hand side of this expression may be either observed or (in the case of A_{rd}) assumed *a priori*. Note that if reference text r contains occurrences of word w and no other text contains word w, then $P_{wr} = 1$. If we are reading word w, then we conclude from this that we are certainly reading text r. In this event the score of word w on dimension d is the position of reference text r on dimension d: thus $S_{wd} = A_{rd}$. If all reference texts contain occurrences of word w at precisely equal frequencies, then reading word w leaves us none the wiser about which text we are reading and S_{wd} is the mean position of all reference texts.

To continue with our simple example, imagine that Reference Text A is assumed from independent sources to have a position of -1.0 on dimension d, and Reference Text B is assumed to have a position of $+1.0$. The score of the word "choice" is then

$$0.25(-1.0) + 0.75(1.0) = -0.25 + 0.75 = +0.5.$$

Given the pattern of word usage in the reference texts, if we knew only that the word "choice" occurs in some text, then this implies that the text's expected position on the dimension under investigation is +0.5. Of course we will update this expectation as we gather more information about the text under investigation by reading more words.

used with equal relative frequencies in all reference texts, they convey no useful information, but they do not systematically bias our results. Where such words are systematically used with unequal relative frequencies in reference texts, we assume that this is because they are conveying information about differences between texts.

Scoring Virgin Texts

Having calculated scores for all words in the word universe of the reference texts, the analysis of any set of virgin texts V of any size is very straightforward. First, we must compute the relative frequency of each virgin text word, as a proportion of the total number of words in the virgin text. We call this frequency F_{wv}. The score of any virgin text v on dimension d, S_{vd}, is then the mean dimension score of all of the scored words that it contains, weighted by the frequency of the scored words:

$$S_{vd} = \sum_w (F_{wv} \cdot S_{wd}). \quad \textbf{(3)}$$

This single numerical score represents the expected position of the virgin text on the *a priori* dimension under investigation. This inference is based on the assumption that the relative frequencies of word usage in the virgin texts are linked to policy positions in the same way as the relative frequencies of word usage in the reference texts. This is why the selection of appropriate reference texts—discussed at some length above—is such an important matter.

Interpreting Virgin Text Scores

Once raw estimates have been calculated for each virgin text, we need to interpret these in substantive terms, a matter that is not as straightforward as might seem at first sight. Because different texts draw upon the same word universe, relative word frequencies and hence word scores can never distinguish perfectly between texts. Words found in common to all or most of the reference texts hence tend to take as their scores the mean overall scores of the reference texts. The result is that, for any set of virgin texts containing the same set of nondiscriminating words found in the reference texts, the raw virgin text scores tend to be much more clustered together than the reference text scores. While the mean of the virgin scores will have a readily interpretable meaning (relative to the policy positions of the reference texts), the dispersion of the virgin text scores will be on a different scale—one that is much smaller. To compare the virgin scores directly with the reference scores, therefore, we need to transform the scores of the virgin texts so that they have same dispersion metric as the reference texts. For each virgin text v on a dimension d (where the total number of virgin texts $V > 1$), this is done as follows:

$$S^*_{vd} = (S_{vd} - S_{\bar{v}d})\left(\frac{SD_{rd}}{SD_{vd}}\right) + S_{\bar{v}d}, \quad \textbf{(4)}$$

where S_{vd} is the average score of the virgin texts, and the SD_{rd} and SD_{vd} are the sample standard deviations of the reference and virgin text scores, respectively. This preserves the mean and relative positions of the virgin scores but sets their variance equal to that of the reference texts. It is very important to note that this particular approach to rescaling is not fundamental to our word-scoring technique but, rather, is a matter of substantive research design unrelated to the validity of the raw virgin text scores. In our case we wish to express the estimated positions of the virgin texts on the same metric as the policy positions of the reference texts because we wish to compare the two sets of numbers to validate our technique. Further development to interpret raw virgin scores can and should be done, yet the simple transformation (Eq. 4) provides excellent results, as we demonstrate below. Other transformations are of course possible, for example, by analysts who wish to compare estimates derived from text analysis with policy positions estimated by other sources but expressed in some quite different metric. For these reasons we recommend that raw scores always be reported, in addition to any transformed values of virgin scores.

Estimating the Uncertainty of Text Scores

Our method for scoring a virgin text on some policy dimension generates a precise point estimate, but we have yet to consider any *uncertainty* associated with this estimate. No previous political science work estimating policy positions using quantitative content analysis deals systematically with the uncertainty of any estimate generated. The seminal and widely used CMP content analysis data, for example, are offered as point estimates with no associated measures of uncertinty. There is no way, when comparing the estimated positions of two manifestos using the CMP data, to determine how much the difference between estimates can be attributed to "real" differences and how much to coding unreliability.[6] Notwithstanding this, the time series of party policy positions generated by the CMP data has been seen in the profession as one of its great virtues, and "movements" of parties over time have typically been interpreted as real policy movements rather than as manifestations of coding unreliability.

Here we present a simple method for obtaining uncertainty estimates for our estimates of the policy positions of virgin texts. This allows us for the first time to make systematic judgments about the extent to which differences between the estimated policy positions of two texts are in fact significant.[7] Recall that each virgin text score S_{vd} is the weighted mean score of the words in text v on dimension d. If we can compute a mean for any set of quantities, then we can also compute a variance. In this context our interest is in how, for a given text, the scores S_{wd} of the words in the text vary around this mean. The variance of S_{wd} for a given text measures how dispersed the individual word scores are around the text's mean score. The less this variance, the more the words in the text all correspond to the final score and hence the lower our uncertainty about that score. Because the text's score S_{vd} is a weighted average, the variance we compute also needs to be weighted. We therefore compute V_{vd}, the variance of each word's score around the text's total score, weighted by the frequency of the scored word in the virgin text:

$$V_{vd} = \sum_{w} F_{wv}(S_{wd} - S_{vd})^2. \quad \mathbf{(5)}$$

This measure produces a familiar quantity directly analogous to the unweighted variance, summarizing the "consensus" of the scores of each word in the virgin text.[8] Intuitively, we can think of each scored word in a virgin text as generating an independent prediction of the text's overall policy position. When these predictions are tightly clustered, we are more confident in their consesus than when they are scattered more widely.

As with any variance, we can use the square root of V_{vd} to produce a standard deviation. This standard deviation can be used in turn, along with the total number of scored virgin words N^v, to generate a standard error $\sqrt{V_{vd}}/\sqrt{N^v}$ for each virgin text's score S_{vd}.[9] As we will see below, this standard error can then be used to perform standard statistical tests, such as the difference between means, to evaluate the significance of any difference in the estimated positions of two texts.[10]

[6] In large part this is because most manifestos in the data set were coded once only by a single coder, making it impossible to provide specific indications of inter- or intracoder reliability. The CMP has not yet published any test of intracoder reliability (Volkens 2001, 39). Intercoder reliability checks have been performed by correlating the frequency distribution of an "official" coding of a single standard text with the codings of hired researchers. The average correlation found for 39 "thoroughly trained" hired coders was 0.72, with correlations running as low as 0.34 (Volkens 2001, 39). Thus we can be certain that there is intercoder unreliability in the CMP data but have no precise way of knowing whether or not the difference between the estimated positions of two texts is statistically significant.

[7] Previous approaches to content analysis typically refer to *reliability*, but that is different from the notion of uncertainty we use here. Reliability refers to the stability of measures across repeated codings, as with the intercoder reliability of hand-coded content analysis. *Uncertainty* in our usage is consistent with the statistical notion of uncertainty, representing confidence that an estimate reflects the true position rather than variation due to chance or other uncontrollable factors, since we regard the generation of texts by political actors to be a stochastic process.

[8] Note that while we have employed the weighted formula here because our representation of words thus far has been as frequency distributions, this formula is equivalent to computing a population variance of the score of every (nonunique) word in the text. Each word hence contributes once for each time it occurs.

[9] This standard error applies to the raw virgin scores but not directly to the transformed scores. In the tables that follow (Tables 2–7), we also computed a standard error for the transformed scores along with 95% confidence intervals for the transformed scores, to make more straightforward the task of interpreting the uncertainty of the transformed scores on the original policy metric. The procedure for obtaining the upper and lower bounds of the transformed score confidence interval was straightforward. First, we computed the untransformed 95% confidence interval, calculated as the untransformed score S_{vd} plus and minus two standard errors (computed as explained in the text). These upper and lower confidence intervals, in the metric of the raw scores, were then transformed using exactly the same rescaling procedure as applied to the raw scores S_{vd}. The transformed standard error was then taken to be half of the distance between the transformed score and the bounds.

[10] We note that this measure is only one of a number of possible approaches to representing the uncertainty of our estimates of the positions of virgin texts and that numerous alternative measures can be developed to gauge the accuracy and robustness of final scores. In this introductory treatment of the word scoring method, we have deliberately chosen a form that will be familiar to most readers as well as being simple to compute. Diagnostic analysis of the word scoring technique is something to which we will return in future work.

TABLE 1. Word Scoring Example Applied to Artificial Texts

	Word Count						Probability of Reading Text r, Given Reading Word w					Score	Virgin Score		
	Reference Text					Virgin									
Word w	r_1	r_2	r_3	r_4	r_5	Text	P_{w1}	P_{w2}	P_{w3}	P_{w4}	P_{w5}	S_{wd}	F_{wv}	$F_{wv} * S_{wd}$	$F_{wv}(S_{wd} - S_{vd})^2$
A	**2**	**0**	**0**	**0**	**0**	**0**	1.00	0.00	0.00	0.00	0.00	−1.50	0.0000	0.0000	0.0000
B	**3**	**0**	**0**	**0**	**0**	**0**	1.00	0.00	0.00	0.00	0.00	−1.50	0.0000	0.0000	0.0000
C	**10**	**0**	**0**	**0**	**0**	**0**	1.00	0.00	0.00	0.00	0.00	−1.50	0.0000	0.0000	0.0000
D	**22**	**0**	**0**	**0**	**0**	**0**	1.00	0.00	0.00	0.00	0.00	−1.50	0.0000	0.0000	0.0000
E	**45**	**0**	**0**	**0**	**0**	**0**	1.00	0.00	0.00	0.00	0.00	−1.50	0.0000	0.0000	0.0000
F	**78**	**2**	**0**	**0**	**0**	**0**	0.98	0.03	0.00	0.00	0.00	−1.48	0.0000	0.0000	0.0000
G	**115**	**3**	**0**	**0**	**0**	**0**	0.97	0.03	0.00	0.00	0.00	−1.48	0.0000	0.0000	0.0000
H	**146**	**10**	**0**	**0**	**0**	**2**	0.94	0.06	0.00	0.00	0.00	−1.45	0.0020	−0.0029	0.0020
I	**158**	**22**	**0**	**0**	**0**	**3**	0.88	0.12	0.00	0.00	0.00	−1.41	0.0030	−0.0042	0.0028
J	**146**	**45**	**0**	**0**	**0**	**10**	0.76	0.24	0.00	0.00	0.00	−1.32	0.0100	−0.0132	0.0077
K	**115**	**78**	**2**	**0**	**0**	**22**	0.59	0.40	0.01	0.00	0.00	−1.18	0.0220	−0.0261	0.0119
L	**78**	**115**	**3**	**0**	**0**	**45**	0.40	0.59	0.02	0.00	0.00	−1.04	0.0450	−0.0467	0.0156
M	**45**	**146**	**10**	**0**	**0**	**78**	0.22	0.73	0.05	0.00	0.00	−0.88	0.0780	−0.0687	0.0146
N	**22**	**158**	**22**	**0**	**0**	**115**	0.11	0.78	0.11	0.00	0.00	−0.75	0.1150	−0.0863	0.0105
O	**10**	**146**	**45**	**0**	**0**	**146**	0.05	0.73	0.22	0.00	0.00	−0.62	0.1460	−0.0904	0.0043
P	**3**	**115**	**78**	**2**	**0**	**158**	0.02	0.58	0.39	0.01	0.00	−0.45	0.1580	−0.0712	0.0000
Q	**2**	**78**	**115**	**3**	**0**	**146**	0.01	0.39	0.58	0.02	0.00	−0.30	0.1460	−0.0437	0.0032
R	**0**	**45**	**146**	**10**	**0**	**115**	0.00	0.22	0.73	0.05	0.00	−0.13	0.1150	−0.0150	0.0116
S	**0**	**22**	**158**	**22**	**0**	**78**	0.00	0.11	0.78	0.11	0.00	0.00	0.0780	0.0000	0.0157
T	**0**	**10**	**146**	**45**	**0**	**45**	0.00	0.05	0.73	0.22	0.00	0.13	0.0450	0.0059	0.0151
U	**0**	**3**	**115**	**78**	**2**	**22**	0.00	0.02	0.58	0.39	0.01	0.30	0.0220	0.0066	0.0123
V	**0**	**2**	**78**	**115**	**3**	**10**	0.00	0.01	0.39	0.58	0.02	0.45	0.0100	0.0045	0.0081
W	**0**	**0**	**45**	**146**	**10**	**3**	0.00	0.00	0.22	0.73	0.05	0.62	0.0030	0.0019	0.0034
X	**0**	**0**	**22**	**158**	**22**	**2**	0.00	0.00	0.11	0.78	0.11	0.75	0.0020	0.0015	0.0029
Y	**0**	**0**	**10**	**146**	**45**	**0**	0.00	0.00	0.05	0.73	0.22	0.88	0.0000	0.0000	0.0000
Z	**0**	**0**	**3**	**115**	**78**	**0**	0.00	0.00	0.02	0.59	0.40	1.04	0.0000	0.0000	0.0000
AA	**0**	**0**	**2**	**78**	**115**	**0**	0.00	0.00	0.01	0.40	0.59	1.18	0.0000	0.0000	0.0000
BB	**0**	**0**	**0**	**45**	**146**	**0**	0.00	0.00	0.00	0.24	0.76	1.32	0.0000	0.0000	0.0000
CC	**0**	**0**	**0**	**22**	**158**	**0**	0.00	0.00	0.00	0.12	0.88	1.41	0.0000	0.0000	0.0000
DD	**0**	**0**	**0**	**10**	**146**	**0**	0.00	0.00	0.00	0.06	0.94	1.45	0.0000	0.0000	0.0000
EE	**0**	**0**	**0**	**3**	**115**	**0**	0.00	0.00	0.00	0.03	0.97	1.48	0.0000	0.0000	0.0000
FF	**0**	**0**	**0**	**2**	**78**	**0**	0.00	0.00	0.00	0.03	0.98	1.48	0.0000	0.0000	0.0000
GG	**0**	**0**	**0**	**0**	**45**	**0**	0.00	0.00	0.00	0.00	1.00	1.50	0.0000	0.0000	0.0000
HH	**0**	**0**	**0**	**0**	**22**	**0**	0.00	0.00	0.00	0.00	1.00	1.50	0.0000	0.0000	0.0000
II	**0**	**0**	**0**	**0**	**10**	**0**	0.00	0.00	0.00	0.00	1.00	1.50	0.0000	0.0000	0.0000
JJ	**0**	**0**	**0**	**0**	**3**	**0**	0.00	0.00	0.00	0.00	1.00	1.50	0.0000	0.0000	0.0000
KK	**0**	**0**	**0**	**0**	**2**	**0**	0.00	0.00	0.00	0.00	1.00	1.50	0.0000	0.0000	0.0000
Total	1,000	1,000	1,000	1,000	1,000	1,000							1.00	**−0.45**	**0.14**
	−1.50	−0.75	0.00	0.75	1.50	**−0.45**									
							A priori positions of reference texts								
							Estimated score for virgin text S_{vd}							**−0.45**	
							Estimated weighted variance V_{vd}							**0.14**	
							Estimated SD $\sqrt{V_{vd}}$							**0.38**	
							Estimated SE $\sqrt{V_{vd}}/\sqrt{1000}$							**0.018**	

Illustration Using a Sample Text

The method we have outlined can be illustrated by working though the calculation of word scores on an artificial text. Table 1 shows the results of analyzing a very simple hypothetical data set, shown in columns 2–7 in the table (in bold face), containing word counts for 37 different words observed in five reference texts, $r_1 - r_5$, as well as counts for the same set of words in a hypothetical "virgin" text whose position we wish to estimate. The policy positions of the reference texts on the dimension under investigation are estimated or assumed *a priori* and are shown at the bottom of the table as ranging between −1.50 and +1.50. Table 1 shows that, in this hypothetical data set, nearly all words can be ranked from left to right in terms of the extent to which they are associated with left- or right-wing parties. Within each individual text, the observed pattern of word frequencies fits a normal distribution. We also indicate the "real" position of the virgin text, which is unknown to the hypothetical analyst but which we know to be −0.45. This is the essential quantity to be estimated by comparing the distribution of the word frequencies in the virgin texts with that in the reference texts.

The columns headed P_{w1}–P_{w5} show the conditional probabilities (Eq. 1) necessary for computing word scores from the reference texts—this is the matrix of probabilities that we are reading reference text r given that we are reading word w. Combined with the *a priori* positions of the reference texts, these allow us to calculate scores, S_w, for each word in the word universe of the reference texts (Eq. 2). These scores are then used to score the virgin text by summing the scores of words used in the virgin text, weighting each score by the relative frequency of the word in question (Eq. 3). The resulting estimate, and its associated uncertainty measure, is provided at the bottom right of Table 1, together with its associated standard error. From this we can see that, in this perfectly behaved data set, our technique perfectly retrieves the position of the virgin text under investigation.

While this simple example illustrates the calculations associated with our technique, it of course in no way shows its efficacy with real-world data, in which there will be much more heavily overlapping patterns of word usage in reference texts, large numbers of very infrequently used words, volumes of words found in virgin texts that do not appear in reference texts and therefore cannot be scored, and so on. The true test of the technique we propose lies in applying it to texts produced by real-world political actors, to see if we can reproduce estimates of their policy positions that have been generated by more traditional means.

ESTIMATING ECONOMIC POLICY POSITIONS OF BRITISH AND IRISH PARTIES

We now test our technique using real-world texts, by attempting to replicate previously published findings on the policy positions of political parties in Britain and Ireland. We compare our own findings with three sets of independent estimates of the economic policy positions of British and Irish political parties at the time of the 1997 general elections in each country. These are the results of 1997 expert surveys of party policy positions (Laver 1998 a, b) and of the hand-coding and deterministic computer-coding of 1997 party manifestos (Laver and Garry 2000).

British Party Positions on Economic Policy

The first task is to calculate word scores on the economic policy dimensions for British party manifestos in the 1990s. We selected the 1992 British Labour, Conservative, and LD party manifestos as reference texts. For independent estimates of the economic policy positions of these manifestos, we use the results of an expert survey of the policy positions of the parties that wrote them, on the scale "increase public services vs. cut taxes," reported in Laver and Hunt 1992.[11] The first stages in the analysis are to observe frequency counts for all words used in these reference texts[12] and to calculate relative word frequencies from these.[13] Using these relative frequencies and the reference text policy positions, we then calculated a word score on the economic policy dimension for every word used in the reference texts, using the procedures outlined above (Eqs. 1 and 2).

Having calculated word scores on the economic policy dimension for each of the 5,299 words used in the 1992 reference texts, we use these to estimate the positions of three "virgin" texts. These are the Labour, LD, and Conservative manifestos of 1997. Note that this is a tough substantive test for our technique. Most commentators, backed up by a range of independent estimates, suggest that the ordering of the economic policy positions of the British parties changed between the 1992 and the 1997 elections, with Labour and the LDs exchanging places, leaving Labour in the center and the LDs on the left in 1997. This can be seen in 1997 expert survey findings (Laver 1998a) that we set out to replicate using computer word scoring, reported in the third row of the top panel in Table 2. We are particularly interested to see whether our technique can pick up this unusual and significant movement.

We can only score virgin texts on the words that they share with the universe of reference texts. The 1997 British manifestos used a total of 1,573 words that did not appear in the 1992 texts and these could not be scored.[14] We thus applied the word scores derived from

[11] It is very important to note that such expert survey estimates are convenient to use as reference scores in this context but are not in any way intrinsic to our technique. What we require are independent estimates of, or assumptions about, the positions of the reference texts in which we can feel confident. The expert survey scores we use are reported in the first row in the lower half in Table 2. Both in terms of their face validity and because these scores report the mean judgments of a large number of British political scientists, we consider these estimated positions of the reference texts to represent a widely accepted view of the of the British policy space in 1992.

[12] While, for reasons discussed above, we included every single word used in the 1992 manifestos, even common words without substantive political meaning such as "a" and "the," we did exclude all "non-words," which we took to be character strings not beginning with letters.

[13] Any computer-coded content analysis software (for example, Textpack) can perform simple word counting. To process large numbers of texts simultaneously and quickly perform all subsequent calculations on the output, however, we wrote our own software. Easy-to-use software—entitled WORDSCORES—for implementing the methods described in this paper is freely available from `http://www.politics.tcd.ie/wordscores/`. A full replication data set for this paper, using the WORDSCORES software, is also available at that web site. Installation or updating of WORDSCORES can be accomplished by any computer connected to the Internet by executing a single command from within the Stata statistical package: `net install http://www.politics.tcd.ie/wordscores/wordscores`. Version information prior to installation can be obtained by executing the Stata command `net describe http://www.politics.tcd.ie/wordscores/wordscores`.

[14] Most of the 1997 words not used in 1992 were used very infrequently, with a median occurrence of 1 and a mean occurrence of between 1.2 and 1.9 (see Table 2). For this reason they would have contributed very little weight to the virgin text scores. Overall for

TABLE 2. Raw and Standardized Estimated Economic Policy Positions of 1997 British Party Manifestos

Party	Liberal Democrat	Labour	Conservative	Mean Absolute Difference
Estimates				
1997 transformed virgin text scores	**5.00**	**9.17**	**17.18**	
SE	0.363	0.351	0.325	
1997 expert survey	**5.77**	**10.30**	**15.05**	
SE ($n = 117$)	0.234	0.229	0.227	
1997 standardized comparison scores				
Word scores	−0.88	−0.21	1.09	**0.13**
Expert survey	−0.99	−0.02	1.01	—
Hand-coded content analysis	−0.83	−0.28	1.11	**0.17**
Dictionary-based computer-coding	−1.08	0.18	0.90	**0.13**
Raw data				
1992 reference texts				
A priori positions	**8.21**	**5.35**	**17.21**	
SE ($n = 34$)	0.425	0.377	0.396	
Length in words	17,077	11,208	28,391	
No. of unique words	2,911	2,292	3,786	
1997 virgin texts				
Raw mean word scores (S_{vd})	10.2181	10.3954	10.7361	
SE	0.015	0.015	0.014	
Length in words	13,709	17,237	20,442	
Unique words scored	1,915	2,211	2,279	
% words scored	94.9	96.2	95.5	
Unique unscorable words	423	697	714	
Mean frequency of unscorable words	1.23	1.26	1.29	

Sources: A priori positions 1992 (Laver and Hunt 1992); expert survey scores 1997 (Laver 1998a); hand-coded content analysis and deterministic computer-coding (Laver and Garry 2000).
Note: Standardized scores are reported raw scores for 1997 standardized within each data source. For hand- and deterministic computer-codings, these have been recalculated to facilitate comparison from data presented by Laver and Garry (2000), who standardized their raw score across all observations for Britain and Ireland. The mean absolute difference reports the mean of the absolute differences for the three parties between the standardized party scores for each text-based estimate and the standardized expert survey party score. Standard errors are computed as described in the text.

the 1992 reference texts to the 1997 manifestos, calculating a "raw" score for each of the three manifestos (Eq. 3) and transforming (Eq. 4) it in the way described above. Finally, we calculate the standard errors of our estimates (Eq. 5 and associated discussion).

The key results of this analysis are presented in the top panel in Table 2. The first row reports our estimated positions of the 1997 party manifestors, transformed to the same metric as the 1992 expert survey scores that were used as points of reference. Our first point of comparison is with a set of 1997 expert survey scores, expressed in the same metric, highlighting the shift of the Labour Party to the center of this policy dimension (Laver 1998a). These scores are reported in the third row in Table 2. The comparison is very gratifying. Our word-scored estimates clearly pick up the switch in Labour and LD economic policy positions and are remarkably close, considering that they derive from an utterly independent source, to the expert survey estimates for 1997. Note particularly that the word scores we used were calculated from 1992 reference positions that locate the LDs between Labour and the Conservatives on economic policy, so that it was simply the changing relative frequencies of word use between the 1992 and the 1997 manifestos that caused the estimated positions of these two parties to reverse, in line with independent estimates.

Table 2 also reports the standard errors associated with our raw estimates, from which we can conclude that differences among the estimated economic policy positions of the three manifestos are statistically significant. Note that this availability of standard errors, allowing such judgments to be made, is unique among published estimates of policy positions based on the content analysis of political texts.

To compare our results with those generated by other content analysis techniques, the last four rows in the top panel in Table 2 report, in addition to our own estimates and those of the 1997 expert survey, two other text-based estimates of the 1997 economic policy positions of the British parties. One of these derives from hand-coded content analysis, and the other from dictionary-based computer-coding, of the 1997 manifestos that we have treated here as virgin texts (both reported in Laver and Garry 2000). Since different published sets of scores had different metrics, all scores have been standardized to facilitate comparison.[15] The main

the 1997 virgin texts, the bottom panel in Table 2 shows that the percentages of virgin words scoreable were 96.2%, 94.9%, and 95.5% for the LDs, Labour, and the Conservatives, respectively.

[15] All sets of standardized estimates in Table 2 have been standardized within country and time period in the tables that follow, to

substantive difference between different estimates of British party positions in 1997 concerns the placement of the Labour Party. All scales locate Labour between the LDs and the Conservatives. The dictionary-based scale places Labour closer to the Conservatives, and the other text-based scales place Labour closer to the LDs, while the independent expert survey locates Labour midway between the two other parties.

As a summary of the fit between the various text-based estimates of party positions and the expert survey, the final column in the top panel in Table 2 reports the mean absolute difference between the estimated positions of the parties on each standardized scale and the positions of the same parties in the expert survey. This confirms our *prima facie* impression that our word-scored estimates are somewhat closer than the hand-coded content analysis to the expert survey estimates (representing the consensus among British political scientists about British party positions in 1997) and are about as close to these as the more traditional dictionary-based computer-coded scale. This is a remarkable achievement considering that, in stark contrast to all other methods, our word scoring technique treats words as data without reading or understanding them in any way, uses no knowledge of English, and does not require a predetermined computer-coding dictionary when analyzing the texts.

Irish Party Positions on Economic Policy

We now report a similar analysis for the Irish party system. As our reference texts for Irish politics in the 1990s, we take the manifestos of the five main parties contesting the 1992 election—Fianna Fáil, Fine Gael, Labour, the Progressive Democrats (PDs), and the Democratic Left (DL). For our independent estimate of the positions of these reference texts, we use an expert survey taken at the time of the 1992 Irish election (Laver 1994). Having used these data in a preliminary analysis to calculate word scores for the economic policy dimension in Ireland in the 1990s, we then analyze 1997 Irish party manifestos as virgin texts. Our aim is once more to replicate independent published estimates of Irish party policy positions in 1997—the results of an expert survey conducted at the time of the 1997 election (Laver 1998b), as well as estimates based on hand-coded content analysis and dictionary-based computer-coding (Laver and Garry 2000). The results of this analysis are listed in Table 3, which has the same format as Table 2.

facilitate comparison of estimates originally reported using different units of analysis. (Thus the 1997 British estimates, for example, are standardized against themselves.) This differs from the practice adopted by Laver and Garry (2000), who standardized across both countries and time periods. This was because they were evaluating the application of a single expert coding scheme and computer-coding dictionary to all observations. In contrast, we use the 1992 manifestos to generate separate sets of words scores for Britain and Ireland and apply these separately to virgin texts taken from subsequent time periods in each country. The standardized figures in Tables 2–5 thus differ from those reported by Laver and Garry (2000) but are calculated directly from them.

Substantively, while nothing as dramatic happened in Ireland between 1992 and 1997 as the vaunted dash to the center by the British Labour Party under Tony Blair, there was a major coalition realignment that we expect to show up in the economic policy positions of the parties. The government that formed immediately after the 1992 election was the first-ever coalition between Fianna Fáil and the Labour Party. As the bottom panel in Table 3 shows, these parties were judged by expert survey respondents in 1992 to be adjacent, though by no means close, on the economic policy dimension. This government fell in 1994 and was replaced without an intervening election by a "rainbow" coalition of Fine Gael, Labour, and DL—so called because of major policy differences among what was essentially a coalition of Fianna Fáil's opponents. By the time of the 1997 election, the three parties of the Rainbow Coalition presented a common front to the electorate and sought reelection. While promoting independent policy positions, they were nonetheless careful to ensure that their respective party manifestos did not contain major policy differences that would embarrass them on the campaign trail. Confronting the Rainbow Coalition at the election, Fianna Fáil and the PDs formed a pact of their own, promising to go into government together if they received enough support and, also, taking care to clean up any major policy incompatibilities in their respective manifestos that would have been exploited by opponents during the campaign. The 1997 election was thus fought between two rival coalitions—the Fine Gael, Labour, and DL rainbow, on one side, and Fianna Fáil and the PDs, on the other—who published independent but coordinated policy programs.

The top panel in Table 3 shows that the main manifestation of these changes in expert survey data is a collective judgment that Fine Gael Shifted to the left in 1997 as a result of its membership in the Rainbow Coalition with Labour and DL. The experts did not consider Fianna Fáil to have shifted right, despite the fact that the 1997 Fianna Fáil manifesto was designed not to conflict with that of the PDs and that immediately after the election Fianna Fáil agreed to a joint program of government with the right-wing PDs, subsequently governing harmoniously with them for the first full-term coalition government in the history of the Irish state. This is intriguing because, as the last four lines in the top panel in Table 3 show, both expert survey and hand-coded content analyses continue to show Fine Gael to the right of Fianna Fáil in 1997, while both dictionary-based computer-coding and our own word scoring techniques, which proceeded without expert intervention, find Fine Gael to the left of Fianna Fáil. Both sets of computer-coded results reflect the pattern of actual coalitions in the legislature, so we may speculate here that we are seeing signs of experts—whether survey respondents or human text coders—reading between the lines of the published texts and inferring that, in a coalition environment such as this, stated policy positions are not entirely sincere.

Be that as it may, the results in Table 3 show that our approach, while generating results with good face

TABLE 3. Raw and Standardized Estimated Economic Policy Positions of 1997 Irish Party Manifestos

	DL	Labour	Fianna Fáil	Fine Gael	PD	Mean Absolute Difference
Estimates						
1997 transformed virgin text scores	**3.79**	**6.78**	**15.32**	**13.18**	**16.44**	
SE	1.908	0.503	0.461	0.593	0.797	
1997 expert survey	**5.47**	**7.77**	**12.07**	**12.30**	**17.27**	
SE ($n=30$)	0.325	0.330	0.398	0.363	0.310	
1997 standardized comparison scores						
Word scores	−1.32	−0.78	0.79	0.37	0.94	**0.27**
Expert survey	−1.21	−0.70	0.24	0.29	1.38	
Hand-coded content analysis	−1.10	−0.72	−0.02	0.38	1.46	**0.11**
Deterministic computer-coding	−1.22	−0.52	0.36	−0.06	1.45	**0.15**
Raw data						
1992 reference texts						
A priori positions	**4.50**	**6.88**	**13.13**	**15.00**	**17.63**	
SE ($n=28$)	0.40	0.37	0.57	0.47	0.30	
Length in words	6,437	16,373	3,782	3,679	3,523	
No. of unique words	1,763	2,768	1,186	1,019	1,136	
1997 virgin texts						
Raw mean word scores (S_{vd})	10.9205	10.9954	11.2087	11.1552	11.2367	
SE	0.048	0.013	0.012	0.015	0.020	
Length in words	2,549	32,171	38,659	24,026	13,922	
Unique words scored	748	2,348	2,609	2,098	1,721	
% words scored	92.4	92.4	89.7	92.1	92.9	
Unique unscorable words	172	1,492	2,203	1,902	991	
Mean frequency of unscorable words	1.13	1.64	1.82	1.59	1.13	

Sources: A priori positions 1992 (Laver 1994); expert survey 1997 (Laver 1998b); expert-coded content analysis and deterministic compute coding (Laver and Garry 2000).
Note: See Note to Table 2.

validity in terms of subsequent coalition alignments, does not correspond as well as the other text-based techniques with expert survey. The key difference between our scale and the others is the convergence of Fianna Fáil and the PDs indicated by our technique, followed as we have seen by a coalition between the two parties. While this convergence is substantively plausible, an alternative possibility is that our estimates are less accurate than the others in this case.

One possible source of such a problem is that the 1997 Irish manifestos were on average considerably longer than their short 1992 progenitors, using many words that were not used in 1992. The Fianna Fáil manifesto, in particular, burgeoned dramatically in length. We scored the 4,279 different words in the 1992 manifestos, but a total of 4,188 new words appeared in 1997, albeit many of them only once.[16] There was thus much less overlap than in Britain between the word pools used in 1992 and 1997, leaving more of the 1997 Irish manifestos necessarily unscored. This is reflected in noticeably higher standard errors for our Irish estimates than for the British ones. The short DL manifesto in 1997, for example, generates a word-scored estimated economic policy position of 3.79 on the 1–20 metric of the expert survey with which it is being compared, but the very high associated standard error tells us that this position might be anything from 0.0 to 7.6 on this scale (its 95% confidence interval). The PD manifesto has a standard error that implies that we cannot statistically distinguish its economic policy position from that of Fianna Fáil. In other words, the standard errors generated by the word scoring technique are telling us that we should not feel as confident with its estimates for Ireland as we feel with those for Britain. We consider this to be an interesting and important result in itself—bearing in mind that all previous content analysis policy estimates of which we are aware report point estimates with no estimate whatsoever of associated error and, thus, are effectively blind to the potential problems arising from short texts we have diagnosed in the Irish case.

ESTIMATING THE POLICY POSITIONS OF BRITISH AND IRISH PARTIES ON THE LIBERAL–CONSERVATIVE SOCIAL POLICY DIMENSION

A range of techniques has been used to estimate economic policy positions in Britain and Ireland and has been found to have good face validity. When setting out to cross-validate economic policy estimates produced by our word scoring method, therefore, we are working in well-explored territory. We turn now to a more difficult and interesting problem. This is the estimation of policy positions on the "liberal–conservative" dimension of social policy, taken as the second most important

[16] The Fianna Fáil manifesto in 1997 contained more than 10 times as many total words as the 1992 manifesto. Because the pool of reference texts included manifestos from four other parties, however, we were able to score 89.7% of the words in the 1997 manifesto (see Table 3). Results for the other virgin texts were all above 92% words scored.

TABLE 4. Raw and Standardized Estimated Social Policy Positions of 1997 British Party Manifestos

	Liberal Democrat	Labour	Conservative	Mean Absolute Difference
Estimates				
1997 transformed virgin text scores	**5.17**	**8.96**	**15.06**	
SE	0.285	0.272	0.254	
1997 expert survey	**6.75**	**8.28**	**13.26**	
SE ($n = 116$)	0.240	0.228	0.253	
1997 standardized comparison scores				
Word scores	−0.91	−0.15	1.07	**0.12**
Expert survey	−0.79	−0.34	1.13	
Hand-coded content analysis	−1.07	−0.15	0.91	**0.33**
Deterministic computer-coding	−1.06	−0.12	0.93	**0.31**
Raw data				
1992 reference texts				
A priori positions	**6.87**	**6.53**	**15.34**	
SE ($n = 34$)	0.410	0.358	0.451	
1997 virgin texts				
Raw mean word scores (S_{vd})	9.5285	9.6956	9.9649	
SE	0.013	0.012	0.011	

Note: Sources as in Table 2. All statistics for the word counts and frequencies of reference and virgin texts are the same as in Table 2.

dimension of competition in many European party systems, a general perception for which Warwick (2002) found support when extracting common policy spaces from party manifesto and expert survey data.

Traditional techniques of content analysis have been very much less effective at providing reliable and stable estimates of policy positions on this dimension, a conclusion confirmed in a careful study by McDonald and Mendes (2001). Having found a number of economic policy scales to be highly reliable, they found the reliability of content analysis-based social policy scales to be "not so filled with noise as to be completely unreliable" but "below a . . . reliability that we would take as minimally acceptable" (McDonald and Mendes 2001, 111).

In applying our word scoring approach to a new policy dimension, we also reveal one of its chief advantages of flexibility, ease of use, and susceptibility to tests using different *a priori* conditions. Once the reference texts have been converted into the matrix of word probabilities P_{wr}, it is straightforward to compute word scores for a new dimension d' simply by changing the *a priori* set of reference scores to $A_{rd'}$. We can then very easily apply these new word scores to the virgin texts and thereby estimate their positions on d', which in most cases takes under one second of computing time. In contrast to other computer-coding techniques, there is no need for the labor-intensive development and testing of a new coding dictionary for each new policy dimension considered. We demonstrate this by rerunning the analysis for the social policy dimension in Britain and Ireland in a manner identical to that for economic policy, except that the reference scores were taken from expert survey estimates of the social policy positions of the authors of these reference texts (Laver 1994; Laver and Hunt 1992). The social policy positions we estimate are defined *a priori* in terms of promoting liberal policies on matters such as abortion and homosexuality, at one end, and opposing such policies, at the other.

British Party Positions on Social Policy

The results of rescoring of the 1997 virgin texts for Britain are reported in Table 4, which has the same format as Table 2 without repeating raw data unnecessarily. As before, we begin by comparing our estimates with those generated by the completely independent expert survey conducted at the time of the 1997 election. Substantively, the main party movement reported by the expert surveys is a shift from estimates in 1992 that found the social policy positions of Labour and the LDs to be statistically indistinguishable, to one in 1997 in which Labour occupied a statistically distinct position on the conservative side of the LDs. This finding is clearly replicated by our word-scored estimates.

As before, the last four rows in the top panel in Table 4 compare standardized estimates from our word scoring method with those derived from the 1997 expert survey, as well as both hand- and dictionary-based computer-coded content analyses of the 1997 manifestos. These results, summarized by the mean absolute differences, show that computer word scoring performs extraordinarily well in this previously troublesome area, far better than any other content analysis technique. Substantively this is because, according to the expert survey that summarizes the judgments of British political scientists on this matter, the situation in 1997 was one in which Labour and the LDs were relatively close to each other in the more liberal half of the social policy dimension, with the Conservatives firmly on the right. This configuration is retrieved from the 1997 manifestos by our language-blind word scoring technique—it can be seen in the negative standard scores for the Labour Party. The more traditional techniques of content analysis, whether hand- or computer-coded, place Labour much closer to the Conservatives on social policy than to the LDs, a finding that does not seem to have good face validity.

The mean absolute differences between the results of the various content analyses and those of the expert survey show that our word scoring technique did as well on the liberal–conservative dimension in Britain as it did for economic policy. What is striking, however, is that it did distinctly better than more traditional text analysis techniques in what has previously been a very problematic area for content analysis.

Irish Party Positions on Social Policy

We reran the analysis in the same way to estimate the social policy positions of the 1997 Irish party manifestos, treating these as virgin texts. The results are reported in Table 5. The most important substantive pattern to watch for in the Irish case is the relative position of Fianna Fáil and the PDs. Since the PDs are regarded by many as a classical liberal party, their right-wing economic policy position is widely perceived to be combined with a relatively leftist position on social issues. As Table 5 shows, this received wisdom is reflected in expert survey estimates. Fianna Fáil, in contrast, is typically seen as the guardian of traditional Catholic social values in Ireland. This pattern can be seen clearly in the expert surveys, which place Fianna Fáil very firmly on the right of the liberal–conservative dimension of social policy.

In contrast to the situation in Britain, therefore, the relative positions of parties on the liberal–conservative social policy dimension in Ireland differ in important substantive ways from those on the economic policy dimension. The top row in Table 5 shows that our language-blind word scoring techniques picks this difference up very well, coming close to the 1997 expert survey results in its analysis of the 1997 manifestos as virgin texts. As the last four rows in the top panel in Table 5 show, the more traditional content analysis techniques cannot replicate independent estimates of the social policy position of the PDs, (mis)placing the PDs, with high positive standard scores, on the conservative side of the social policy dimension at a position much more conservative than that of Fianna Fáil. This neither corresponds to the consensus of political scientists reflected in the expert judgments nor has good face validity.

The mean absolute differences again summarize the relative performance of the three content analysis techniques. These show that our word scoring technique, despite the fact that it uses no knowledge of the English language, performs strikingly better than the other content analysis techniques, performing remarkably well on a dimension that has previously presented content analysts with considerable problems.

Overall Fit with Expert Surveys

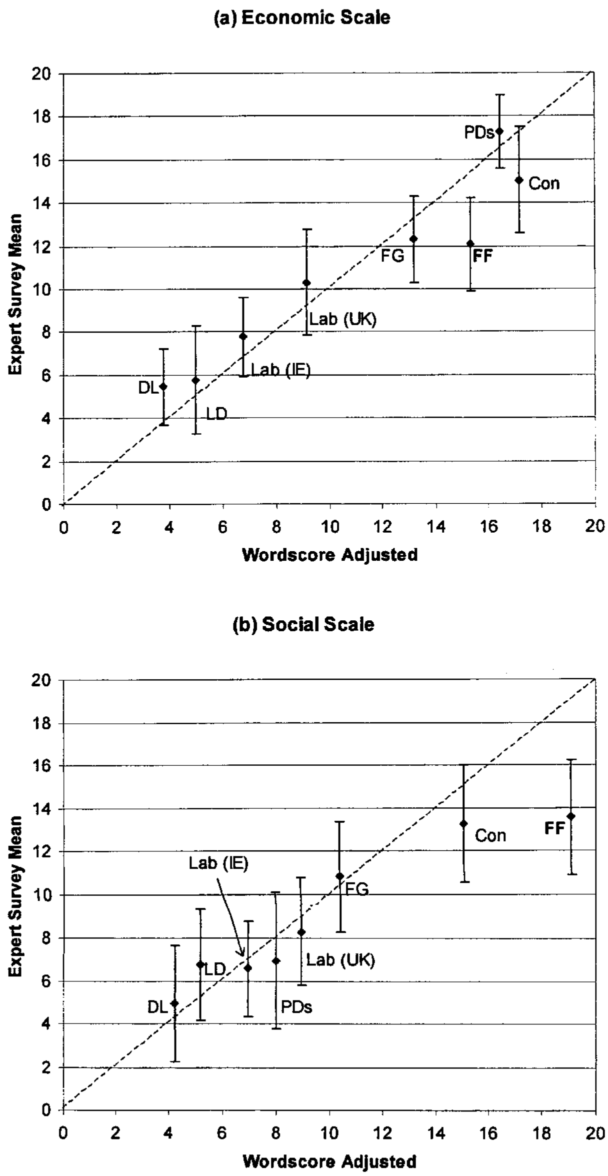
Figure 2 summarizes the fit between independent expert survey findings and our rescaled estimates of the policy positions of virgin texts, using computer word scoring. The X axis gives the word-scored estimates for 1997 virgin texts; the Y axis, expert survey estimates for 1997 of the positions of the authors of those texts. The vertical bars on each point represent a single standard deviation among the expert survey results. These bars may be interpreted as the range within which a single standard deviation of experts ranked the party on each scale. Where this bar crosses the vertical line of perfect correspondence, it indicates that approximately the middle 65% of the experts surveyed could easily have chosen the policy position estimated by the word scoring procedure. Of all of the texts we analyzed, on two policy dimensions, the only text for which word-scored estimates were more than a single standard deviation away from expert survey results was the Fianna Fáil manifesto in Ireland. And this difference, as we have argued, could possibly have been the result of

TABLE 5. Raw and Standardized Estimated Social Policy Positions of 1997 Irish Party Manifestos

	DL	Labour	Fianna Fáil	Fine Gael	PD	Mean Absolute Deviation
Estimates						
1997 transformed virgin text scores	**4.23**	**6.96**	**19.07**	**10.37**	**8.01**	
SE	1.178	0.319	0.339	0.378	0.474	
1997 expert survey	**4.97**	**6.57**	**13.55**	**10.82**	**6.93**	
SE ($n = 30$)	0.495	0.405	0.491	0.467	0.577	
1997 standardized comparison scores						
Word scores	−0.97	−0.49	1.65	0.12	−0.31	**0.21**
Expert survey	−1.02	−0.57	1.42	0.64	−0.47	
Hand-coded content analysis	−1.31	−0.62	0.09	1.23	0.62	**0.67**
Deterministic computer-coding	−1.07	−1.02	0.75	0.25	1.09	**0.62**
Raw data						
1992 reference texts						
A priori positions	**3.50**	**6.00**	**17.50**	**13.71**	**9.43**	
SE ($n = 28$)	0.416	0.404	0.391	0.554	0.809	
1997 virgin texts						
Raw mean word scores (S_{vd})	9.4960	9.6098	10.1157	9.7523	9.6537	
SE	0.049	0.013	0.014	0.016	0.020	

Note: Sources as in Table 3. Word statistics and counts as in Table 3.

FIGURE 2. Agreement Between Word Score Estimates and Expert Survey Results, Ireland and United Kingdom, 1997, for (a) Economic and (b) Social Scales



Note: The diagonal dashed line shows the axis of perfect agreement. Vertical bars represent one standard deviation of the expert scores (Ireland, $N=30$; UK, $N=117$).

contextual judgments made by experts about the "real" position of Fianna Fáil, rather than of error in the computer analysis of the actual text of the party manifesto. Put in a slightly different way, the technique we propose performed, in just about every case, equivalently to a typical expert—which we take to be a clear confirmation of the external validity of our technique's ability to extract meaningful estimates of policy positions from political texts.

CODING NON-ENGLISH-LANGUAGE TEXTS

Thus far we have been coding English-language texts, but since our approach is language-blind it should work equally well in other languages. We now apply it to German-language texts, analyzing these using no knowledge of German. Our research design is essentially similar to that we used for Britain and Ireland. As reference texts for Germany in the 1990s, we take the 1990 manifestos of four German political parties—the Greens, Social Democratic Party (SPD), Christian Democrats (CDU), and Free Democrats (FDP). Our estimates of the *a priori* positions of these texts on economic and social policy dimensions derive from an expert survey conducted in 1989 by Laver and Hunt (1992). Having calculated German word scores for both economic and social policy dimensions in precisely the same way as before, we move on to analyze six virgin texts. These are the manifestos of the same four parties in 1994, as well as manifestos for the former Communists (PDS) in both 1990 and 1994. Since no expert survey scores were collected for the PDS in 1990, or for any German party in 1994, we are forced to rely in our evaluation upon the face validity of our estimated policy positions for the virgin texts. However, the corpus of virgin texts presents us with an interesting and taxing new challenge. This is to locate the PDS on both economic and social policy dimensions, even though no PDS reference text was used to calculate the German word scores. We are thus using German word scores, calculated using no knowledge of German, to locate the policy positions of the PDS, using no information whatsoever about the PDS other than the words in its manifestos, which we did not and indeed could not read ourselves. The top panel in Table 6 summarizes the results of our analysis.

The first row in Table 6 reports our rescaled computer estimates of the economic policy positions of the six virgin texts. The main substantive pattern for the economic policy dimension is a drift of all established parties to the right, with a sharp rightward shift by the SDP. Though this party remains between the position of the Greens and that of the CDU, it has moved to a position significantly closer to the CDU. The face validity of this seems very plausible. Our estimated economic policy positions of the 1990 and 1994 PDS manifestos locate these firmly on the left of the manifestos of the other four parties, which has excellent face validity. The rescaled standard errors show that the PDS is indeed significantly to the left of the other parties but that there is no statistically significant difference between the 1990 and the 1994 PDS manifestos. In other words, using only word scores derived from the other four party manifestos in 1990 and no knowledge of German, the manifestos of the former Communists were estimated in both 1990 and 1994 to be on the far left of the German party system. We consider this to be an extraordinarily good result for our technique.

The third row in Table 6 reports our estimates of the social policy positions of the virgin texts. As in the

TABLE 6. Estimated Economic and Social Policy Positions of German Party Manifestos, 1990–94

Party	1990 PDS	1994 PDS	1994 Green	1994 SDP	1994 CDU	1994 FDP
Estimates						
1994 transformed economic policy virgin text scores	**4.19**	**3.98**	**7.47**	**10.70**	**13.67**	**17.15**
SE	0.436	0.511	0.259	0.365	0.391	0.220
1994 transformed social policy virgin text scores	**1.16**	**1.93**	**4.09**	**11.07**	**13.65**	**8.12**
SE	0.306	0.421	0.221	0.325	0.368	0.182
			1990 Green	1990 SDP	1990 CDU	1990 FDP
Raw data						
1990 reference texts						
Economic policy						
A priori positions	—	—	**5.21**	**6.53**	**13.53**	**15.68**
SE ($n = 19$)	—	—	0.652	0.436	0.544	0.613
Social policy						
A priori positions	—	—	**2.90**	**6.68**	**14.42**	**6.84**
SE ($n = 19$)	—	—	0.908	0.856	0.537	0.603
Length in words	—	—	6,345	9,768	7,322	42,446
No. of unique words	—	—	1,838	2,517	1,987	6,594
1994 virgin texts						
Economic policy						
Raw mean word scores (S_{vd})	10.3048	10.2802	10.4459	10.5997	10.7407	10.9059
SE	0.020	0.024	0.012	0.017	0.019	0.010
Social policy						
Raw mean word scores (S_{vd})	7.4136	7.5096	7.6076	7.9257	8.0420	7.7909
SE	0.016	0.019	0.010	0.015	0.017	0.008
Length in words	15,296	10,078	36,419	16,341	14,562	50,452
Unique words scored (N_v)	2,031	1,674	3,455	2,466	2,281	4,168
% words scored	86.7	86.8	86.1	89.8	90.2	87.0
Unique unscorable words	1,294	945	5,064	1,669	1,236	4,707
Mean frequency of unscorable words	1.57	1.41	1.40	1.18	1.16	1.39

Sources: As in previous tables. The rescaled values for PDS 1990 are in the context of the virgin scores for the non-PDS 1994 parties, using the four 1990 texts as references. This procedure has no effect on the value of the raw scores.

Irish case, an important matter to watch for is whether the word scoring technique can pick up what is widely perceived as the classical liberal position of the FPD—on the right of the economic policy dimension and on the liberal side of the social policy dimension—a perception confirmed by the expert survey results reported in the bottom panel in Table 6. The results again suggest a general conservative shift among the establishment parties, most marked with the SDP. Language-blind word scoring also picks up the liberal positions of the FDP, putting this party on the liberal side of the social policy dimension and the right-wing side of the economic policy dimension. Again providing strong face validity for our general approach, the word-scored estimates place the PDS very firmly at the liberal end of the liberal–conservative dimension of social policy. Again, the standard errors imply that, while the position of the PDS is significantly to the left of the other parties, there is no statistically significant difference between the PDS manifesto of 1990 and that of 1994.

Overall, we take these results to show that our word scoring technique can migrate effectively into a non-English-language environment. They illustrate the enormous payoffs available from using language-blind text coding, since our technique allowed us to analyze very quickly and effectively texts written in a language that we do not speak!

USING THE WORD SCORING TECHNIQUE TO ANALYZE LEGISLATIVE SPEECHES

Our demonstration of the word scoring technique thus far has been limited to estimating party policy positions from party manifestos. However, computer word scoring also offers the important prospect of moving into completely new areas of text analysis for which the effort involved has previously been prohibitive. Our technique removes this effort, making it possible to analyze the speeches of all members of a given legislature, for example, opening up the prospect of generating policy spaces that locate politicians in a time series and, thereby, the possibility of much more sophisticated analyses of intra- and interparty politics. Moving beyond party politics, there is no reason the technique should not be used to score texts generated by participants in any policy debate of interest, whether these are bureaucratic policy documents, the transcripts of speeches, court opinions, or international treaties and agreements.

To demonstrate the applicability of our computer word scoring technique to texts other than party manifestos, we now use word scoring to analyze legislative speeches. Although most legislatures have long preserved such speeches as part of the written parliamentary record, these speeches have become

TABLE 7. Mean Raw and Standardized Scores of Speakers in 1991 Confidence Debate on "Pro- versus Antigovernment" Dimension, by Category of TD

		Median		Raw		Standardized	
Group	*N*	Total Words	Unique Words	Mean	SD	Mean	SD
Reference texts							
FF Prime Minister Haughey	1	6,711	1,617	1.0000	—	—	—
FG opposition leader Bruton	1	4,375	1,181	−1.0000	—	—	—
DL leader de Rossa	1	6,226	1,536	−1.0000	—	—	—
Virgin texts							
FF ministers	12	3,851	727	−0.2571	0.0383	1.15	0.66
PD minister	1	2,818	593	−0.2947	—	0.50	—
FF	10	1,553	397	−0.2999	0.0721	0.41	1.24
Independent	1	3,314	582	−0.3360	—	−0.21	—
Greens	1	1,445	415	−0.3488	—	−0.43	—
WP	2	2,001	455	−0.3501	0.0423	−0.46	0.73
FG	21	1,611	394	−0.3580	0.0306	−0.59	0.53
Labour	7	2,224	475	−0.3599	0.0220	−0.62	0.38

highly amenable to computerized analysis as they are increasingly published electronically. While the analysis of speeches holds considerable promise, it also raises new challenges for content analysis—whether computerized or traditional—because such speeches differ substantially from party manifestos in several key respects. First, manifestos are typically comprehensive documents addressing a wide range of policy issues, while speeches tend to be much more restricted in focus. Second, manifestos are published in a political context that is fairly well defined. Greater care must be taken in establishing the political context of speeches if we are to justify the comparison of different speeches in the same analysis. Third, because manifestos and speeches use different language registers and lexicons, the analysis of speeches requires types of reference text different from those used in the analysis of manifestos. Finally, political speeches tend to be much shorter than manifestos. With fewer words to analyze, statistical confidence in the results is likely to be reduced. In these respects, the analysis of legislative speeches will be more problematic than the analysis of party manifestos, and therefore we expect this final test of the word scoring technique to be particularly difficult. If successful, however, we would consider it a major confirmation of the ability of our technique to extract political positions from texts using word frequencies as data.

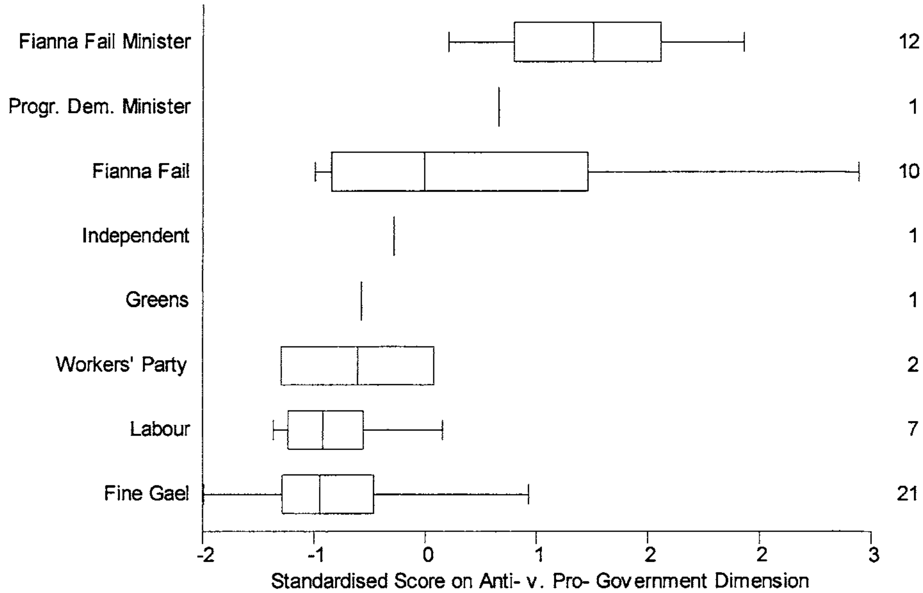
Laver and Benoit (2002) analyzed the acrimonious confidence debate in the Irish Dáil that took place in October 1991 over the future of the incumbent Fianna Fáil–PD coalition government. The matter of interest is the extent to which each individual legislator speaking in this debate was pro- or antigovernment. The texts analyzed were the set of 58 set-piece speeches, extracted from the verbatim transcript of the debate published on the web site of the Irish legislature (http://www.irlgov.ie/oireachtas). The tightly structured debate gave each member of the 166-member Dáil a single opportunity to speak. Including speeches by each party leader, the 58 speeches generated a written record of just over 167,000 words.

For reference positions in the debate, we postulated *a priori* the location of certain party leaders on the "pro- versus antigovernment" dimension. The speech of the *Taoiseach* (prime minister) was assumed axiomatically to be progovernment and assigned a reference position of +1.0. The speech of the Fine Gael leader of the day and leader of the opposition, John Bruton, was assumed axiomatically to be antigovernment and assigned a reference position of −1.0, as was the speech of Prionsias de Rossa, then leader of the opposition Workers' Party. This allowed the calculation of word scores for all different words used in the debate in at least one of the reference texts—a total of 2,856 different words in all. Having calculated word scores from the reference texts, it was then possible to estimate the positions of 55 other speakers on the pro- versus antigovernment dimension, scoring their speeches as virgin texts.

Table 7 presents the results of this analysis, along with descriptive statistics about each text or group of texts from the same party. The three reference speeches were relatively long, as indicated by the median numbers of total and unique words. The virgin texts were typically short, however, with medians of 2,224 total and 508 unique words—much shorter than the typical manifesto analyzed in previous examples, which ranged in length from about 3,000 to 28,000 total words. The bottom half of Table 7 groups legislators by party and, in the case of governing Fianna Fáil and PD parties, by whether or not the legislator was a government minister. Taking the standardized scores as the most interpretable results, our expectation is that members of the coalition parties should be relatively progovernment, that government ministers should be more strongly progovernment then backbenchers, and that opposition legislators should be relatively antigovernment.

The results give us strong encouragement about the possibility of extending the use of our word scoring technique to the analysis of political speeches. Figure 3 shows these results graphically, generating a scale of "support versus opposition" to the government that is readily recognizable by any observer of Irish politics.

FIGURE 3. Box Plot of Standardized Scores of Speakers in 1991 Confidence Debate on "Pro- versus Antigovernment" Dimension, by Category of Legislator



Note: Values at the right indicate the number of legislators in each category.

Fianna Fáil ministers were overwhelmingly the most progovernment speakers in the debate, with Fianna Fáil TDs (members of parliament) on average less progovernment in their speeches. At the other end of the scale, Labour, Fine Gael, and Workers' Party TDs were the most systematically antigovernment in their speeches, closely followed by the sole Green TD.

Not only does the word scoring plausibly locate the party groupings, but also it yields interesting information about individual legislators, whose scores may be compared to those of the various groupings. The position of government minister and PD leader Des O'Malley, for instance (the sole PD minister in Table 7), was less staunchly progovernment than that of his typical Fianna Fáil ministerial colleagues. This may be evidence of the impending rift in the coalition, since in 1991 the PDs were shortly to leave the coalition with Fianna Fáil.

We already noted that the word scoring of relatively short speeches may generate estimates of a higher uncertainty than those for relatively longer party manifestos. This is because our approach treats words as data and reflects the greater uncertainty that arises from having fewer data. In the point estimates of the 55 individual speeches we coded as virgin texts (not shown), greater uncertainty about the scoring of a virgin text was directly represented by its associated standard error. For the raw scores (with a minimum of −0.41 and a maximum of −0.25), the standard errors of the estimates derived from speeches ranged from 0.020, for the shortest speech of 625 words, to 0.006, for the longest speech of 6,396 words, delivered by the Labour Party leader Dick Spring. These errors are indeed larger than those arising in our manifesto analyses. However, substantively interesting distinctions between speakers are nonetheless possible on the basis of the resulting confidence intervals. Considering policy differences within Fine Gael, for example, the raw estimates (and 95% confidence intervals) of the positions of for former FG *Taoiseach* Garrett FitzGerald were −0.283 (−0.294, −0.272), while those of future party leader Enda Kenny were −0.344 (−0.361, −0.327). This allows us to conclude with some confidence that Kenny was setting out a more robustly antigovernment position in the debate than party colleague Fitzgerald. Thus even when speeches are short, our method can detect strong variations in underlying positions and permit discrimination between texts, allowing us to infer how much of the difference between two estimates is due to chance and how much to underlying patterns in the data.

Overall we consider the use of word scoring beyond the analysis of party manifestos to be a considerable success, reproducing party positions in a no-confidence debate using no more than the relative word frequencies in speeches. This also demonstrates three important features of the word scoring technique. First, in a context where independent estimates of reference scores are not available, *assuming* reference text positions using substantive local knowledge may yield promising and sensible results. Second, we demonstrate that our method quickly and effortlessly handles a large number of texts that would have presented a daunting task using traditional methods. Third, we see that the method works even when texts are relatively short and provides estimates of the increased uncertainty arising from having less data.

CONCLUSIONS AND FUTURE WORK

Our word scoring approach to placing political texts on policy dimensions has been demonstrated to be effective at replicating the results of content analysis techniques based on human- or computer-coding. The scores produced by our technique are both substantively plausible and congruent with independent estimates—even when parties made dramatic moves on policy positions, as with the British Labour party in 1997. Furthermore, it avoids many of the problems of traditional techniques of content analysis. First, it produces policy estimates for texts whose positions are unknown, at a low cost and terrific speed—typically completing the analysis in a matter of seconds. In an analysis reported elsewhere, we were able to estimate the policy positions of the Irish political parties during the 2002 Irish general election, updating the analysis the same day each party released its election manifesto on-line (Benoit and Laver Nd.). Second, unlike traditional methods of content analysis, our technique provides quantitative measures of uncertainty for text-based estimates of policy positions. These allow analysts to make informed judgments, when comparing two estimated policy positions, about whether differences between them can be viewed as significant or merely as products of chance or measurement error—something that has not been possible before. Finally, because it treats words simply as data rather than requiring any knowledge of their meaning as used in the text, our word scoring method works irrespective of the language in which the texts are written. In other words, while our method is designed to analyze the content of a text, it is not necessary for an analyst using the technique to understand, or even read, the texts to which the technique is applied. The primary advantage of this feature is that the technique can be applied to texts in any language.

Given these advantages, the computer word scoring approach to text analysis opens up exciting possibilities for the rapid analysis and reanalysis of large and complex text data sets. As political texts become ever more easily available electronically, for example, it is now possible to analyze party manifestos and other election addresses before the election concerned has even taken place. It is worth reiterating that the great leap forward in efficiency made possible by our computational text analysis approach is made possible by a no less dramatic shift from previous applications of content analysis in political science. Our crucial move is to abandon the notion, which runs throughout most political science content analysis, that the objective of an analyst coding a text is to identify its meaning. Indeed, this notion has been so much taken for granted that it is seldom even recognized as an assumption. It is also why many early attempts to computerize content analysis within political science have in effect attempted to automate tasks otherwise performed by human experts, rather than cashing in on the things computers do really well. The results have been rather like the early robots designed in the 1960s—remarkable more because they could do anything at all than because they actually did anything better or faster than real people. As with dictionary-based computer-coding applications, these early robots required frequent human intervention, close monitoring, and occasional direct control to make their behavior realistic. Furthermore, neither robots nor computer algorithms to analyze texts can understand meaning "in context," something easily, if unreliably, performed by humans.[17] Consider an attempt to computer-code the following text: "Some say that I am not averse to the argument that it would be dangerous not to raise taxes. They have every right to say this and nobody would deny them this right, but in this case it is impossible not to conclude they are wrong." While everyone agrees that this would be a wonderful thing to do, no published work has yet reported success at coding large volumes of political text in context, in this sense. Our approach avoids these pitfalls by circumventing them entirely, by treating individual words simply as data rather than attempting to use computerized algorithms to ascribe meaning to these words in an emulation of a human reader.

Nonetheless, sensitive to the issue of analyzing words in context while retaining our insistence on an essentially statistical method, we intend in future work to extend our approach to allow us to analyze word pairs, triples, and indeed n-tuples, as a way of taking one step toward a probabilistic analysis of the context in which individual words are located. Two comments are in order here, however. The first is the purely arithmetical point that, in a text with a total of m words, we must find $m-1$ word pairs, $m-2$ word triples, and $m-n$ word n-tuples. In other words, the number of short word strings in a text is effectively the same as the number of words. But, if there are d different words in a given text, then there are d^2 different possible word pairs and d^n different possible n-tuples. In short, the number of different possible word n-tuples increases exponentially with n, meaning that the relative frequencies of even short word strings in a text are likely to be very, very much lower than the relative frequencies of individual words. Much lower relative frequencies will combine with a much higher probability of unscoreable word strings in virgin texts, meaning that our estimates of the policy positions of the virgin texts will be more uncertain when we move from scoring individual words to scoring word n-tuples. But this will nonetheless be an interesting and important matter to explore.

The second comment on scoring short word strings concerns why our technique appears to work so well without doing this at present. As Laver and Garry (2000) point out when discussing the dictionary-based computer-coding of individual words, this almost certainly has to do with the way that words are used in

[17] Recall the first published reliability tests of the expert coders used by the Comparative Manifestos Project (CMP), Volkens 2001, in which a significant number of coders produced codings that correlated with an "official" coding in the 30–60% range. This is almost certainly the most professionally run and prestigious content analysis project in political science to date. We have seen no other published tests of intercoder reliability in relation to political science content analysis, but we know informally from our own experiences with this that it is a major unspoken problem.

practice in the advocacy of particular policy positions. With regard to our own technique, take the individual word used in our earlier example—"choice." Of course the word "choice" has several meanings, while each meaning can also be qualified with a negative or even a double negative. Someone coming to computational text analysis for the first time might reasonably feel for these reasons that the relative frequency of the word "choice" in a given text does not convey substantive information. This might well be true if our frame of reference were all possible texts written in the English language, read in all possible contexts, but this is very precisely not the frame of reference we propose here. For a given virgin text dealing with a given policy debate in a given political context at a given time—all of these things crucially defined by our selection of a set of reference texts—our approach works because particular words do, empirically, tend to have policy-laden content. Thus, in post-Thatcher Britain, those using the word "choice" in relation to education or health policy, for example, tended to be advocating greater choice of schools or health providers and correspondingly less central control. Those opposing such policies tended, as a matter of empirical observation, not to argue for "no choice" or "less choice" but rather to talk about the benefits of central planning and coordination. This is why the use of the word "choice" in this precise context conveyed substantive information about policy positions. Of course, if the political context changes, the information content of words may well change too—perhaps "citizens now face a stark choice and must sweep out this corrupt administration." If the context changes, however, so must the set of reference texts and hence all word scores—highlighting once more the role of the expert analyst in ensuring that the reference texts reflect in a valid way the political context of the virgin texts to be analyzed. It is patterns in the relative word frequencies observed in the reference texts that define the information content of the words to be analyzed.

In short, our technique works as well as we have shown it to work because, in practice and in a precisely defined context, individual words convey information about policy positions, information revealed in the preliminary analysis of the reference texts. Of course we will almost certainly not always be right when we apply a given word score to a given virgin text. However, provided that we are right more often than we are wrong, a function of choosing good reference texts, and provided that we analyze a large enough number of words, the slender pieces of information we extract by scoring individual words compound to allow us to make what we have shown to be valid estimates.[18]

Computerized word scoring offers the potential for a huge increase in the scope and power of text analysis within political science, but there is still no such thing as a methodological free lunch. While the word scoring technique automates much of the dreary and time-consuming mechanical tasks associated with traditional text analysis, it in no way dispenses with the need for careful research design by an analyst who is an expert in the field under investigation. The key to our *a priori* approach is the identification of an appropriate set of reference texts for a given research context and the estimation or assumption of policy positions for these reference texts with which everyone can feel comfortable. This is by no means a trivial matter, since the word scores for each policy dimension, and hence all subsequent estimates relating to virgin texts, are conditioned on the selection of reference texts and their *a priori* positions on key policy dimensions. This is thus something to which a considerable amount of careful and well-informed thought must be given before any analysis gets under way. In this, our method shares the "garbage in–garbage out" characteristic of any effective method of data analysis; potential users should, indeed, be comforted by this.[19] The casual or ill-informed choice of reference texts or *a priori* policy positions will result in findings that are unreliable—in the same way as will the choice of inappropriate or poorly worded survey questions or an inappropriate or ambiguously defined content analysis coding scheme. Given a valid set of reference texts, however, and good estimates or assumptions of the policy positions of these, computer word scoring offers the potential to crunch huge volumes of virgin text very rapidly indeed, with an enormous range of intriguing political science applications.

[18] Statistically, there is an analogy with the Condorcet Jury Theorem—if we treat individual words as jurors deciding on the policy content of texts.

[19] If they are not, they should consider what they would feel about a method offering "garbage in, gold out."

REFERENCES

Adams, J. 2001. "A Theory of Spatial Competition with Biased Voters: Party Policies Viewed Temporally and Comparatively." *British Journal of Political Science* 31: 121–58.

Bara, Judith. 2001. "Tracking Estimates of Public Opinion and Party Policy Intentions in Britain and the USA." In *Estimating the Policy Positions of Political Actors*, ed. Michael Laver.London: Routledge, 217–36.

Baron, David. 1991. "A Spatial Bargaining Theory of Government Formation in Parliamentary Systems." *American Political Science Review* 85: 137–64.

Baron, David. 1993. "Government Formation and Endogenous Parties." *American Political Science Review* 87: 34–47.

Benoit, Kenneth, and Michael Laver Nd. "Estimating Irish Party Positions Using Computer Wordscoring: The 2002 Elections." *Irish Political Studies*.

Blais, A., D. Blake, and S. Dion. 1993. "Do Parties Make a Difference—Parties and the Size of Government in Liberal Democracies." *American Journal of Political Science* 37: 40–62.

Budge, Ian, and Denis Farlie. 1977. *Voting and Party Competition.* London: Wiley.

Budge, Ian, Hans-Dieter Klingemann, Andrea Volkens, Judith Bara, and Eric Tannenbaum, with Richard Fording, Derek Hearl, Hee Min Kim, Michael McDonald, and Silvia Mendes. 2001. Mapping Policy Preferences: Parties, Electors and Governments, 1945–1998; Estimates for Parties, Electors and Governments, 1945–1998. Oxford: Oxford University Press.

Budge, Ian, David Robertson, and Derek Hearl, eds. 1987. *Ideology, Strategy and Party Change: Spatial Analyses of Post-War Election Programmes in 19 Democracies*. Cambridge: Cambridge University Press.

Castles, Francis and Peter Mair. 1984. "Left-Right Political Scales: Some Expert Judgments." *European Journal of Political Research* 12: 83–88.

de Vries, Miranda, Daniela Giannetti, and Lucy Mansergh. 2001. “Estimating Policy Positions from the Computer Coding of Political Texts: Results from Italy, The Netherlands and Ireland.” In *Estimating the Policy Positions of Political Actors*, ed. Michael Laver. London: Routledge, 193–216.

Gabel, Matthew, and John Huber. 2000. “Putting Parties in Their Place: Inferring Party Left-Right Ideological Positions from Party Manifesto Data.” *American Journal of Political Science* 44: 94–103.

Kim, H. M., and R. C. Fording. 1998. “Voter Ideology in Western Democracies, 1946–1989.” *European Journal of Political Research* 33: 73–97.

Kitchshelt, Herbert. 1994. *The Transformation of European Social Democracy.* Cambridge: Cambridge University Press.

Kleinnijenhuis, Jan, and Paul Pennings. 2001. “Measurement of Party Positions on the Basis of Party Programmes, Media Coverage and Voter Perceptions.” In *Estimating the Policy Positions of Political Actors*, ed. Michael Laver. London: Routledge, 161–82.

Klingemann, Hans-Dieter, Richard Hofferbert, and Ian Budge, with Hans Keman, Torbjorn Bergman, François Pétry, and Kaare Strom. 1994. *Parties, Policies and Democracy.* Boulder, CO: Westview Press.

Laver, Michael. 1994. “Party Policy and Cabinet Portfolios in Ireland 1992: Results from an Expert Survey.” *Irish Political Studies* 9: 157–64.

Laver, Michael. 1998a. “Party Policy in Britain, 1997: Results from an Expert Survey.” *Political Studies* 46: 336–47.

Laver, Michael. 1998b. “Party Policy in Ireland, 1997: Results from an Expert Survey.” *Irish Political Studies* 13: 159–71.

Laver, Michael, and Kenneth Benoit. 2002. “Locating TDs in Policy Spaces: Wordscoring Dáil Speeches.” *Irish Political Studies* 17 (Summer): 59–73.

Laver, Michael, and Ian Budge, eds. 1992. *Party Policy and Government Coalitions.* London: Macmillan.

Laver, Michael, and John Garry. 2000. “Estimating Policy Positions from Political Texts.” *American Journal of Political Science* 44: 619–34.

Laver, Michael, and William Ben Hunt. 1992. *Policy and Party Competition.* London: Routledge.

Laver, Michael, and Norman Schofield. 1998. *Multiparty Government: The Politics of Coalition in Europe*. Ann Arbor: University of Michigan Press.

McDonald, Michael D., and Silvia M. Mendes. 2001. “The Policy Space of Party Manifestos.” In *Estimating the Policy Positions of Political Actors*, ed. Michael Laver. London: Routledge.

Muller, Wolfgang, and Kaare Strom, eds. 2000. *Coalition Governments in Western Europe.* Oxford: Oxford University Press.

Schofield, N., and R. Parks. 2000. “Nash Equilibrium in a Spatial Model of Coalition Bargaining.” *Mathematical Social Science* 39: 133–74.

Volkens, Andrea. 2001. “Manifesto Research Since 1979: From Reliability to Validity.” In *Estimating the Policy Positions of Political Actors*, ed. Michael Laver. London: Routledge.

Warwick P. 1994. *Government Survival in Parliamentary Democracies.* Cambridge: Cambridge University Press.

Warwick, Paul. 2001. “Coalition Policy in Parliamentary Democracies—Who Gets How Much and Why.” *Comparative Political Studies* 34: 1212–36.

Warwick, Paul. 2002. “Toward a Common Dimensionality in West European Policy Spaces.” *Party Politics* 8: 101–22.