Comparing the Location of Canadian Political Parties Using French and English Manifestos as Textual Data

Benoît Collette

and

François Pétry

Université Laval

Department of Political Science

Centre for the Analysis of Public Policy

Quebec City, Canada

Abstract:

In recent years, we have seen a multiplication of studies estimating the location of political parties on the left-right axis using textual data. The addition of computer-assisted content analysis techniques to the researcher's toolbox has opened new fields of inquiry. So far, studies have focused on one specific language, and little attention has been paid to linguistic differences in party manifestos that are provided to the voters in more than one language. There is no theoretical reason to assume that two party platforms that have the same content but are written in different languages will not locate parties the same way on the left-right axis. However differences between languages are well documented in the field of linguistics. In the case of Canada, there are numerous differences between its two official languages, French and English, that may modify the results of textual analyses depending on which language they are performed. The relative frequency of words in a text is an example: French has gender categories (masculine-feminine) for nouns and adjectives and a more differentiated verb termination system (usually 5 compared to 2 in English). Considering the fact that quantitative content analyses are often based on word frequencies, a comparative, multi-language analysis is a first step to test the impact of linguistic specificities on party positioning. Using Wordscores and Wordfish, plus expert surveys as a benchmark, we compare the spatial location of major Canadian parties on the left-right axis using English and French manifestos over the past 25 years to see if the spatial representation is the same in both languages or if there are fundamental and significant differences between them.

Keywords: Wordscores, Wordfish, Canada, Parallel textual data.

The capacity to locate political parties in a common space allows researchers to compare “like with like”, a pre-requisite to further comparisons of parties and party systems cross-nationally, across federated states within one single federation (the Canadian provinces for example), and over time. Party location on a common space has become the instrument of choice in comparative typologies of parties and party systems (Blondel 1968; Laver and Hunt 1992; Sartori 2005 [1976]) to the point that earlier criteria of comparison, such as whether two parties belong to the same ideological family, or the distinction between mass parties and cadre parties, are now considered obsolete by many comparative party scholars. The capacity to locate political parties in a common space allows us to better assess how representative governments really are. By locating political parties within a common space, and by comparing their position to the preferences expressed by voters, researchers can obtain a measure of congruence between the two, which can be used to test the relative merits of various voting models (Budge and Farlie 1983; Downs 1957; Rabinowitz and MacDonald 1989).

Word-based parallel content analysis

Over the past thirty years, the methods employed by researchers to locate political parties have multiplied from hand-coded methods, such as the well-known CMP technique (Budge et al. 2001; Klingemann et al. 2007), to expert surveys (Castles and Mair 1984; Huber and Inglehart 1995; Laver and Hunt 1992) to dictionary methods (Ray 2001; Laver and Garry 2000; Kleinnijenhuis and Pennings 1999), to computer-assisted content analysis methods. These new computer-assisted, quantitative methods for extracting political party positions on the left-right axis or other policy dimensions from political texts have been a useful addition to the researcher’s toolbox. They rely on objective textual data; can be used to with nearly unlimited flow of data and make it possible to isolate policy preferences from behaviour (see Benoit and Laver 2007b; Laver and Garry 2000; Marks et al. 2007).

The description of stochastic text generation process by Benoit, Laver, and Mikhaylov (2009: 497-501) could be generalized from the CMP, as illustrated in the article, to content analysis as a whole. The “true” preferences and the intended message of authors are unobservable and uncertain. Only the text generated by authors is observable and certain. All methods try to infer, logically and statistically, the “true” preferences of authors using different models. CMP does so by coding texts and devising scales to measure the position of texts in policy space. “Word-based” techniques, such as Wordscores (Laver et al. 2003) and Wordfish (Slapin and Proksch 2008), analyse the vocabulary, and more specifically the distribution of words, of political actors in a text to extract policy positions. We call them “word-based” techniques because the analytical unit are the words in a text, not paragraphs, sentences, locutions or topics. This particularity has two main advantages. By chopping texts into words, word-based techniques gain the advantage of analytical simplicity, because words can be automatically identified and treated without human intervention.[1] Second, since words are treated like quantitative data, the knowledge of a language is no longer necessary to extract and then compare policy positions from texts written in different languages.

The disadvantage of word-based techniques is that they do not take into account the meaning or the grammatical structure of sentences and words that make them up. Focusing exclusively on the relative frequency of mention of words can lead to some linguistic nonsense when the logic is pushed to its limits. As an extreme illustration, it is possible to extract a policy position from a random bunch of words that no human reader could make sense of, or from a freely reorganized text in, say, alphabetical order. Also, it is impossible to measure the positive or negative direction of a policy

[1] In practise things are more complicated and experience showed us that, for example, hyphenated locutions can be sometimes treated as a single word or as separated words depending on the method used to produce frequency matrix.

preference in a text (Monroe et al. 2008). For example, the meaning of the sentence “We will raise taxes” is the exact opposite of “We will not raise taxes”. But if we cut these sentences into separate words “we” (2 times) “will” (2) “not” (1) “raise” (2), and “taxes” (2), the difference between the two is “not”, a relatively meaningless word which is likely to be overlooked. The difference in the meanings of the two sentences will be blurred as a result.

How crippling is this disadvantage when comparing political texts written in different languages? It is this paper’s objective to find out. We do this by checking whether Wordscores and Wordfish extract the same policy positions on the left-right axis from parallel texts. Parallel texts are original documents written in different languages, not translations. They can be use to benchmark automatic translation quality (see Jian-Yun et al. 1999 for applications of parallel texts).

Do Wordscore and Wordfish extract the same policy positions on the left-right axis from parallel documents. In theory, there is no reason to believe that they would not. Parallel documents are rigorously the same. Their format is identical, they include the same topics and a bilingual reader will consider those documents to be almost exactly the same. Studies analyzing parallel textual data are interesting because they give an opportunity to test the validity and the reliability of word-based textual analysis methods in a more rigorous way than repeated studies focusing on a single language and/or a single party system. Wordscores has been tested with languages other than English, such as Dutch (Klemmensen et al. 2007), German (Hug and Schulz 2007; Magin et al. 2009; Bräuninger and Debus 2008), and French (Laver et al. 2006). But we could find only one study comparing Wordscores results using bi-texts as input. Debus (2009: 53-54) uses Wordscores to compare Flemish and French coalition agreements in Belgium and finds no significant difference between them for economic policy positions. In their analysis of European Parliament speeches Slapin and Proksch (2008) compare Wordfish results for speeches in English, French and German. They find remarkable similarities between languages (English and French especially):

> The comparison of the results across languages suggests that the position estimation technique is in fact highly robust to the choice of language (the correlation coefficient is 0.86 or higher). The highest correlation is between positions estimated from the English and French translations. These two languages are so similar to each other with regard to the information contained in words that they produce virtually identical position estimates (Proksch and Slapin 2009: 13).

It should be noted that due to the large number of official languages spoken in the European Union, speeches in the European Parliament are delivered in one language and then translated. Slapin and Proksch had to rely on translations instead of original texts. This suggests a fortiori that the positions extracted from parallel documents such as party manifestos should be very similar across different languages.

Methodology

We test Wordscores and Wordfish on Canadian parallel manifestos and compare the results on the left-right dimension with expert surveys to estimate the cross-language reliability of these two techniques. Canada, as an officially bilingual country, is a pertinent case for this test. At the federal level manifestos from major parties are bilingual and both versions are considered official and public. With the exception of the Bloc québécois, Canadian party manifestos can qualify as perfect parallel texts. Also, Canada has clearly identified polarized parties, the New Democratic Party (NDP) and the Conservative Party, on the left-right dimension (Irvine 1987; Johnston 2008). It is necessary to have a right and a left end point in the Canadian federal partisan spectrum in order to use Wordscores and Wordfish because both these techniques need to set values on reference texts.[2]

There is a debate over the merits of word stemming (reducing words top their root form) and removing stopwords (the, who, that, le, qui, quoi, etc.) with computer-assisted content analysis methods (see Lowe 2008). Stemming words and removing stopwords tends to reduce the size of texts and word diversity. That is supposed to lead to better results as meaningless words are removed and family of words reduced to stems. On the other hand, these operations artificially skew the word distribution in a text that could distort results. Although word stemming and removing stopwords is a common practice with Wordfish, we will compare results with non-stemmed and stemmed texts.[3] This comparison will be a pertinent addition in the discussion. Jfreq has built-in stemming and stopword removal options that have been used to create the different matrixes needed.[4]

The corpus is composed of Canadian manifestos from the 2000, 2004, 2006, and 2008 elections. The relatively short period of time will minimize the problem of political vocabulary change, a concern for both Wordscores and Wordfish (Slapin and Proksch 2008; Budge and Pennings 2007). Three major parties are included in the analysis: the New Democratic Party (NDP), the Liberal Party and the Conservative Party. The Bloc québécois has been excluded because it did not produce sufficiently reliable English versions of its manifestos to qualify as parallel texts. Original manifestos have been tailored and standardized to facilitate content analysis. Messages from the leaders, financial annexes, tables, and quotations from other works (newspaper clips or OECD Reports for example) have been excluded from the original texts. For Worsfish, we followed Proksch and Slapin's (2008) recommendation and eliminated unique party words (words that are used by only one party) from the texts analyzed.

Parallel texts differ only by the language in which they are written. Is language difference a factor that might provoke Wordscore and Wordfish to extract noticeably different positions from two party manifestos that are exactly similar in all other respects? Some specific features of languages can have a significant impact on the distribution of words in a text. French and English differ on many levels: syntax, grammar, and style (see Lederer 1994; Vinay and Darbelnet 2003 [1977]). An important syntaxic difference between French and English is the usage of articles. In French, they are more frequent than in English, because they are quasi-mandatory in before nouns. We can expect to find some common articles (de, du, la, les, etc.) to have a very high frequency, higher than the equivalent in English.

Looking at grammar, we find several of significant differences. In French gender tends to duplicate adjectives, pronouns, and articles as they can be declined in either masculine of feminine form. French texts tend to have relatively higher differentiations of those words when compared to English and that means expressing the same concepts with more words that are relatively less frequent than in English. While in English nouns may be singular or plural, in French adjectives can also be singular or plural, in addition of being masculine or feminine. So it is possible to have four words to express the same concept instead of one in English. As a simple illustration, take the adjective "pretty" in English. In French it can be declined in masculine singular form ("beau"), in feminine singular ("belle"), in masculine plural form ("beaux"), and feminine plural ("belles") depending on the object. Inevitably "pretty" will be more frequent in a text than one of the four French equivalent adjectives.

[2] Contrary to what Slapin and Proksch (2008: 708) state, we need reference texts with Wordfish. Among the parameters, the user has to set omega scores for references texts representing both extremes of the dimension on which political texts will be located.
[3] Stemming and removing is more complicated with Wordscores because raw texts need to be treated instead of word matrix with Wordfish. Results with Wordscores will be included in a future version of this paper.
[4] To remove French stopwords, we used the Snowball list (http://snowball.tartarus.org/algorithms/french/stop.txt).

Another important difference between French and English is the verb termination system. In English, there are two forms in the present tense, one for the third person singular (he/she/it) and one for the rest, while most of the time in past tense you have only one. Future, conditional, and subjective have only one. In French, you usually find five forms, one for each person plural (we/you/they) and two for the singular persons (either I/you or I/he-she-it), depending on the verb family, and those forms are applied to future, conditional, subjunctive, and past tenses. Therefore, the same verbs tend to be declined in more different words in French, modifying again the word distribution.

One consequence of these differences will appear in the number of different words in a text. French texts have more words than English texts. Unsurprisingly, between 2000 and 2008 the manifestos from major parties contain 11 717 different words in French, and 8 966 in English.[5] The average frequency of words is higher in English than in French. Overall, the mean frequency is 19.1 in French and 18.7 in English, but the standard deviation is much larger in French (218) than in English (176). Thirteen French words (du pour, un, en, d', l', le, a, des, la, et, les, de), many of them articles, appear more than 2000 times, while in English only eight words appear more than 2000 times (for, will, in, a, of, to, and, the). When we exclude those words the mean frequency becomes 12.8 in French and 14.3 in English. Finally, we find more unique words in French (words used only once). There are 4 625 unique words in French and 3 361 in English.

Canadian expert surveys

Expert surveys will serve as a benchmark to compare the face validity of the Wordscores and Wordfish estimates. Expert surveys are widely used to locate political parties in a policy space and often serve as a benchmark for textual analysis (Benoit and Laver 2007b; Keman 2007; Marks et al. 2007; Ray 2007; Volkens 2007; Whitefield et al. 2007; Laver and Garry 2000).[6] This approach was made popular thanks to a well publicized study by Castles and Mair (1984) who asked political scientists in 17 countries to locate political parties in their own country on an 11-point left-right scale. This was followed by a more ambitious study of left-right party location in 42 countries (Huber and Inglehart 1995). These two studies have inspired a large body of country-specific studies using expert surveys to locate political parties on a range of issues (for a sample see Laver 1998; Laver and Hunt 1992).

In constructing our own expert survey, we replicate the well-tested methodology already used by Laver and Hunt (1992), Laver (1998), Laver and Benoit (2005), and Benoit and Laver (2007b). Our online survey questionnaire included 18 questions on a 0-10 scale. Respondents were asked to position the Bloc, Conservatives, Liberals, NDP, and Greens using the 11 point scale on the left-right axis and other policy dimensions. For all dimensions 0 represents the left position and 10, the right position. The expert survey was conducted immediately after the Canadian general election, held on October 14 2008. The electronic questionnaire was sent to every political science professors in Canadian universities. 163 experts sent back the questionnaire.[7] Of the total, 127 questionnaires (78%) were filled in English and 27 (22%) in French.

Table 1 about here

[5] The frequency matrix from which data are taken has been produced with Will Lowe's jfreq program, available at http://www.williamlowe.net/software/.

[6] For reviews of the data on party placement and on self placement on policy spaces using the mass or expert survey method see Knutsen (1998); Budge (2000); Laver and Garry (2000); Mair (2001).

[7] The data from the 2008 expert survey can be obtained upon request from the authors.

Table 1 indicates how the experts who responded to our survey located the parties on the left-right policy dimension. The first column of numbers gives the mean score for the entire sample of experts. But what is of direct interest here is the mean score for each subsample. The mean score for the location of the Conservative Party among English speaking respondents is 7.94 with a 95% confidence interval ranging from 7.7 to 8.1. If our sample of experts were a random sample (which it is not) we would say that 19 times out of 20 English speaking experts in the population located the Conservative Party between 7.7 and 8.1 points on the left-right axis in 2008. The mean score for the location of the Conservative Party among French speaking respondents is 8.39 with a 95% confidence interval ranging from 8.0 to 8.8. Note that the confidence interval for French speaking experts is much larger than for English speaking experts (this is of course due to the much smaller size of the French sample) and that the two intervals barely overlap each other. In other words, it is likely that French experts located the Conservative Party significantly more to the right than English experts in 2008. The mean scores for the location of the Liberal Party and the NDP are 5.2 for English experts and 5.5 for French experts, respectively. The small overlap between the two confidence intervals makes it once again likely that French speaking experts in the population locate the Liberals significantly more to the right than English speaking experts. The means scores for the NDP differ very little in the sample and the large overlap of the confidence intervals indicate that the location of the NDP on a left-right axis by French speaking experts is indistinguishable from the location by English speaking experts. It should be noted that the scores for the Conservatives, the Liberals and the NDP are all statistically different from each other among both English and French speaking experts.

To summarize, both French and English speaking experts distinctly locate the Conservatives, the Liberals and the NDP. French speaking experts locate the Conservatives and the Liberals more to the right than English speaking experts, although the difference is not large for the Liberals. There is no difference in the way English and French experts locate the NDP. Let us now examine how Wordscores and Wordfish extract the locations of the Conservatives, Liberals and NDP on a left-right dimension from their manifestos and compare the results with the results from the expert survey. .

Wordscores

The first computer-assisted method selected to locate the positions of political parties is the Wordscores computer program developed by Michael Laver and his co-researchers (Laver, Benoit and Garry 2003).[8] Unlike the CMP and dictionary methods that treat texts as discourse to be understood and interpreted for meaning either by a human coder or by a computer, Wordscores treats texts (more precisely the words contained in those texts) as data containing information about the position of the texts' authors on predefined policy dimensions. Starting from a set of "reference" texts whose policy positions are determined a priori, the technique extracts data from these reference texts in the form of word frequencies and uses this information to estimate the policy positions of "virgin" texts about which nothing is known.

Wordscores has the advantage of producing a distribution of scores around an estimated mean score. This makes it possible to come up with a standard error and therefore to establish a confidence interval around the estimated mean score. Wordscores provides a statistical measure of how different two virgin texts are from one another in their vocabulary. In fact, two texts are statistically different if their confidence intervals do not overlap. Of course, the scores are all the more valid if one has confidence in the choice of the references texts and in the measure used to decide what their

[8] For review of Wordscores, see Benoit and Laver (2007a; 2008) ; Klemmensen *et al.* (2007); Volkens (2007) Lowe (2008); Martin and Vanberg (2008); Slapin and Proksch (2008).

positions are on a given scale or cleavage. Laver and collaborators make several important recommendations concerning the selection and scoring of references texts, one of them being that the virgin texts and the reference texts must share a similar frame of reference (Laver et al. 2003: 313-315). In our case, all documents are party manifestos that share the same structure. The recommendation is therefore fulfilled.

Here is a summary of how we used the Wordscores program to measure the positions of the Canadian parties on policy scales. There is a debate over the proper score transformation method that should be used (see Martin and Vanberg 2008; Benoit and Laver 2008). In this paper, we will overlook this debate and keep the original method.

We first select the reference texts which will be used to represent the extreme positions on the a priori defined policy scale. As Laver, Benoit and Garry (2003) point out, it is important that the reference texts are directly relevant to the virgin texts under analysis. We use manifestos from the previous election as the reference texts.[9] The NDP manifestos are arbitrarily coded 0 (left), and the Conservative manifestos 10 (right) for the left-right dimension.

Each virgin text (that is each party platform at each election) is then coded by the Wordscores program which gives to each word in each virgin text a score between 0 and 10 according to the relative frequency of its appearance in the reference texts. For example, if the word "healthcare" appears one percent of the time in the NDP reference text, and 0.9 percent of the time in the Conservative reference text word obtains a score equal to (0.01*-10) + (0.009*10) = -0.01. By dividing the sum of the scores associated with each word by the total number of words in a text, we obtain an average which corresponds to the total score of the text. From the wordscores in each reference text, we computed the textscores in each virgin text, and then transformed the virgin textscores to their original metric to be able to locate the positions of each platform at each election in our pre-defined space.

Figure 1-2 about here

English and French Wordscores estimates correlate at 70.5%.[10] Figure 1 and 2 show confidence intervals for English and French position estimates on the left-right dimension. Note that the confidence intervals do not overlap most of the time: The locations of party manifestos are most often statistically distinct. This finding is consistent with the results of the expert survey. It should be noted that Wordscores tend to produce statistically distinct estimates of party positions (Klemmensen et al. 2007; Laver et al. 2003; Hug and Schulz 2007; Magin et al. 2009; Laver et al. 2006). The left-right ordering of the parties is also consistent with the results of the expert survey. Note that the left-right ordering is stable over time: As expected, the NDP is positioned on the left and the Conservative party on the right with the Liberal party somewhere between them. The only exception is the crossing of Liberals toward the left and Conservatives toward the right between 2000 and 2004 in French. In English the position of the parties remains stable relative to one another throughout the period.

A significant difference is the mean interval size. It is significantly higher for French manifestos (3.5), than English (2.7). This result is surprising at first glance because French manifestos are longer in general than English manifestos. They have, in average, more unique words (1898 vs. 1529) and more total scored words (17 310 vs. 11 548). We would expect longer documents to be more precise, as they contain more information. In English, the correlation between unique words and confidence

[9] For 2000, we had to use the 1993 manifestos as no French manifesto was available in 1997 to be the right position reference text.

[10] Debus (2009) does not indicate correlation levels for French and Flemish coalition agreements. Therefore we cannot establish a comparison.

interval size (-86.1%) is almost the same as in French (-80.7%). But there is a significant difference in the correlation between total scored words in English (-76%) and French (-60%). Since we used raw text documents without stemming and removing stopwords that could mean although French manifestos contain more words, a significant number of them may have median scores that tend to neutralize each other overall. For example, words like “the” or “and” are found multiple times in all texts. They will be given a median wordscore because of their ubiquity. Those wordscores will have a relatively low effect on textscores. This difference is coherent with the linguistic hypothesis formulated earlier that there are more "meaningless" words in French than in English. When looking at specific words, we note it is indeed the case. For example, “the” wordscore will vary from 4.7 to 5.7; “and” from 4.3 to 4.9; “de” from 4.6 to 5.2; and “les” from 4.1 to 5.1.

Wordfish

The Wordfish method takes a different approach. Borrowing from linguistics, it uses a naïve Bayes assumption to infer the process by which words are processed in a text.[11] A text is represented as a vector of word counts (occurrences) and individual words are assumed to be distributed at random. The probability that each words occur in a text is independent of the position of other words in the text. While empirically false, naïve Bayes often performs well for classification (McCallum and Nigam 1998). Wordfish assume the word frequencies are generated by a Poisson process. Slapin and Proksch (2008) chose this particular distribution because of its estimation simplicity. The single paramater, λ, is both the mean and the variance. The functional form of the model is as follows:

$$\lambda_{ijt} = \exp\left(\alpha_{it} + \psi_j + \beta_j * \omega_{it}\right)$$

Where λ_{jit} is the count of word *j* in party *i*’s manifesto at time *t*, α is a set of party-election fixed effects, ψ is a set of word fixed effects, β is an estimate of a word specific weight capturing the importance of word *j* in discriminating between party position, and ω is the estimate of party *i*’s position in election year *t*. Each platform is treated as separate party position and all positions are estimated simultaneously. That means there is no temporal constraint on the position of a party *i*’s manifesto in election *t*. So if a party uses words in similar relative frequencies over time, its position will remain the same. Party movement is due to changes in word frequencies, not to change in word signification.

To calculate a 95% confidence interval, a parametric bootstrap is required. The parametric bootstrap is a computer-intensive re-sampling method. By running the EM algorithm, λ_{jit} is calculated for each cell in the dataset that represent word frequencies. 500 new datasets are then generated, taking random draws from a Poisson distribution with parameter λ_{jit} for each cell. Using the maximum likelihood estimates as starting value, the algorithm is rerun on each of these new datasets and 500 new party positions are estimated.[12] By using the 0,025 and 0,0975 quintiles of the simulated party positions, confidence interval is approximated. With a parametric bootstrap, confidence intervals shrink as the number of words increase.

Wordfish gives two sets of results. The first one is an estimation of the location of political parties on an axis that corresponds to the selected axis, with a confidence interval for every manifesto and the second for each word found in selected texts. The document scores make it possible to position

[11] See Lowe (2008) and Monroe *et al.* (2008) for a discussion over potential problems affecting Wordfish.
[12] Due to time constraints, we had to reduce to reduce the number of simulations to 50 for the left-right dimension. Usually it takes more than 72 hours to run 500 simulations with a standard desktop computer (2GHz processor with 1.75 GMb RAM running R onLinux).

parties on policy dimensions and compare these positions with estimates of other methods. For reasons of comparability, we set the same values as Wordscores for reference texts in each dimension (0 for NDP 2004 and 10 for Conservative 2004).

Figure 3 to 6 about here

Figure 3 and 4 compare the location of English and French manifestos with raw documents (no stemming, no stopword removal). NDP and Conservative party show a similar pattern but the Liberal party is located differently. While in English scores are close to 0 in all four elections, in French scores are close to -10. There is a clear opposition between the Liberals and the Conservatives in French, with the NDP in between, while in English NDP and Liberals occupy the same position. If we look at Figure 5 and 6 that compare the location of English and French manifestos with stemmed documents, that distinctive Liberal pattern appears also in English, with Liberal score being close to -10. Stemming had a different impact on confidence interval size. It has a negative impact in English and a positive impact in French. However, a higher number of simulations (500 instead of 50) would probably cancel this effect, as confidence intervals will shrink.

Table 2 about here

Table 2 reports the correlation of French and English estimates. Correlation between French and English is 81,5% with raw documents and 93,9% with stemmed documents. Such figures are comparable with Proksch and Slapin (2009: 19), who reported a 86% correlation between English and French stemmed speeches.

Conclusion

Do grammatical differences between languages threaten the validity and the reliability of left-right party position results generated by word-based textual analysis techniques such as Wordscores and Wordfish? There are some differences between languages but they appear minor with the exception of the switch between Liberals and Conservatives from 2000 to 2004 in English. This case points to possible problems with the French versions of Conservative and Liberal manifestos in 2000. The English and French results generated by Wordscores appear at least as reliable and valid as those generated by expert survey. In fact the confidence intervals for Wordscores do not appear larger (they often appear smaller) than the confidence intervals for expert survey. Wordfish generated highly correlated estimates that are inconsistent with experts, but some important variables have not been controlled, such as document length. On average, Liberal manifestos are almost twice as long as Conservative manifestos. This asymmetry could have skewed Wordfish estimates. Canadian results do not clearly indicate a fundamental problem with Wordscores and Wordfish, but provide an argument in favor of stemmed documents. We agree with Slapin and Proksch (2008) that stemming documents should be done before using Wordfish, as estimates from stemmed documents are more correlated and have smaller confidence intervals.

So far, we have limited our scope to European languages that are relatively close to each other. Including non-European languages in comparative research would be an interesting addition and a test for the universality of computer-assisted content analysis. Research in comparative linguistic is progressing fast (for example, see Yang and Li 2003) and its applications could be useful for political textual analysis.

References

Benoit, Kenneth, and Michael Laver. 2007a. Benchmarks for text analysis: A response to Budge and Pennings. *Electoral Studies* 26 (1), pp. 130-5.
———. 2007b. Estimating party policy positions: Comparing expert surveys and hand-coded content analysis. *Electoral Studies* 26 (1), pp. 90-107.
———. 2008. Compared to What? A Comment on "A Robust Transformation Procedure for Interpreting Political Text" by Martin and Vanberg. *Political Analysis* 16 (1), pp. 101-11.
Benoit, Kenneth, Michael Laver, and Slava Mikhaylov. 2009. Treating Words as Data with Error: Uncertainty in Text Statements of Policy Positions. *American Journal of Political Science* 53 (2), pp. 495-513.
Blondel, J. 1968. Party Systems and Patterns of Government in Western Democracies. *Canadian Journal of Political Science / Revue canadienne de science politique* 1 (2), pp. 180-203.
Bräuninger, Thomas, and Marc Debus. 2008. Der Einfluss von Koalitionsaussagen, programmatischen Standpunkten und der Bundespolitik auf die Regierungsbildung in den deutschen Ländern. *Politische Vierteljahresschrift* 49 (2), pp. 309-38.
Budge, I., H.-D. Klingemann, A. Volkens, J. Bara, and E. Tanenbaum eds). 2001. *Mapping Policy Preferences. Estimates for Parties, Electors, and Governments 1945-1998*. Oxford: Oxford University Press.
Budge, Ian. 2000. Expert Judgements of Party Policy Positions: Uses and Limitations in Political Research *European Journal of Political Research* 37.
Budge, Ian, and Dennis J. Farlie. 1983. *Explaining and Predicting Elections. Issue Effects and Party Strategies in Twenty-Three Democracies*. Londres: George Allen & Unwin.
Budge, Ian, and Paul Pennings. 2007. Do they work? Validating computerised word frequency estimates against policy series. *Electoral Studies* 26 (1), pp. 121-9.
Castles, Francis G., and Peter Mair. 1984. Left-Right Political Scales: Some 'Expert' Judgments. *European Journal of Political Research* 12 (1), pp. 73-88.
Debus, Marc. 2009. Pre-electoral commitments and government formation. *Public Choice* 138 (1), pp. 45-64.
Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper.
Huber, John D., and Ronald Inglehart. 1995. Expert Interpretations of Party Space and Party Locations in 42 Societies. *Party Politics* 1 (1), pp. 73-111.
Hug, Simon, and Tobias Schulz. 2007. Left--Right Positions of Political Parties in Switzerland. *Party Politics* 13 (3), pp. 305-30.
Irvine, William. 1987. Canada 1945-1980: Party Platforms and Campaign Strategies. In I. Budge, D. Robertson and D. Hearl (ed.), *Ideology, Strategy, and Party Change: Spatial Analysis of Post-War Elections Programmes in Nineteen Democracies*. Cambridge: Cambridge University Press, pp. 73-94.
Jian-Yun, Nie, Michel Simard, Pierre Isabelle, and Richard Durand. 1999. "Cross-language information retrieval based on parallel texts and automatic mining of parallel texts from the Web." In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. Berkeley, California, United States: ACM.
Johnston, Richard. 2008. Polarized Pluralism in the Canadian Party System: Presidential Address to the Canadian Political Science Association, June 5, 2008. *Canadian Journal of Political Science/Revue canadienne de science politique* 41 (04), pp. 815-34.
Keman, Hans. 2007. Experts and manifestos: Different sources - Same results for comparative research? *Electoral Studies* 26 (1), pp. 76-89.
Kleinnijenhuis, J., and P. Pennings. 1999. "A Probabilistic Keyword Approach to Textual Analysis." In

Mannheim Joint Sessions of the ECPR. Mannheim.

Klemmensen, Robert, Sara Binzer Hobolt, and Martin Ejnar Hansen. 2007. Estimating policy positions using political texts: An evaluation of the Wordscores approach. *Electoral Studies* 26 (4), pp. 746-55.

Klingemann, Hans-Dieter, Andrea Volkens, Judith Bara, Ian Budge, and Michael D. McDonald. 2007. *Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Central and Eastern Europe*: Oxford University Press.

Knutsen, Oddbjorn. 1998. The Strength of the Partisan Component of Left-Right Identity: A Comparative Longitudinal Study of Left-Right Party Polarization in Eight West European Countries. *Party Politics* 4 (1), pp. 5-31.

Laver, Michael. 1998. Party Policy in Britain 1997: Results from an Expert Survey. *Political Studies* 46 (2), pp. 336-47.

Laver, Michael, and Kenneth Benoit. 2005. Estimating Party Policy Positions: Japan in Comparative Context. *Japanese Journal of Political Science* 6 (2), pp. 187-209.

Laver, Michael, Kenneth Benoit, and John Garry. 2003. Extracting Policy Positions from Political Texts Using Words as Data. *The American Political Science Review* 97 (2), pp. 311-31.

Laver, Michael, Kenneth Benoit, and Nicolas Sauger. 2006. Policy Competition in the 2002 French Legislative and Presidential Elections. *European Journal of Political Research* 45 (4), pp. 667-97.

Laver, Michael, and John Garry. 2000. Estimating Policy Positions from Political Texts. *American Journal of Political Science* 44 (3), pp. 619-34.

Laver, Michael, and W. Ben Hunt. 1992. *Policy and Party Competition*. New York: Routledge.

Lederer, Marianne. 1994. *La traduction aujourd'hui : le modèle interprétatif*. Vanves: Hachette F.L.E.

Lowe, Will. 2008. Understanding Wordscores. *Political Analysis* 16 (4), pp. 356-71.

Magin, Raphael, Markus Freitag, and Adrian Vatter. 2009. Cleavage Structures and Voter Alignments within Nations. *Zeitschrift für Vergleichende Politikwissenschaft* 3 (2), pp. 231-56.

Mair, Peter. 2001. Searching for the positions of political actors: a review of approaches and a critical evaluation of expert surveys. In M. Laver (ed.), *Estimating the Policy Positions of Political Actors*. New York: Routledge, pp. 10-29.

Marks, Gary, Liesbet Hooghe, Marco R. Steenbergen, and Ryan Bakker. 2007. Crossvalidating Data on Party Positioning on European Integration. *Electoral Studies* 26 (1), pp. 23-38.

Martin, Lanny W., and Georg Vanberg. 2008. A Robust Transformation Procedure for Interpreting Political Text. *Political Analysis* 16 (1), pp. 93-100.

McCallum, Andrew, and Kamal Nigam. 1998. "A Comparison of Event Models for Naive Bayes Text Classification." In *AAAI-98 Workshop on Learning for Text Categorization*.

Monroe, Burt L., Michael P. Colaresi, and Kevin M. Quinn. 2008. Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis* 16 (4), pp. 372-403.

Proksch, Sven-Oliver, and Jonathan B. Slapin. 2009. Position Taking in European Parliament Speeches. *British Journal of Political Science* First View, pp. 1-25.

Rabinowitz, George, and Stuart Elaine MacDonald. 1989. A Directional Theory of Issue Voting. *The American Political Science Review* 83 (1), pp. 93-121.

Ray, Leonard. 2001. A Natural Sentence Approach to the Computer Coding of Party Manifestos. In M. Laver (ed.), *Estimating the Positions of Political Actors*. New York: Routledge, pp. 149-61.

———. 2007. Validity of measured party positions on European integration: Assumptions, approaches, and a comparison of alternative measures. *Electoral Studies* 26 (1), pp. 11-22.

Sartori, G. 2005 [1976]. *Parties and Party Systems. A Framework for Analysis*. Colchester: ECPR Press.

Slapin, Jonathan B., and Sven-Oliver Proksch. 2008. A Scaling Model for Estimating Time-Series Party Positions from Texts. *American Journal of Political Science* 52 (3), pp. 705-22.

Vinay, J.-P., and J. Darbelnet. 2003 [1977]. *Stylistique comparée du français et de l'anglais : méthode de traduction*. Laval: Groupe Beauchemin.
Volkens, Andrea. 2007. Strengths and weaknesses of approaches to measuring policy positions of parties. *Electoral Studies* 26 (1), pp. 108-20.
Whitefield, Stephen, Milada Anna Vachudova, Marco R. Steenbergen, Robert Rohrschneider, Gary Marks, Matthew P. Loveless, and Liesbet Hooghe. 2007. Do expert surveys produce consistent estimates of party stances on European integration? Comparing expert surveys in the difficult case of Central and Eastern Europe. *Electoral Studies* 26 (1), pp. 50-61.
Yang, Christopher C., and Kar Wing Li. 2003. Automatic Construction of English/Chinese Parallel Corpora. *Journal of the American Society for Information Science and Technology* 54 (8), pp. 730-42.

Appendix

Table 1: Canadian Expert Survey 2008, 0-10 left-right scale

Party	**Mean both**	**Conf. interval**	**Mean English**	**Conf. interval**	**Mean French**	**Conf. interval**
Conservative	8.0	7.9-8.2	7.9	7.7-8.1	8.4	8.0-8.8
Liberal	5.3	5.0-5.6	5.2	5.0-5.4	5.5	5.2-5.8
NDP	2.9	2.7-3.1	2.9	2.7-3.1	3.0	2.5-3.6

Number of respondents = 163

Figure 1: English Left-Right Confidence Intervals of Canadian Manifestos with Wordscores

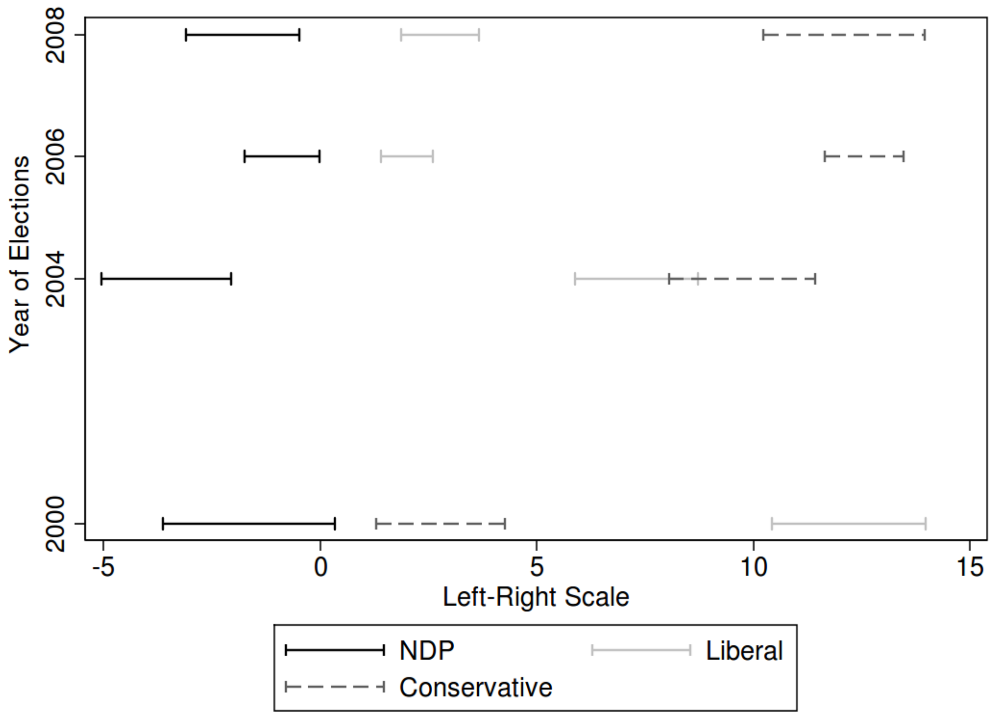

Figure 2: French Left-Right Confidence Intervals of Canadian Manifestos with Wordscores

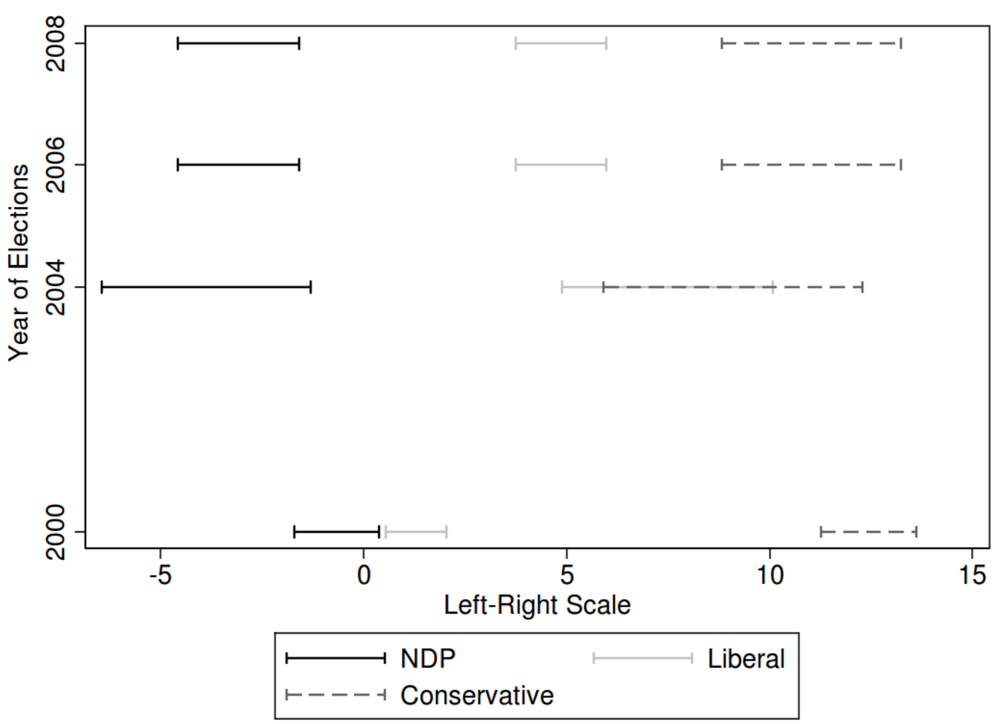

Figure 3: English Left-Right Scores of Canadian Manifestos with Wordfish

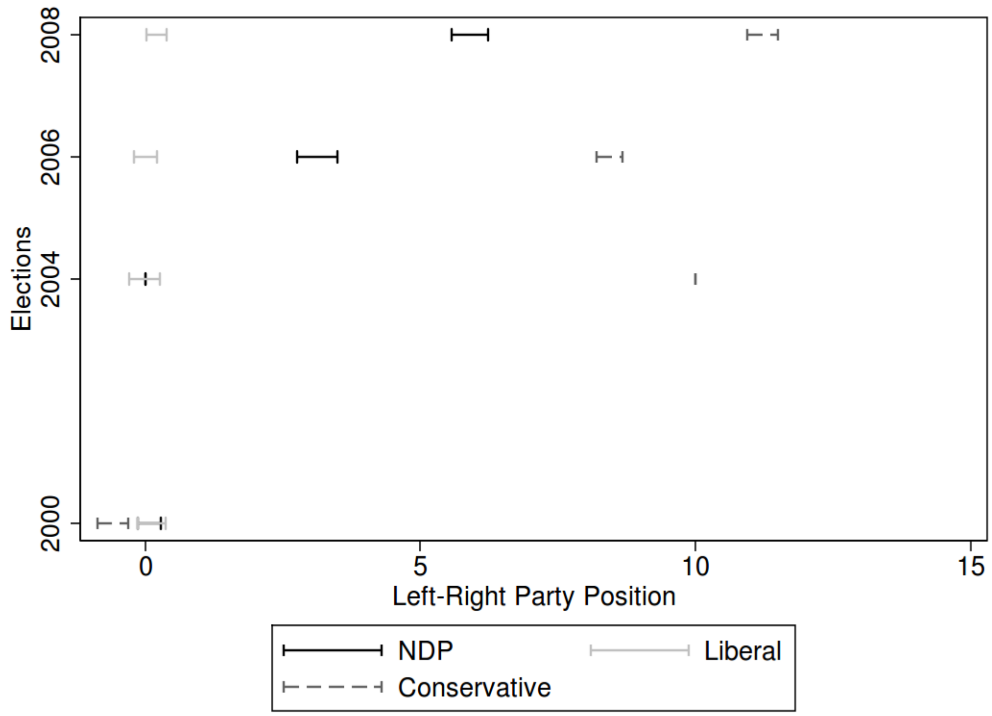

Figure 4: French Left-Right Scores of Canadian Manifestos with Wordfish

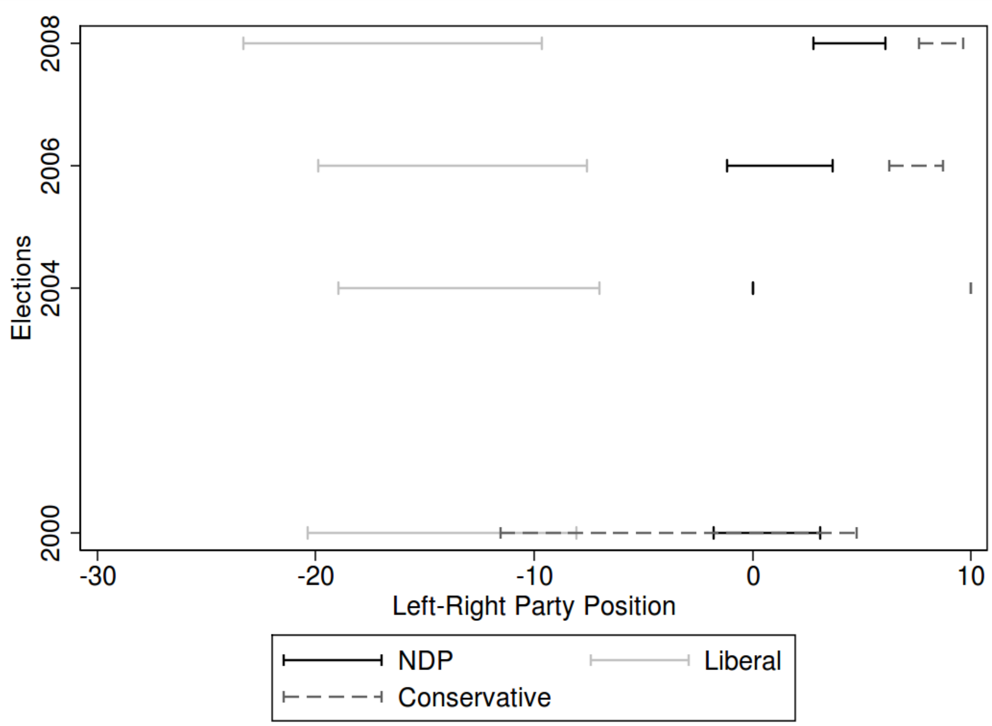

Figure 5: English Left-Right Scores of Canadian Stemmed Manifestos with Wordfish

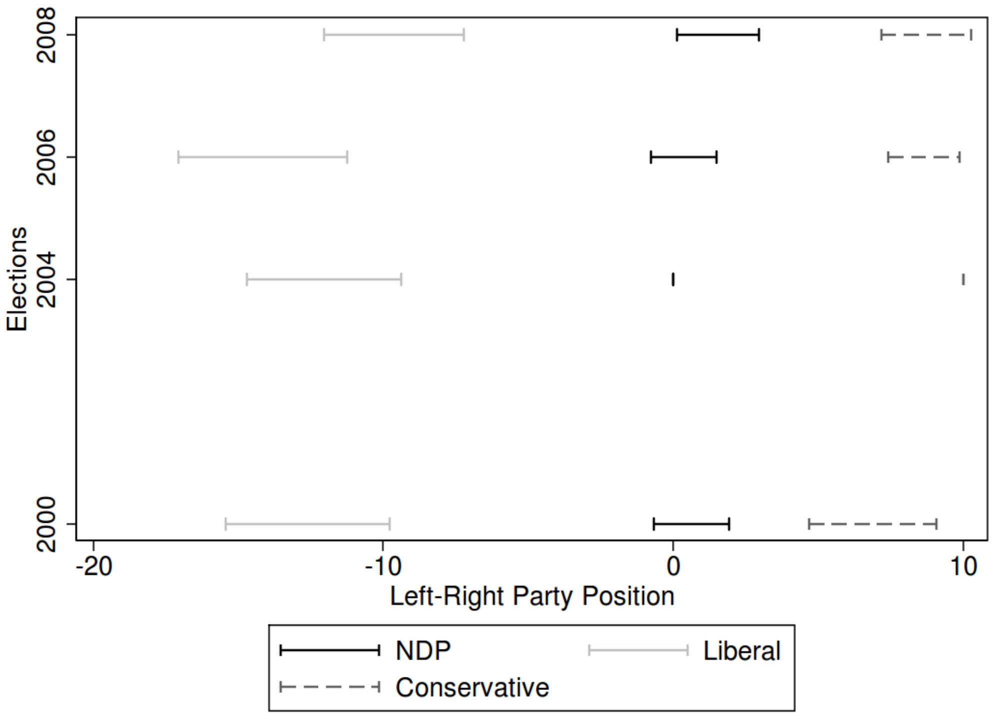

Figure 6: French Left-Right Scores of Canadian Stemmed Manifestos with Wordfish

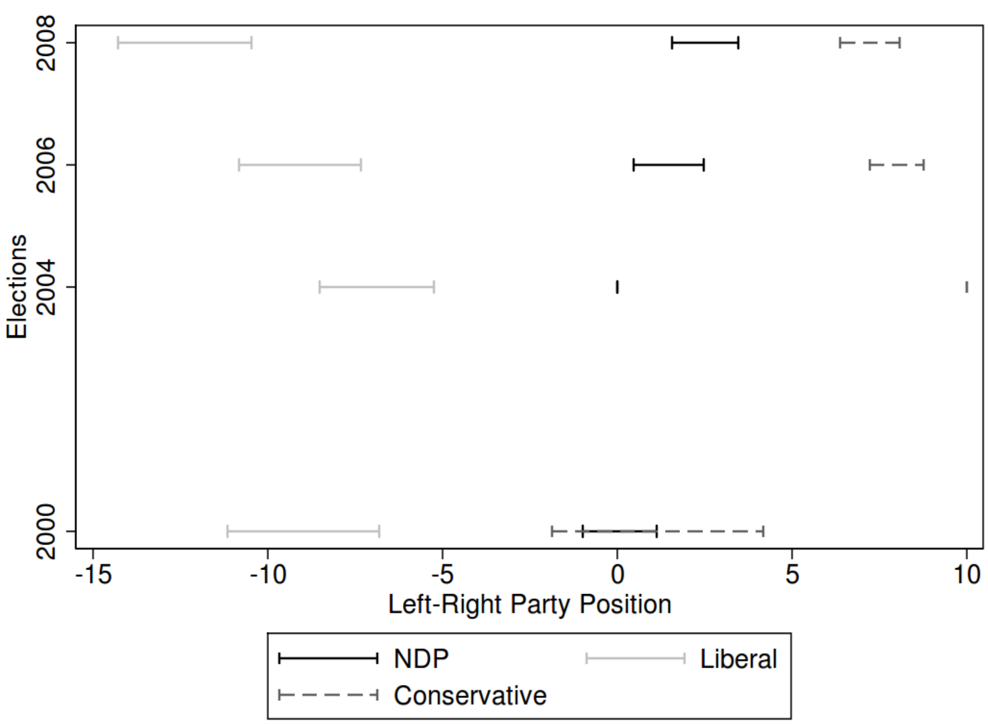

Table 2: Wordfish Document Scores Correlations

	English (Stem)	English	French (Stem)	French
English (Stem)	1.0000			
English	0.7020	1.0000		
French (Stem)	0.9390	0.8009	1.0000	
French	0.9242	0.8154	0.9843	1.0000