Computational Identification of Ideology in Text: A Study of Canadian Parliamentary Debates

Yaroslav Riabinin
Dept. of Computer Science, University of Toronto,
Toronto, ON M5S 3G4, Canada

February 23, 2009

In this study, we explore the task of classifying members of the 36$^{\text{th}}$ Canadian Parliament by ideology, which we approximate using party membership. Earlier work has been done on data from the U.S. Congress by applying a popular supervised learning algorithm (Support Vector Machines) to classify Senatorial speech, but the results were mediocre unless certain limiting assumptions were made. We adopt a similar approach and achieve good accuracy — up to 98% — without making the same assumptions. Our findings show that it is possible to use a bag-of-words model to distinguish members of opposing ideological classes based on English transcripts of their debates in the Canadian House of Commons.

1 Introduction

Internet technology has empowered users to publish their own material on the web, allowing them to make the transition from readers to *authors*. For example, people are becoming increasingly accustomed to voicing their opinions regarding various products and services on websites like Epinions.com and Amazon.com. Moreover, other users appear to be searching for these reviews and incorporating the information they acquire into their decision-making process during a purchase. This indicates that modern consumers are interested in more than just the facts — they want to know how other customers *feel* about the product, which is something that companies and manufacturers cannot, or will not, provide on their own.

Although no monetary transaction takes place when we cast our vote, one could argue that the opinions of others are just as important to us in the *political* marketplace. In order to make a responsible decision regarding the electability of a particular candidate, voters must look beyond appearances and be able to judge the *character* of the politician in question. This includes evaluating their intelligence and leadership abilities, but it also involves learning about the candidate's stance on various issues. However, many people forego this process and turn to the Internet for instant answers. By relying on public opinion — which is expressed in political blogs and discussion groups and so on — users put their faith in other voters to get it right.

Of course, politicians have their own opinions, and they normally act in a manner that is consistent with their *ideology*. Some researchers have argued that identifying a person's underlying "belief system" gives us insight into the views and attitudes of that individual and allows us to predict their outlook on a variety of issues (B. Yu, Diermeier, & Kaufmann, 2008). In politics, this information could be quite valuable — for example, it might help voters make more informed decisions by exposing the *true* beliefs of a particular candidate, which are often obscured by ambiguous campaign promises and the use of deceitful language.

However, identifying someone's political ideology may be a complicated task. Researchers have noted that while it is usually easy to determine which political party a person belongs to, the ideology of that individual is not directly observable (Diermeier et al., 2007). For example, there may be members of the Liberal Party of Canada who have conservative views on certain issues, which means that voters cannot rely on party affiliation alone to decide which politician will uphold the same beliefs as them. Therefore, a computational analysis of the words spoken by a political figure, especially in the context of a discussion or debate, will be particularly useful for this task, since it can yield unique insights about a person's underlying belief system.

In this work, we lay the foundation for a broader understanding of political opinions using natural language processing techniques. Our goal is to determine a person's political leanings from transcripts of their parliamentary speeches. To do this, we build on the research of Diermeier et al. (2007), which will be discussed in more detail in the next section. Briefly, the authors use supervised learning techniques to classify U.S. Senators as either Liberal or Conservative, based on what they say in Congress. Our approach differs from theirs in that we classify politicians with respect to *party affiliation*, which is an approximation of ideology. However, we believe that our methodology is valid, since the results of Diermeier et al.'s study indicate that there is a high correlation between party membership and ideology. In fact, errors in the classification would raise the question of whether the individuals that are misclassified are "true" Liberals or Conservatives, since the language they use does not identify them with other members of the party that they belong to.

Moreover, we explore the classification task in a different setting — the Canadian parliament, which is a political system that is fundamentally distinct from that of the United States. This allows us to investigate whether methods that are successful on U.S. data will produce the same results when tested on their Canadian counterparts. Specifically, given the dynamics of the parliamentary system — the fact that there is one party in power and another party that is the official opposition — we consider the potentially confounding effect this arrangement has on the classification. We expect that this factor will have the greatest impact on the results, although there may be other differences between the two political systems.

The rest of this paper is organized as follows. Section 2 provides an overview of relevant work on sentiment analysis and describes prior research on political opinion classification in particular. Section 3 discusses the central ideas of this study and the approach we take. Section 4 gives a detailed account of how the study is conducted. In Section 5, we present the outcome of our tests and offer possible intepretations of the results. In Section 6, we summarize the contributions made by this study and suggest future research directions.

2 Background and Related Work

Given the growing popularity of online reviews for movies, hotels, restaurants, automobiles, and so on, there is a pressing need for software that will help users make sense of all the available data. The computational treatment of such evaluative text is discussed in great detail by Pang and Lee (2008) in a recent survey of *opinion mining* and *sentiment analysis*. Notably, they emphasize that the task of extracting opinions is often reduced to many *classification* sub-problems — for example, one might first need to classify a sentence or paragraph as expressing any opinion at all (p. 24). What follows is a brief discussion of some of these approaches. Then, we present a more comprehensive account of existing work on the computational analysis of *politically-oriented* text.

2.1 Analyzing Sentiment

A fundamental task in this area of research is to separate *facts* from *opinions* and then determine whether the opinions in question convey positive or negative sentiment. Yu and Hatzivassiloglou (2003) show that it is possible to successfully classify documents as mostly *subjective* or mostly *objective*, using a common supervised machine-learning algorithm — Naïve Bayes. They achieve up to 97% precision and recall (F-measure) on a test set of 4,000 articles from the Wall Street Journal of type *News*, *Business*, *Editorial*, and *Letter to editor*. They use an equal number of such articles to train the classifier.

In the same (2003) work, Yu and Hatzivassiloglou also explore the task of determining the *polarity* of opinions at the sentence-level. They classify sentences as expressing positive, negative, or neutral sentiment and achieve up to 90% accuracy on a manually-annotated test set of 400 sentences. They require no training data, however, since they use a *content-based* method that consists of looking at the number and strength of semantically-oriented words in the sentence. Specifically, they begin with a *seed* set of known positive and negative words and calculate their co-occurrence with all the words in the given sentence to build a larger "affect dictionary", which can then be used to estimate the general sentiment of the text. In recent years, some researchers have developed more efficient ways of constructing such *emotional lexicons* — for example, by mining blogs for information about relationships between words and emotions (Yang et al., 2007), or by using a massive collection of HTML documents for this purpose (Nobuhiro & Kitsuregawa, 2007).

An alternative to using *opinion words* to measure sentiment polarity is to use a supervised learning algorithm. Mullen and Collier (2004) investigate the effectiveness of this approach by performing 10-fold cross-validation experiments on 1,380 Epinions.com movie reviews. They classify texts as positive or negative using Support Vector Machines (SVMs)[1] and report an accuracy of 83.5% with the standard "bag-of-words" model, where the features are plain unigrams. A slightly higher accuracy of 85.7% is achieved if *lemmatized* unigrams are used instead. Although the authors report

[1]This method will be described in more detail in Section 4.3.

that the best results are produced by a Hybrid SVM model, where lemmas are combined with additional real-valued features, the highest accuracy achieved by the classifier is 86%, which is hardly an improvement over the bag-of-lemmas model. These findings support the conclusion that supervised learning algorithms perform well on the sentiment polarity classification task, even when a basic unigram model is used.

2.2 Political Opinion Mining

We now consider the computational treatment of text that is of a *political* nature.

2.2.1 Counting the Words

In a popular blog entitled *Wordwatchers*, James Pennebaker, a distinguished psychologist, tracked the words used by Democratic and Republican candidates in the 2008 U.S. Presidential election. He attempts to answer the following question: what can we learn about the personality and governing styles of the politicians from what they say in public speeches, interviews, and debates? The methodology behind this approach is simple: word tokens are counted and grouped into categories (e.g., function words) using software that was developed specifically for this purpose.[2] Once the numbers are computed, the results can be interpreted. For example, Pennebaker concludes that since John McCain used first person singular (*I*, *me*, *my*) more than Barack Obama, this might signal an openness and honesty about the Republican candidate (Pennebaker, 2008).

In regards to Canadian politics, a similar strategy was employed by Skillicorn and Little (2006) in their coverage of the 2006 Canadian Federal election. They analyzed speeches of the three English-speaking party leaders at the time (Stephen Harper, Paul

[2] This program is called the Linguistic Inquiry and Word Count (Pennebaker et al., 2007) and will be described further in Section 4.4, since it is used in our research.

Martin, and Jack Layton), in order to determine who had the most *spin* — text whose apparent meaning is not the true beliefs of the person saying or writing it. Consequently, by raising the question of which candidate is more *trustworthy*, the issue of electability was addressed rather directly in their work. However, their research was based on only four simple assumptions, which were derived from Pennebaker's psychological model of deception in text — e.g., a smaller usage of first-person pronouns indicates greater spin. Therefore, given the questionable foundation of their methodology, their results should be treated with an appropriate amount of skepticism.

The work of Laver et al. (2003) can be seen as another example of how basic natural language processing techniques can be applied to political text. The authors examine the manifestos of several British and Irish parties in 1992 and calculate the relative frequencies of all the words present in the data. Having *a priori* knowledge about the **social** and **economic** policies of these "reference" texts, they look at the 1997 manifestos of the same parties and attempt to extract their policy positions simply by comparing word frequencies. Moreover, Laver et al. apply this technique to legislative speeches made by Irish party members in 1991 to estimate their individual positions on the "pro- vs. anti-government" dimension. Although the authors report that their method was successful on both of the aforementioned tasks, they fail to support this conclusion with any standard evaluation measures, such as accuracy, precision, or recall.

2.2.2 Informal Politics

Many Internet users voice their personal opinions regarding political issues by maintaining a blog or by participating in online debates. Mullen and Malouf (2006) examine this form of *informal discourse* and test the effectiveness of standard text classification methods for predicting the party affiliation of users, based on their posts. They use a Naïve Bayes classifier to label 185 members of the `www.politics.com` discussion group as either *left* (Democrats and liberals) or *right* (Republicans and conservatives). They perform 10-fold cross-validation experiments on the dataset and achieve an accuracy of 60.37%, which is a modest improvement over the 51.89% majority baseline. They conclude that traditional word-based methods are inadequate for the task of political sentiment analysis and propose that it might be fruitful to exploit the *quoting relationships* between posters. This suggestion is based on the observation that users tend to quote other users who are at the *opposite* end of the political spectrum.

In subsequent work (2007, 2008), Mullen and Malouf attempt to improve their classifier's performance on the same task by incorporating information about the *social properties* of the online discussion community into their algorithm. Specifically, they construct a graph representing the citation patterns of individual posters and use it to cluster them into "teams". Combined with Naïve Bayes, this approach yields an accuracy of up to 73%, which is significantly higher than the previous result. This seems to indicate that exploiting the relationships between discourse participants is a worthwhile endeavour.

2.2.3 Electronic Rulemaking

The goal of electronic rulemaking (*eRulemaking*) is to use technology to facilitate the process of creating and adopting new government regulations, and to increase public participation in all aspects of this activity. Some researchers in this field have worked on developing methods that would help rule-writers analyze a large number of *public comments* on proposed legislation (Kwon et al., 2006). The authors extract various pieces of information from the text, such as the topic of discussion, the argument structure, and the opinions being expressed. In another study, Kwon et al. (2007) focus on the problem of identifying the main claim of the writer and classifying it as "support", "oppose", or "propose a new idea". Using a supervised machine learning method called *BoosTexter*, they achieve a significant improvement over the baseline on both of these tasks.

2.3 Towards Ideology Extraction

In order to gain more insight into the political attitudes of U.S. Congressmen, Thomas et al. (2006) examine the speeches made by members of the House of Representatives in 2005. Specifically, they address the problem of determining, from transcripts of Congressional floor-debates, whether the author of each "document" (continuous single-speaker segment of text) supports or opposes the proposed piece of legislation that is under discussion. They use a minimum-cut classification framework that combines SVMs with information about *speaker agreement* (similar to Mullen and Malouf's (2007) work on exploiting relationships between discourse participants). This approach yields an accuracy of around 70% on a test set of 860 speech segments (grouped by debate, with 10 debates in total), which is a modest improvement over the 58% majority baseline. Greene (2007) reports a statistically significant increase in performance on the same task, obtaining an accuracy of up to 74.19% using an algorithm he developed for detecting *implicit sentiment* in text. Bansal et al. (2008) incorporate information about speaker *disagreement* into the same framework that was adopted by Thomas et al. in their original work. This addition increases the accuracy of the classifier to 78%.

However, the task of uncovering the underlying belief system of an individual is *not* equivalent to the sentiment polarity classification problem described above. Lin et al. (2006) note the difference between the two:

> A positive or negative opinion toward a particular movie or product is

> fundamentally different from an overall perspective. One's opinion will change from movie to movie, whereas one's perspective can be seen as more static, often underpinned by one's ideology or beliefs about the world.

Following this claim, the authors attempt to use standard text classification methods — Naïve Bayes and SVMs — to identify the "ideological perspective" from which a given document is written. They look at 594 articles about the Israeli-Palestinian conflict, published on the `www.bitterlemons.org` website, and classify each article as being written from an Israeli or a Palestinian perspective. They achieve an accuracy of up to 99% on a subset of the data — articles that were written by Editors of the website, as opposed to Guests. In a related study (Lin & Hauptmann, 2006), the authors develop a statistical test for determining if two document collections convey *opposing* ideological perspectives. Although their methodology is less sophisticated and their evaluation is less rigorous, they examine several corpora, one of which consists of transcripts of three Bush-Kerry Presidential debates in 2004. Studying such data is an important step towards extracting the ideology of politicians based on their speeches.

Yu, Kaufmann, and Diermeier (2008) conduct a series of experiments to explore the *characteristics* of political text. Based on their findings, they also conclude that identifying sentiment is *not* sufficient for general-purpose political opinion classification. In a related work, the authors attempt to classify members of the U.S. Senate by *ideology* (Diermeier et al., 2007). They use SVMs to label each speaker as a Liberal or a Conservative and achieve up to 94% accuracy on a dataset consisting of 350 training documents and 50 test documents, where each document is a concatenation of all the speeches made by one Senator in a given time period (e.g., the 101st Senate).

However, in these experiments, the authors focus exclusively on "extreme" Senators — the 25 most conservative and the 25 most liberal ones in each Senate. If the task is to classify "moderate" Senators, the results are significantly worse (only up to 80% accuracy). Later work by Yu, Diermeier, and Kaufmann (2008) makes no distinction between moderates and extremes — rather, they rely on *party affiliation* for the "truth" about the political views of the individuals in question. The goal of their study is to examine the classifier's person- and time-dependency by using speeches from both the Senate and the House of Representatives and by comparing the results. They find that party classifiers trained on House speeches can be generalized to Senate speeches of the same year, but not vice versa. They also observe that classifiers trained on House speeches perform better on Senate speeches from recent years than older ones, which indicates the classifiers' time-dependency.

3 Classification by Party Membership

It is clear from our discussion of related work that the task of analyzing political text is a difficult one to perform computationally. Some researchers have even put forward the hypothesis that the language used in political discussions does not identify the affiliation of the author, since both sides are likely to be using largely the *same* vocabulary (Mullen & Malouf, 2006). However, others have argued that despite all the challenges faced by natural language processing techniques, success on this task can be achieved due to the fact that people with different perspectives tend to *emphasize different words* from a shared vocabulary (Lin et al., 2006).

Our research focuses on the task of distinguishing Liberal politicians from Conservative ones, based on their speeches in the Canadian Parliament. We view this as a binary classification problem and use SVMs to identify features that are most indicative of each ideology. Diermeier et al. (2007) explore the same task, but in a different setting — the U.S. Congress — and with several important assumptions.

First, as mentioned in Section 2.3, they use a ranking system[3] to establish the "ground truth" about Senators — the direction and strength of their political leanings (e.g., "extreme" Liberal) — and group them according to this measure. In our work, we treat Members of Parliament (MPs) as simply "Liberal" or "Conservative", depending on which political party they belong to. This makes our task more difficult, because there is a greater diversity of views within each party.

Second, there is considerable overlap between the **train** and **test** portions of their dataset, since they extract content from multiple Senates (101st - 108th) and since members of Congress tend to preserve their beliefs over time. Specifically, 44 of the 50 "extreme" Senators in their test set are represented in the training data, which means that the classifier is already trained on speeches made by these particular individuals. The trouble with this fact is that the classifier might be learning to discern speaking styles, rather than ideological perspectives. In order to avoid the bias this approach may introduce, we focus on one time period — the 36th Parliament — so that there is a one-to-one mapping between MPs and documents in our dataset. In other words, each training and test document is a concatenation of all the speeches made by a *unique* speaker, such that no other document contains text spoken by that person.

Another difference between our work and previous attempts to classify politicians by ideology has to do with the *properties* of the speech being analyzed. The goal of Diermeier et al.'s study was to test the hypothesis that low dimensionality in voting[4] is explained by institutional constraints, such as party leadership. For this purpose, they chose to examine Senatorial speech, because members of the U.S. Senate can speak out

[3] DW-NOMINATE scores, available at `http://voteview.com/dwnomin.htm`.

[4] This means that voting patterns could be explained using just one or two dimensions, such as the traditional left-right dimension that is associated with the government's role in the economy and economic re-distribution.

of turn and discuss matters that are completely unrelated to the topic at hand. The authors claim that this is an ideal setting for assessing whether the voting patterns of Senators are shaped by agenda control or by an underlying ideology that is shared by members of the same party.

In contrast, we have elected to focus on a different type of text — transcripts of debates in the Canadian House of Commons. Generally speaking, adherence to party doctrine is much more evident in Parliamentary systems. For example, when it comes to *voting*, there is a strict division between government and opposition members, rather than between those on the Left and on the Right of the political spectrum. What this means is that opposition members tend to vote against the government to signal their *opposition*, rather than their discontent with a particular proposal (Godbout & Hoyland, 2008). From this it follows that parliamentary *speech* is also likely to be highly partisan, which can make our task easier. We test this hypothesis by examining debates that take place in the context of the **Oral Question Period** — a time when the Opposition can hold the Government accountable for its actions and policies. Hence, these speeches tend to be opinionated and address a wide range of topics, such as the economy and the enviroment (see Table 1 for an example). It should also be mentioned that the Canadian Parliament is bilingual, such that French or English could be spoken at any time and speakers can switch languages whenever they choose. Consequently, the proceedings include a professional translation of everything that is said into the other official language.

However, this form of discourse raises some concerns about the behaviour of the classifier. For example, in the 36$^{\text{th}}$ Parliament, the Conservatives occupy the role of the Opposition party and the Liberals form the Government. As a result, it is not clear whether a classifier that is trained on such data would be detecting ideology (Liberal

Table 1: An exchange between a Conservative and a Liberal MP.

THE ECONOMY **Mr. Monte Solberg (Medicine Hat, Ref.):** Mr. Speaker, first it was the Prime Minister's in-law, Paul Desmarais, who said that high taxes were strangling Canada's economy. Today Doug Young, his former cabinet minister, chaired a whole conference on plummeting Canadian productivity. At the conference the Prime Minister's own pollster admitted that Canadians are upset with our declining standard of living, and the weak dollar proves it. If top Liberals do not buy the Prime Minister's low dollar-high tax argument, then why should the rest of us? **Right Hon. Jean Chrétien (Prime Minister, Lib.):** Mr. Speaker, the Canadian people are quite happy with the economic policies of this government. The Canadian people are very happy that we have taken unemployment from 11.4% to 7.8%. The Canadian public is quite happy about the fact that we have reduced the Conservative's deficit from $42 billion to zero and we are still going. The Canadian people are pretty happy to see that the Financial Times of London has called Canada the top dog of financial managers.

vs. Conservative) or party status (Government vs. Opposition). Moreover, the particular *format* of the Oral Question Period may also introduce bias into the classification. During this part of the daily debates — and regardless of who is in power — members of the opposition parties, as well as some members of the governing party, pose questions to the Cabinet Ministers[5] and their representatives, usually asking them to explain their stance on issues that are within the realm of their responsibility. Therefore, our ideological classifier may be learning — at least in part — to distinguish questions

[5]These are members of the the governing party that have been selected by the Governor General, on the advice of the Prime Minister, to be responsible for some aspect of government — e.g., Minister of Finance.

from answers, which is ultimately not the goal of our research.

We address these issues in our work and attempt to determine what effect, if any, these properties of the Oral Question Period have on the results of the classification. To do this, we extract additional data from the **Government Orders** portion of the House of Commons proceedings. During this time period, MPs debate bills and motions that are put on the agenda by the governing party. These speeches tend to be much greater in length and, most importantly, their format is uniform for all participants. We perform the same classification experiments with this new dataset in order to test whether the change in discourse affects the accuracy of the classifier. We also combine the original Oral Question Period data with the new speeches to create a third dataset, which we use to further explore the classification task.

4 Materials and Methods

We now describe the methodology behind our study.

4.1 Data Preparation

Some researchers have used transcripts of Canadian parliamentary debates in their work on *machine translation* (Brown et al., 1990; Fraser & Marcu, 2007), since the corpus is bilingual. However, our goal is to *classify* text by ideology, which is why we focus exclusively on the English portion of the data. This means that we make no distinction between a speaker's **original** words in English, and the **translation** of a speaker's speeches from French.

Our first dataset (referred to as **OQP-Speakers**) consists of 200 documents, where each *document* is a concatenation of all the speeches made by a unique speaker during the Oral Question Period (OQP), over the course of the 36$^{\text{th}}$ Parliament (September

22, 1997, to October 22, 2000). Of these 200 speakers, 79 are Conservative MPs and 121 are Liberal MPs (see Table 16 and Table 17 in Appendix A for the names of these individuals). Initially, we extracted speeches for 149 Liberals, but we discarded 28 with the lowest word counts, in order to avoid having documents with too few words. The total number of words for the Conservatives in this dataset is 487,000, and 885,231 for the Liberals.

In addition to this speaker-based dataset, we also extracted speech *segments* — uninterrupted pieces of text spoken by a single individual — for all Liberal and Conservative MPs who participated in the OQP, and placed them in separate documents, with no regard for who the author of the speech is. After removing 428 documents with the lowest word counts, 20,000 documents remained: 6,666 Conservative speeches and 13,334 Liberal speeches.[6] This dataset will be referred to as **OQP-Segments**. The total number of words for the Conservatives in this dataset is 541,605, and 996,261 for the Liberals.

Our second speaker-based dataset (referred to as **GOV-Speakers**) also consists of 200 documents, divided into 79 Conservative MPs and 121 Liberal MPs, but these speeches were extracted from the Government Orders (GOV) portion of the House of Commons debates. Similarly, although 150 Liberals spoke during this time, we discarded 29 of them, based on their word counts. Moreover, it should be noted that all the Conservative MPs in this dataset are the same as in the OQP data, but 20 out of the 121 Liberals are different. The total number of words for the Conservatives in this dataset is 3,444,315, and 2,145,174 for the Liberals.

Our third speaker-based dataset (referred to as **OQP+GOV**) consists of the same 200 speakers as in the OQP-Speakers dataset. However, their speeches in the Oral

[6]These speeches are quite short, each one generally less than 150 words.

Question Period have been combined with their Government Orders speeches. It should be noted that three of the Liberal MPs in the OQP-Speakers dataset *did not* speak during the Government Orders period, which means that in this dataset their speeches have not been augmented with additional (Government Orders) data. The total number of words for the Conservatives in this dataset is 3,931,237, and 2,662,437 for the Liberals. Please see Appendix A for further information about the corpus.

4.2 Document Representation

For the purposes of our classification task, documents containing speeches are represented as vectors in an n-dimensional space, where each dimension corresponds to some feature, such as *word type* (this is known as the "bag of words" model). In order to do this, we first converted all letters to lowercase and expanded all clitics (e.g., "I'm" becomes "I am" and "won't" becomes "will not"). Then, we extracted strings of consecutive alphabetic characters as valid words, ignoring all punctuation marks (this process is called *tokenization*). We also experimented with *stemming* — removing word suffixes[7] — and using alphanumeric character strings as valid words (i.e., words *and* numbers). Therefore, we have the following feature sets: **words**, **word stems**, and **words+numbers**.

The value of each feature for a given document was calculated by counting the number of times it occurs in the text. Specifically, we use four weighting schemes: **bool** (presence-of-feature), **tf** (term frequency), **tf-norm** (term frequency normalized by document length), and **tf-idf** (term frequency—inverse document frequency). A commonly used measure in document processing tasks is *tf-idf*. Its value is given by the following formula:

[7] We used the Porter Stemming algorithm (Porter, 1980) for this.

$$tf\text{-}idf = \frac{(tf/l)}{\log(n/df)} \tag{1}$$

where n is the total number of documents in the dataset, tf is the number of occurrences of the feature in the current document, which has length l, and df is the number of documents in which the feature occurs. The purpose of the denominator is to *minimize* the importance of features that occur in many documents (i.e., words that many people use).

In addition to this, we experimented with various word removal strategies to reduce the vocabulary size and to eliminate "stopwords" (frequently occurring words that typically contribute nothing to the classification). For example, we removed all instances of *Mr. Speaker* from the text.[8] We also discarded all features with $df < 5$ and $tf < 10$, as well as the top 500 most frequent features, across all documents (further information will be provided in the next section).

4.3 Support Vector Machines

Support Vector Machines (SVMs) are among the most widely used supervised learning algorithms for text classification. The reason for this is that they are perfectly suited for linearly separable problems with a high-dimensional input space (i.e., many features), such as those found in text categorization (see Joachims (1998) for more information about the theory behind SVMs). Briefly, the classifier is *trained* on a number of documents whose category membership is known.[9] During this process, it selects the features that are most indicative of a given category — in our case, Liberal or Conservative. Ultimately, this allows us to see which words characterize a political ideology,

[8] All MPs begin their speeches with this formulaic phrase, so the words *mr* and *speaker* are unlikely to be useful for discriminating Liberals from Conservatives.

[9] For our binary ("yes" or "no") classifcation task, we arbitrarily chose Conservatives to be the "positive" (+1) class and the Liberals to be the "negative" (-1) class.

as approximated by party membership. It also allows us to *rank* MPs that were put in the same class, in order from most representative of that class to least, to find out which MPs were classified as "more Liberal/Conservative" than others.

There are many implementations of the SVM algorithm, but in our study we use the SVM-light package (Joachims, 2008) with default parameter settings.

4.4 Linguistic Inquiry and Word Count

Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2007) is a program designed for the purpose of analyzing text along various dimensions of language. By counting the occurrences of particular words and word stems in over 70 categories, the LIWC determines the degree to which people use different types of words. These categories include: linguistic processes (pronouns, adverbs, prepositions, etc.), psychological processes (positive emotion, negative emotion, etc.), and personal concerns (work, money, religion, etc.).

We use this software in our work in order to check whether a different approach to the problem can yield good classification results. Specifically, we perform the SVM classification experiments with the LIWC categories as features. This reduces the size of the feature set from thousands of words to only a handful of linguistic dimensions.

4.5 Evaluation Measures

One of the most frequently used evaluation criteria is **accuracy**, which is defined as the percentage of correctly classified test instances. We use this as the main measure of our classifier's performance. Other commonly used measures include **precision** and **recall**[10]. However, since our decision to make the Conservatives the positive class was

[10]Precision and Recall are defined as the proportion of documents correctly assigned to the positive class among all the documents that were assigned to the positive class, and among all the documents

arbitrary, we calculate the averages of precision/recall scores for *both* classes (Liberals *and* Conservatives). For practical reasons, we only report accuracy in our work, since our results show that all three measures are very close — in fact, accuracy is roughly the mean of precision and recall.

5 Experimental Results

Speeches in the **Oral Question Period** and the **Statements by Members** portions of the House of Commons debates are organized by topic of discussion. Although we did *not* use Members' Statements as data in our classification experiments, we began our study by extracting the names of these topics from *both* sections, along with the frequency of their occurrence in the 36$^{\text{th}}$ Parliament. Table 2 lists the top 10 most frequent topics.

Overall, 1169 different topics were found in the OQP, and 2537 in the Members' Statements. However, most of these are infrequent. For example, less than 10% of the OQP topics are discussed with reasonable frequency (i.e., more than 10 times), and more than 75% of the Members' Statements topics are brought up only once in the entire corpus.[11] These results indicate that the Statements by Members section contains many speeches that are unlikely to reveal ideological differences between Liberals and Conservatives. Also, since MPs do not engage in debate with other MPs during this time period, we use only the Oral Question Period data for the speaker classification experiments, which will be presented next.

that truly belong to the positive class, respectively.

[11] Although, it is interesting to note that the most frequent topics in Statements by Members are very similar to those in the OQP, which might signal some consistency in the subject matter of the House of Commons debates.

Table 2: Discussion topics and their frequencies of occurrence.

Oral Question Period	**Statements by Members**
1. ABORIGINAL AFFAIRS (234)	1. AGRICULTURE (81)
2. TAXATION (217)	2. HEPATITIS C (48)
3. EMPLOYMENT INSURANCE (215)	3. THE SENATE (48)
4. HEALTH (215)	4. ABORIGINAL AFFAIRS (45)
5. HUMAN RESOURCES DEVELOPMENT (205)	5. THE ENVIRONMENT (42)
6. AGRICULTURE (188)	6. THE BUDGET (41)
7. NATIONAL DEFENCE (178)	7. TAXATION (40)
8. FISHERIES (166)	8. HEALTH CARE (39)
9. HEPATITIS C (152)	9. VIOLENCE AGAINST WOMEN (38)
10. THE ENVIRONMENT (141)	10. JUSTICE (37)

5.1 Classifying Speakers in the Oral Question Period

Given the relatively small size of the OQP-Speakers dataset — only 200 documents — we performed 5-fold cross-validation[12] on the data, with default SVM parameter settings at all times. The overall accuracy was calculated by taking the average of the five accuracy scores produced by each iteration of the experiment. In total, we tested 12 different SVM methods: 4 weighting schemes (bool, tf, tf-norm, tf-idf) $\times$ 3 feature sets (words, word stems, words+numbers) $\times$ 1 word removal strategy (500; 10; 5)[13]. The results of these experiments are shown in Table 3.

These numbers indicate that incorporating stemming and numbers into the feature set does *not* yield any noticeable improvements over plain unigrams (words). In fact, although stemming reduces the size of the vocabulary, it can also be harmful when

[12] The dataset was divided into 5 *balanced* groups of 40 MPs: groups 1–4 contain 24 Liberals and 16 Conservatives, and group 5 contains 25 Liberals and 15 Conservatives. The experiment was repeated 5 times, and each time a different group was "held out" as the test set, while the other groups formed the training set.

[13] This notation is read as follows: remove the top 500 most frequent features; remove features with *tf* < 10; remove features with *df* < 5.

Table 3: 5-fold cross-validation on the OQP-Speakers dataset.

	words	word stems	words + numbers	avg. accuracy
bool	91.5	92.5	91.5	**91.83**
tf	77.5	78.0	79.5	**78.33**
tf-norm	96.0	96.5	96.5	**96.33**
tf-idf	96.0	95.0	95.5	**95.50**
avg. accuracy	**90.25**	**90.50**	**90.75**	

conducting a feature analysis, since it strips off potentially important inflectional and derivational morphemes from words.

Although the differences in performance between the four weighting schemes are likely not significant, the lowest accuracy is achieved by the **tf** weighting scheme, followed by **bool**. This result could be explained by the fact that both of these methods are sensitive to *word count*. Specifically, Conservatives with low word counts are considered "less conservative" by the classifier, and thus tend to be labeled as Liberals. However, the opposite is true for Liberals — the higher their word count, the more likely they are to be misclassified as Conservatives. For example, Jean Chrétien has the highest word count in the entire OQP-Speakers dataset, but he is labeled as a Conservative by the classifier.

This is not the case for the **tf-norm** and **tf-idf** weighting schemes, which appear to be unaffected by word count, as evidenced by the fact that the Liberals and Conservatives that are misclassified using these two methods have varying word counts. The highest accuracy is achieved by the normalized frequency of features (tf-norm) weighting scheme. However, this method outperforms *tf-idf* by a very small margin — less than 1% — when the accuracy is averaged over all three feature sets (the results are identical for the **words** feature set).

Hence, in the remainder of this section, we will focus on the analysis of classifi-

Table 4: MPs misclassified by the *tf-idf*/words/(500; 10; 5) method.

Conservatives (falsely labeled as Liberals)	**Liberals** (falsely labeled as Conservatives)
André Harvey Angela Vautour Charlie Power Diane St-Jacques Jean J. Charest Norman Doyle	Stan Keyes

cation results produced by the *tf-idf* weighting scheme, with words as features. This method yields an accuracy of 96%, which is a substantial improvement over the majority baseline of 60.5%. It is interesting to note that almost all of the errors come from the Conservative side (see Table 4 for the names of the MPs that have been falsely classified by this method). It is also worthwhile to consider the features that have been selected by the classifier as the most indicative of each ideology. Table 5 lists the top 50 most discriminative words for the Liberals, and Table 6 does the same for the Conservatives.[14]

Notice that the Liberal "lexicon" is characterized by words related to Québec (*french*, *francophonie*, *MAI*, *PQ*) and various social issues (*housing*, *violence*, *humanitarian*, *youth*, *society*, *technology*), while the Conservatives tend to focus on monetary concerns (*APEC*, *taxpayer*, *dollar*, *millions*, *paying*, *premiums*), aboriginal affairs (*native*, *indian*, *chief*), and, to a lesser degree, national defense (*military*, *marshall*). Also, notice that the Liberals use language that is generally positive (*congratulate*, *excellent*, *progress*) and that is intended

[14] All acronyms have been recovered to uppercase for the ease of reading. They are as follows: NDP: New Democratic Party; MAI: Montréal Arts Interculturels; PQ: Parti Québécois; HRDC: Human Resources Development Canada; APEC: Asia-Pacific Economic Cooperation; AIDA: Agricultural Income Disaster Assistance; EDC: Export Development Canada; HRD: Human Resources Development; CPP: Canada Pension Plan.

Table 5: Top 50 Liberal features for *tf-idf*/words/(500; 10; 5).

1. opposite	11. improve	21. collective	31. occasions	41. committed
2. housing	12. french	22. comment	32. various	42. society
3. violence	13. water	23. assistance	33. recommend	43. promote
4. operation	14. francophonie	24. assist	34. progress	44. technology
5. closely	15. NDP	25. agreements	35. youth	45. investment
6. humanitarian	16. wish	26. standing	36. correctional	46. suggest
7. discussions	17. parties	27. excellent	37. obligations	47. MAI
8. consultations	18. agri	28. congratulate	38. respond	48. relation
9. established	19. sector	29. developing	39. refers	49. PQ
10. inform	20. organization	30. repeat	40. dialogue	50. additional

Table 6: Top 50 Conservative features for *tf-idf*/words/(500; 10; 5).

1. HRDC	11. EDC	21. HRD	31. traffic	41. mismanage
2. APEC	12. dollar	22. ethics	32. pockets	42. actuary
3. blood	13. millions	23. bureaucrats	33. paying	43. compensate
4. newfoundland	14. indian	24. grant	34. starlight	44. cover
5. AIDA	15. justify	25. prison	35. CPP	45. premiums
6. convicted	16. tainted	26. patronage	36. admit	46. marshall
7. commit	17. columbians	27. resign	37. refusing	47. helicopter
8. port	18. u	28. waiting	38. per	48. ferry
9. taxpayer	19. plutonium	29. lists	39. b	49. shawinigan
10. native	20. military	30. failed	40. promise	50. chief

to create the appearance of the government working for the people (*established*, *inform*, *improve*, *assist*, *developing*, *promote*). In contrast, the Conservatives use negative words that are meant to call the government's competence into question (*justify*, *resign*, *failed*, *admit*, *refusing*, *mismanage*). This might be evidence of the "Government vs. Opposition" confound that was discussed in Section 3. Further support for this claim comes from the fact that the top Liberal word is *opposite*[15].

5.1.1 Omitting Jean Chrétien from the Corpus

As was mentioned earlier, Jean Chrétien is among the several Liberals misclassified when the **tf** weighting scheme is used. In fact, he is the *only* Liberal who is falsely labeled as a Conservative by the classifier with the **bool** weighting scheme. Given that he was the Prime Minister of Canada and the leader of the Liberal Party during the 36$^{\text{th}}$ Parliament, this is a curious result!

It could be the case that Jean Chrétien acts as a "middle party" of his own in the classification. Due to his high word count, he may be using not only a "Liberal" vocabulary, but words that characterize Conservatives as well, perhaps with considerable frequency in some instances. This might affect the performance of the classifier by blurring the line between ideologies, causing some Conservative MPs to be falsely labeled as Liberals. So, if he was omitted from the corpus, there should be a noticeable increase in the accuracy.

In order to test this hypothesis, we repeated the **bool** and **tf** experiments (with words as features), having removed Jean Chrétien from the OQP-Speakers dataset. However, the results show that the hypothesis is false, since there is absolutely no change in the accuracy, other than a slight improvement due to the Prime Minister's

[15] As in the following example: "Members *opposite* keep talking about the health and social transfers to the provinces. Let me try, as many of my colleagues have tried, to clarify this".

absence.

5.1.2 Removing (In-)Frequent Words

We experimented with various word removal strategies, in order to determine whether words that are normally eliminated from the feature set because of their frequency have any effect on the outcome of the classification. The most striking result of our investigation is that keeping more words improves the performance of the classifier. For example, if no words are removed, then an accuracy of 97.5% is achieved using the *tf-idf*/words method. In fact, the accuracy increases to 98% if words with $d_f < 5$ are removed.

Since the (0; 0; 5) removal strategy slightly outperforms (500; 10; 5), even fewer MPs are misclassified using this method.[16] Moreover, given that all of the most frequent words are kept in the feature set (see Table 18 in Appendix A for a list of the top 50 most frequent words), it might be the case that at least some of these words contribute to the classifer's success. A feature analysis reveals that this claim is correct. Table 7 shows the top 10 words that distinguish Liberals from Conservatives.

Despite some similarities with the *tf-idf*/words/(500; 10; 5) method, many of these features are different. Specifically, there appears to be evidence that the question-and-answer format of the Oral Question Period may be responsible for the improvement in accuracy. For the Liberals, the top words are *hon* and *member* (as in "the hon. member for Halifax West"), which is how an MP from the governing party typically addresses an MP that has posed a question in the debate. Also, the word *we* might be used by Liberals to speak on behalf of the entire party when responding to questions. For the Conservatives, the word *why* serves the obvious purpose of posing a question[17], and

[16] They are as follows: Norman Doyle, Angela Vautour, Diane St-Jacques, and André Harvey.
[17] Although, note that some Liberal MPs also ask questions.

Table 7: Top 10 Liberal and Conservative features for *tf-idf*/words/(0; 0; 5).

Liberals	**Conservatives**
1. hon	1. prime
2. member	2. liberal
3. we	3. why
4. bloc	4. finance
5. reform	5. solicitor
6. opposite	6. HRDC
7. housing	7. farmers
8. quebecois	8. her
9. quebec	9. he
10. party	10. hepatitis

the words *he* and *her* are likely used to refer to Liberal MPs that are the targets of the questioning. Again, note the usage of words such as *bloc*, *reform*, *opposite*, and *party* by the Liberals, and *prime* (as in "Prime Minister") by the Conservatives. This lends further support to the hypothesis that the classifier is partially learning to distinguish Government MPs from Opposition MPs.

5.2 Introducing the Government Orders Data

In order to measure the degree to which the format of the Oral Question Period affects the ideological classification results, we repeated the *tf-idf*/words experiments on the GOV-Speakers dataset, which contains speeches that were made during the Government Orders section.[18] We also combined these two datasets to form a third (OQP+GOV) and included it in our experiments.

Table 8 shows the results of 5-fold cross-validation on the aforementioned data, with two different word removal strategies, as well as the results of the OQP-Speakers

[18]The bulk of these speeches are on proposed legislation, with fewer restrictions on the debates than in the Oral Question Period.

Table 8: Classification results for the three 200-speaker datasets.

	GOV-Speakers	**OQP+GOV**	**OQP-Speakers**
(0; 0; 5)	86.5	89.5	98.0
(500; 10; 5)	85.5	88.5	96.0

experiments, for comparison. There is a noticeable drop in accuracy for the GOV-Speakers data, which might indicate that the OQP speeches make it easier to distinguish Liberals from Conservatives, due to factors *other* than ideology. Note also that the accuracy for the OQP+GOV dataset is slightly higher than GOV-Speakers, but still considerably less than OQP-Speakers. This result is not surprising, given that the GOV-Speakers dataset is much larger than OQP-Speakers.

While most of the OQP errors come from the Conservative side, this is *not* the case for the GOV-Speakers and OQP+GOV datasets: more Liberal MPs are falsely labeled as Conservatives than vice versa (see Table 9 for the names of these individuals). However, notice that several Conservative MPs are consistently misclassified as Liberals in all three datasets, which raises the question of what these "errors" might mean. For example, they may expose the true political beliefs of certain individuals. In fact, most of the Conservative MPs who eventually switched parties (see Table 15 in Appendix A) have been consistently misclassified as Liberals in our experiments. These MPs are: André Harvey, David Price, Diane St-Jacques, and Jean J. Charest. Also, notice that Angela Vautour — a Progressive Conservative — was commonly mistaken for a Liberal by our classifier in the OQP experiments, and she used to be a member of the NDP.

Tables 10 and 11 show the top 50 most discriminative Liberal and Conservative features for the GOV-Speakers data.[19] In comparison with the OQP results, we observe

[19]The acronyms in these tables are as follows: CHST: Canada Health and Social Transfer; CMHC: Canada Mortgage and Housing Corporation; NAFO: Northwest Atlantic Fisheries Organization; CWB: Canadian Wheat Board; NASS: National Agricultural Statistics Service; PC: Progressive Conservatives; PSAC: Public Service Alliance of Canada; DEVCO: Cape Breton Development Corporation; GST: Goods

Table 9: MPs misclassified by the *tf-idf*/words/(0; 0; 5) method.

	Conservatives (falsely labeled as Liberals)	**Liberals** (falsely labeled as Conservatives)
GOV-Speakers	André Bachand André Harvey David Chatters Diane St-Jacques Jean J. Charest John Herron Norman Doyle Peter MacKay	Alfonso Gagliano Carolyn Parrish Charles Hubbard David Iftody Eleni Bakopanos Gerry Byrne John Bryden John Cannis John Harvard John McKay John O'Reilly Marcel Massé Roger Gallaway Sheila Copps Stan Dromisky Stan Keyes Steve Mahoney Tom Wappel Wayne Easter
OQP+GOV	André Bachand André Harvey Charlie Power David Price Diane St-Jacques Jean J. Charest John Herron Norman Doyle	Charles Hubbard David Anderson David Iftody George S. Baker Gerry Byrne John Cannis John Harvard John O'Reilly Lawrence MacAuley Roger Gallaway Stan Keyes Steve Mahoney Wayne Easter

Table 10: Top 50 Liberal features for *tf-idf*/words/(500; 10; 5).

1. miramichi	11. guelph	21. producers	31. iraq	41. investments
2. innovation	12. quebecois	22. cape	32. lobsters	42. territorial
3. agri	13. values	23. southwestern	33. london	43. plans
4. customs	14. breton	24. partnership	34. simplistic	44. CWB
5. thornhill	15. labradorians	25. mackenzie	35. consultations	45. accessibility
6. rural	16. sex	26. CMHC	36. post	46. NASS
7. labrador	17. CHST	27. medicare	37. kosovo	47. detain
8. toxic	18. milosevic	28. skills	38. quebeckers	48. pembroke
9. sport	19. unpaid	29. humanitarian	39. foundation	49. committees
10. diversity	20. extradition	30. wing	40. NAFO	50. ceiling

Table 11: Top 50 Conservative features for *tf-idf*/words/(500; 10; 5).

1. band	11. bands	21. taxation	31. refugee	41. scotia
2. immigration	12. PC	22. monopoly	32. bureaucrats	42. jail
3. taxpayer	13. reserves	23. nova	33. DEVCO	43. billions
4. patronage	14. property	24. somalia	34. shut	44. bills
5. auditor	15. native	25. surrey	35. selection	45. gas
6. negatived	16. arbitration	26. payroll	36. housing	46. mismanage
7. bureaucracy	17. appointments	27. advertising	37. GST	47. custody
8. closure	18. supposed	28. PSAC	38. coquitlam	48. pornography
9. conditional	19. nuclear	29. natives	39. pockets	49. rail
10. progressive	20. mint	30. inquiry	40. coal	50. division

fewer verbs and more nouns for both ideologies. Specifically, the Liberals frequently refer to geographical locations (*miramichi*, *thornhill*, *labrador*, *guelph*, *cape breton*, *iraq*, *london*, *kosovo*, *pembroke*), as well as health (*CHST*, *medicare*) and agriculture (*agri*, *CWB*, *NASS*). In contrast, the Conservatives continue to focus on taxation (*taxpayer*, *auditor*, *taxation*, *GST*), money (*payroll*, *pockets*, *billions*), and aboriginal affairs (*band*, *bands*, *reserves*, *native*, *natives*), but also on energy (*nuclear*, *coal*, *gas*) and immigration (*immigration*, *refugee*). It should be noted that the presence of the word *negatived* (as in "the motion was negatived") in the top 10 Conservative features is the result of our failure to remove all the formulaic phrases from the Government Orders speeches.

5.3 Classifying Speech Segments in the Oral Question Period

In addition to classifying speakers, we performed two experiments on the OQP-Segments dataset, which consists of 20,000 short speech segments, made by 6,666 Conservatives and 13,334 Liberals. Given the format of the Oral Question Period, most of the Conservative segments are questions, while the Liberal segments are either questions from fellow Liberal MPs or responses made by Cabinet Ministers and their representatives. The following is a description of the experiments we conducted (with the *tf-idf*/words method):

- Experiment 1: we used 2,000 randomly selected segments as the test set (*proportionally* balanced — 660 Conservative segments and 1,340 Liberal segments), and the remaining 18,000 as the training set.
- Experiment 2: we used an *evenly* balanced training set of 5,916 Conservative and 5,916 Liberal segments, as well as an *evenly* balanced test set of 750 Conservative and 750 Liberal segments.

and Services Tax.

Table 12: OQP-Segments classification results for the two experiments.

	Experiment 1 (proportional test, proportional training)	**Experiment 2** (even test, even training)
majority baseline	67.00	50.00
(0; 0; 5)	93.85	91.93
(500; 10; 5)	84.20	81.53

Table 12 shows the results of these experiments. The highest accuracy is achieved in Experiment 1, which can be explained by observing the composition of the test set: there are fewer Conservative segments than Liberal ones, and since most errors come from the Conservative side, the total number of misclassified MPs is reduced. However, given the marked difference between the baselines of the two experiments, the best results are achieved in Experiment 2, since there is a greater *overall* reduction in the error rate.

Notice the relatively high accuracy that is obtained using the (0; 0; 5) word removal strategy — this is unusual, since the documents being classified are very short, typically no more than 150 words, which means that they are generally harder to classify than documents containing many speeches. Notice also the large drop in accuracy (around 10%) when 500 of the most frequent words are removed. This result seems to support our previous hypothesis that the format of the OQP has a significant impact on the classifier's performance — i.e., the accuracy increases when certain words that are associated with the question-and-answer format of the debates are kept.

5.4 Analyzing the Emotional Content of Speeches

Recall that our feature analysis has shown that Liberals tend to use words that convey a more positive sentiment than those used by the Conservatives. This suggests that it might be possible to distinguish Liberal MPs from Conservative MPs based only on

Table 13: LIWC classification results for all three experiments.

	64 features	**posemo and negemo**	**posemo $-$ negemo**
OQP-Speakers	60.5	80.5	81.0
GOV-Speakers	60.5	79.5	78.5

the emotional content of their speeches. In order to test this hypothesis, we proceeded as follows.

First, we performed 5-fold cross-validation experiments on the OQP-Speakers and GOV-Speakers datasets using 64 LIWC categories as features.[20] We did this to see whether positive and negative emotion are among the top discriminating features for Liberals and Conservatives, respectively. In fact, a feature analysis confirms that positive emotion is among the top 5 Liberal features, while negative emotion is among the top 10 Conservative features. Then, we repeated the same experiments, using only positive emotion and negative emotion (referred to as **posemo** and **negemo**) as features. Finally, we performed a third experiment, where affect was reduced to one feature — positive emotion minus negative emotion. Table 13 shows the results of these experiments.

In the first experiment, we found that the accuracy for both datasets is equal to the majority baseline, because all MPs are classified as Liberals! This result may be explained by the fact that no LIWC category has a *significant* impact on the classification. In other words, even though some categories are listed as the discriminating features for Liberals and the rest as the discriminating features for Conservatives, the difference between these two groups is so slight that the resulting classifier simply labels all test instances as belonging to the majority class.

In contrast, notice that using positive and negative emotion as one, or two, features

[20]These 64 categories and their values (e.g., the percentage of words in the text that are pronouns) were derived from LIWC ouput produced by running the software on all Liberal and Conservative speeches.

yields a substantial improvement of up to 20.5 percentage points over the baseline. However, it is not clear whether this is a result of ideological differences between Liberals and Conservatives, or the difference between members of the government and the opposition. Evidence of this confound has been observed before and it remains to be determined to what degree this property of the Canadian Parliamentary system affects our results.

6 Conclusions and Future Work

The initial motivation behind our study was to explore the task of determining someone's underlying belief system based on the words they use in public speech. Specifically, we focused on the problem of distinguishing Liberals from Conservatives using transcripts of Canadian parliamentary debates. In this section, we summarize the contributions of our research and discuss ideas for future work on this subject.

6.1 Summary of Contributions

We conducted a series of classification experiments on several datasets that consisted of speeches that were made in the House of Commons by Liberal and Conservative MPs. This involved training SVM classifiers on a number of documents — represented as vectors in a multi-dimensional space — and testing their performance in a cross-validation methodology. We achieved an accuracy of up to 98% on speeches from the Oral Question Period, and up to 89.5% on the combination of the OQP data and speeches made during the Government Orders portion of the debates. These results demonstrate that it may be possible to detect ideological differences between speakers on opposite ends of the political spectrum using a simple bag-of-words model.

However, we noticed that the format of the OQP may be affecting the performance of the classifier, which is why we conducted experiments using only the Government Orders speeches. Although there was a noticeable drop in accuracy (in comparison with the OQP results), we still achieved good results — up to 86.5% accuracy, which is well above the 60.5% majority baseline. Moreover, we believe that the errors made by our classifier might indicate that some of the misclassified MPs may not uphold the same political beliefs as other members of the party that they belong to.

Our findings also suggest that there are fundamental differences in the language used by Liberals and Conservatives. For instance, members of the two ideological classes tend to discuss different topics. A feature analysis has shown that the Liberal lexicon is characterized by words related to Québec, health, agriculture, and various social issues, while the Conservatives frequently talk about taxation, money, and aboriginal affairs. Moreover, the words used by Liberals tend to convey positive emotion, as opposed to the generally negative affect expressed by Conservative speech. In fact, we achieved up to 81% accuracy when classifying MPs using the percentages of positive and negative emotion words found in their speeches. However, this result raises the question of whether party status — i.e., Liberals as the governing party and Conservatives as the Official Opposition — plays a role in the classification. This issue has been a concern to us throughout the entire study, given the nature of Parliamentary systems, and we plan on exploring it further in the future.

6.2 Comparison of Work

We compare our research to other work on ideological classification — described in Section 2.3 — that was done on U.S. data, since no one has performed these experiments on Canadian data before. Diermeier et al. (2007) classify U.S. Senators as liberals or conservatives, based on their Congressional speeches. They achieve mediocre results (52% accuracy on "moderate" Senators) unless they make certain limiting assumptions. Specifically, they achieve up to 94% accuracy on a dataset that consists of only "extreme" Senators — those rated as the *most* Liberal and the *most* Conservative. This dataset also contains more documents than unique speakers, which means that many of the Senators in the test set are represented in the training set.

In our work, we make no *a priori* assumptions about the degree to which someone is a Liberal or a Conservative, and we test our classifier on individuals whose speeches have *not* been seen during the training phase. We still achieve up to 89.5% accuracy on a dataset that consists of speeches from both the Oral Question Period and the Government Orders portions of the Canadian House of Commons debates. These findings indicate that there is a sharper distinction between Liberals and Conservatives in the Canadian Parliament, at least based on the language they use. This result may be partially attributed to the fact that members of the government and the opposition might be pressured to adhere to party doctrine in order to create the appearance of strength through unity, since the roles of the parties can easily be reversed after any given federal election.

6.3 Future Directions

As was mentioned earlier, we would like to test the hypothesis that there exists a "Government vs. Opposition" confound that affects our classification results. In order to do this, we intend to acquire additional data from another time period, in which the roles of the Liberals and the Conservatives are reversed (i.e., Conservatives form the Government and the Liberals are in Opposition). The primary candidates for this are the 33$^{\text{rd}}$ and 34$^{\text{th}}$ Parliaments (Conservative majority governments, led by Brian Mulroney), as well as the more recent 39$^{\text{th}}$ and 40$^{\text{th}}$ Parliaments (Conservative minority governments, led by Stephen Harper). Our goal is to train a classifier on speeches made by Liberal and Conservative MPs in one setting (time period), and test it on speeches made by MPs in another setting, where the roles of the two parties are switched. If the resulting accuracy is still high, this would indicate that the distinction is being made on the basis of *ideology*. If the accuracy decreases, this would support the hypothesis that *party status* (Government vs. Opposition) has considerable influence on the performance of the classifier. However, the trouble with this approach is that many of the top Liberal and Conservatives features reflect issues and events that are associated with a certain time period — e.g., Jean Chrétien's Shawinigan business scandal in the 36^{th} Parliament. This means that classifying speeches from a different time period may inevitably yield a lower accuracy as a result of differences in the major topics of debate.

Recall that the Canadian Hansard corpus is bilingual — containing transcripts of debates in English and in French — but in this study we focused exclusively on the English portion of the data. Hence, another future research direction could involve conducting the same classification experiments on Liberal and Conservative speeches in French, and comparing these results to our findings in order to see whether they are consistent. If there are significant differences, this might suggest that the translation from one language to the other is changing the content of the speeches enough to affect the classifier's performance.

Also, there is no reason why the classification task should be limited to a binary distinction between Liberals and Conservatives — it might be useful to include speeches from members of the NDP and the Bloc Québécois in the data and use SVMs to perform a four-way classification. This experiment could yield further insights about the language that characterizes each ideology, which is approximated by party affiliation.

Appendix A

The data we use in our research was extracted from the *Canadian Hansard*, which is a printed record of parliamentary debates. This section contains additional details about our corpus, as well as tables with raw data. All information concerning official government proceedings was obtained from the Parliament of Canada website (`http://www.parl.gc.ca`).

House of Commons

The House of Commons is where MPs debate and vote on proposed legislation, as well as discuss issues that are of national importance. Although speeches in the Commons Chamber can be emotional — especially during exchanges between Government and Opposition members — the general tone of the proceedings is restrained. All discussions are moderated by the Speaker of the House, who enforces the rules of the debate and, where appropriate, a time limit on each participant to make their point.

In our work, we use speeches from the **Oral Question Period** and **Government Orders** portions of the House of Commons debates, because we believe they are the most relevant for our classification task. It is worth noting that there are other parts of the collection that we have not included in our dataset. For example, we considered using the **Statements by Members**, which are short speeches delivered by MPs from every party on a variety of subjects of national, regional, or local importance (e.g., the plight of Canadian farmers). However, since these speeches are unlikely to contain words that distinguish Liberals from Conservatives, we decided to omit such discourse from our corpus.

Table 14: A timeline of the 36^{th} Parliament.

	1^{st} Session		2^{nd} Session	
1997.09.22		1999.09.18 | 1999.10.12		2000.10.22

1^{st} Session: 243 House of Commons Sitting Days
2^{nd} Session: 133 House of Commons Sitting Days

The 36^{th} Parliament

During the 36^{th} Parliament, the Liberal Party of Canada formed a Majority Government, led by Prime Minister Jean Chrétien.[21] This period lasted from September 22, 1997, to October 22, 2000. It was divided into two sessions (see Table 14 for a more detailed timeline). Until March 26, 2000, the Official Opposition to the Liberals was the **Reform Party**, which was led by Preston Manning. However, it then became the **Conservative Reform Canadian Alliance**, which was initially led by Deborah Grey, followed by Stockwell Day. The other parties represented in the House of Commons were the **Progressive Conservatives (PC)**, the **New Democratic Party**, and the **Bloc Québécois**.

Working with the Data

In order to extract the data we needed, we first downloaded the Aligned *Hansards* of the 36^{th} Parliament of Canada[22]. The format of this corpus is very convenient: for each House of Commons Sitting Day there are two files — one in English and the other in French — containing transcripts of debates, such that each sentence in a file is on a separate line. However, there are only 350 of these file pairs, which means that this

[21] This was also the case for the preceding (35^{th}) Parliament, as well as for the following (37^{th}).
[22] `http://www.isi.edu/natural-language/download/hansard/`.

corpus does not cover the entire 36^{th} Parliament. Specifically, all days after May 10, 2000, and up until October 22, 2000, are missing.

We discarded the French files and focused on the English data, extracting speeches for all Liberals (MPs labeled with the abbreviation **Lib.**) and all Conservatives (MPs labeled with the abbreviations **Ref.**, **PC**, and **Canadian Alliance**). Whenever a new speaker enters a debate, their speeches are prepended with a heading such as the following:

Mr.	Maurizio Bevilacqua	(Vaughan-King-Aurora,	Lib.):
title	name	riding	party

These headings make it easy to locate individuals in the corpus. A *speech* is defined as an uninterrupted segment of text that is spoken by that individual.[23] However, it should be noted that of the 350 files in the corpus, we were able to use only 331 to gather the Oral Question Period data, because some files did not contain speeches from this part of the debates. Similarly, we were able to use only 328 files to gather the Government Orders data.

There exist official records that list all the members of the House of Commons at the time when Parliament was dissolved. It is worth mentioning some differences between these records, for the 36^{th} Parliament, and our coverage of Conservative MPs. First, both Stockwell Day and Charles Joseph Clark — leaders of the Canadian Alliance and PC, respectively — are not represented in our dataset, because they entered the House of Commons after May 10, 2000, which is beyond the scope of the Aligned *Hansards* corpus. Also, several MPs switched parties at some point during the 36^{th} Parliament (see Table 15 for more details about these transitions).

[23]We use various heuristics to determine where a speech ends.

Table 15: Conservative MPs who switched parties during the 36th Parliament.

Name	Party at Opening	Party at Dissolution	Comments
Jake E. Hoeppner	Ref.	Ind.	Became Independents late in the Parliament, so treated as *Conservative*.
Jack Ramsay	Ref.	Ind.	
André Harvey	PC	Lib.	Became Liberals late in the Parliament, so treated as *Conservative*.
David Price	PC	Lib.	
Diane St-Jacques	PC	Lib.	
Jean J. Charest	PC	Lib.*	Resigned as leader of the PC in the 1st Session, so treated as *Conservative*. *Later became a Liberal in provincial politics, in Quebec.
Scott Brison	PC	Lib.**	Resigned his seat as PC in July, 2000, so treated as *Conservative*. **Eventually became a Liberal in 2003.
Angela Vautour	NDP	PC	Became PC in the 2nd Session, but had no speeches as NDP in the 1st Session, so treated as *Conservative*.

Table 16: Conservative MPs in the OQP-Speakers and GOV-Speakers datasets.

Allan Kerpan	Angela Vautour	Art Hanger
Bill Gilmour	Bob Mills	Charlie Penson
Chuck Cadman	Chuck Strahl	Cliff Breitkreuz
Dale Johnston	Darrel Stinson	David Chatters
Deborah Grey	Deepak Obhrai	Derrek Konrad
Diane Ablonczy	Dick Harris	Eric Lowther
Garry Breitkreuz	Gary Lunn	Gerry Ritz
Grant Hill	Grant McNally	Gurmant Grewal
Howard Hilstrom	Inky Mark	Jack Ramsay
Jake E. Hoeppner	Jason Kenney	Jay Hill
Jim Abbott	Jim Gouk	Jim Hart
Jim Pankiw	John Cummins	John Duncan
John Reynolds	John Williams	Keith Martin
Ken Epp	Lee Morrison	Leon E. Benoit
Maurice Vellacott	Mike Scott	Monte Solberg
Myron Thompson	Paul Forseth	Peter Goldring
Philip Mayfield	Preston Manning	Rahim Jaffer
Randy White	Reed Elley	Richard M. Harris
Rick Casson	Rob Anders	Roy Bailey
Ted White	Val Meredith	Werner Schmidt
André Bachand	André Harvey	Bill Casey
Charlie Power	David Price	Diane St-Jacques
Elsie Wayne	Gerald Keddy	Gilles Bernier
Greg Thompson	Jean Dubé	Jean J. Charest
Jim Jones	John Herron	Mark Muise
Norman Doyle	Peter MacKay	Rick Borotsik
Scott Brison		

Table 17: Liberal MPs in the OQP-Speakers dataset.

Aileen Carroll	Alex Shepherd	Alfonso Gagliano
Allan Rock	Andrew Telegdi	Andy Mitchell
Andy Scott	Anne McLellan	Arthur C. Eggleton
Bernard Patry	Beth Phinney	Bill Graham
Bob Speller	Bob Wood	Bonnie Brown
Brent St. Denis	Bryon Wilfert	Carmen Provenzano
Carolyn Bennett	Carolyn Parrish	Charles Caccia
Charles Hubbard	Christine Stewart	Claude Drouin
Claudette Bradshaw	Colleen Beaumier	David Anderson
David Iftody	David Kilgour	David M. Collenette
David Pratt	Denis Coderre	Denis Paradis
Derek Lee	Diane Marleau	Don Boudria
Eleni Bakopanos	Elinor Caplan	Ethel Blondin-Andrew
Eugene Bellemare	Fred Mifflin	George S. Baker
Gerry Byrne	Gilbert Normand	Gurbax Singh Malhi
Guy St-Julien	Harbance Singh Dhaliwal	Hec Clouthier
Hedy Fry	Herb Gray	Ian Murray
Jacques Saada	Jane Stewart	Janko Peric
Jean Augustine	Jean Chrétien	Jim Peterson
Joe Jordan	Joe McGuire	John Cannis
John Finlay	John Harvard	John Maloney
John Manley	John McKay	John O'Reilly
John Richardson	Joseph Volpe	Judi Longfield
Julian Reed	Karen Kraft Sloan	Karen Redman
Larry McCormick	Lawrence D. O'Brien	Lawrence MacAulay
Lloyd Axworthy	Lucienne Robillard	Lyle Vanclief
Lynn Myers	Mac Harb	Marcel Massé
Maria Minna	Mark Assad	Marlene Jennings
Martin Cauchon	Mauril Bélanger	Murray Calder
Nancy Karetak-Lindell	Nick Discepola	Paddy Torsney
Pat O'Brien	Paul Martin	Paul Szabo
Pierre S. Pettigrew	Ralph E. Goodale	Raymond Chan
Raymonde Folco	Reg Alcock	Robert Bertrand
Robert D. Nault	Roger Gallaway	Ronald J. Duhamel
Rose-Marie Ur	Roy Cullen	Sarkis Assadourian
Sarmite Bulte	Sergio Marchi	Sheila Copps
Sheila Finestone	Sophia Leung	Stan Dromisky
Stan Keyes	Steve Mahoney	Stéphane Dion
Sue Barnes	Susan Whelan	Ted McWhinney
Tony Valeri	Walt Lastewka	Wayne Easter
Yvon Charbonneau		

Table 18: Top 50 most frequent words in the OQP-Speakers dataset.

1. the	11. it	21. with	31. there	41. an
2. to	12. this	22. has	32. by	42. our
3. of	13. for	23. be	33. s	43. very
4. that	14. minister	24. he	34. all	44. his
5. is	15. have	25. was	35. would	45. been
6. in	16. are	26. as	36. at	46. when
7. and	17. not	27. canada	37. do	47. about
8. a	18. will	28. they	38. hon	48. canadians
9. we	19. government	29. member	39. from	49. house
10. i	20. on	30. what	40. prime	50. can

References

Bansal, M., Cardie, C., & Lee, L. (2008). The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In *Proceedings of the international conference on computational linguistics (COLING).* (Poster paper)

Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., et al. (1990). A statistical approach to machine translation. *Computational Linguistics*, *16*(2), 79–85.

Diermeier, D., Godbout, J.-F., Yu, B., & Kaufmann, S. (2007). Language and ideology in congress. In *Annual Meeting of the Midwest Political Science Association.*

Fraser, A., & Marcu, D. (2007). Measuring word alignment quality for statistical machine translation. *Computational Linguistics*, *33*(3), 293–303.

Godbout, J.-F., & Hoyland, B. (2008). Legislative voting in the Canadian Parliament. In *Meeting of the Canadian Political Science Association.*

Greene, S. (2007). *Spin: Lexical Semantics, Transitivity, and the Identification of Implicit Sentiment*. Unpublished doctoral dissertation, University of Maryland.

Joachims, T. (1998). Text categorization with suport vector machines: Learning with many relevant features. In *Proceedings of the 10th european conference on machine learning (ECML)* (pp. 137–142). London, UK: Springer-Verlag.

Joachims, T. (2008). *SVM-light: Support Vector Machine (Version 6.02).* Available from `http://svmlight.joachims.org/`

Kwon, N., Shulman, S., & Hovy, E. (2006). Multidimensional text analysis for eRulemaking. In *Proceedings of the 2006 international conference on Digital government research (dg.o 2006)* (pp. 157–166).

Kwon, N., Zhou, L., Hovy, E., & Shulman, S. (2007). Identifying and classifying subjective claims. In *Proceedings of the 2007 international conference on Digital government research (dg.o 2007)* (pp. 76–81).

Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, *97*(2), 311–331.

Lin, W.-H., & Hauptmann, A. (2006). Are these documents written from different perspectives? A test of different perspectives based on statistical distribution divergence. In *Proceedings of the international conference on computational linguistics (COLING)* (pp. 1057–1064). Association for Computational Linguistics.

Lin, W.-H., Wilson, T., Wiebe, J., & Hauptmann, . (2006). Which side are you on? Identifying perspectives at the document and sentence levels. In *Proceedings of the conference on natural language learning (CoNLL).*

Mullen, T., & Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)* (pp. 412–418). (Poster paper)

Mullen, T., & Malouf, R. (2006). A preliminary investigation into sentiment analysis of informal political discourse. In *AAAI symposium on computational approaches to analysing weblogs (AAAI-CAAW)* (pp. 159–162).

Mullen, T., & Malouf, R. (2007). Graph-based user classification for informal online political discourse. In *Proceedings of the 1st workshop on information credibility on the web (WICOW).*

Mullen, T., & Malouf, R. (2008). Taking sides: User classification for informal online political discourse. *Internet Research*, *18*, 177–190.

Nobuhiro, N. K., & Kitsuregawa, M. (2007). Building lexicon for sentiment analysis from massive collection of HTML documents. In *Proceedings of the joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)* (pp. 1075–1083).

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1–135.

Pennebaker, J. W. (2008). *The Meaning of Words: Obama versus McCain* [web page]. World Wide Web electronic publication. Available from `http://wordwatchers.wordpress.com/2008/10/12/the-meaning-of-words-obama-versus-mccain`

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2007). *Linguistic Inquiry and Word Count (LIWC2007).*

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, *14*(3), 130–137.

Skillicorn, D., & Little, A. (2006). *Spinning the Election* [web page]. World Wide Web electronic publication. Available from `http://research.cs.queensu.ca/home/skill/election/election.html`

Thomas, M., Pang, B., & Lee, L. (2006). Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)* (pp. 327–335).

Yang, C. H., Lin, K. H. Y., & Chen, H. H. (2007). Building emotion lexicon from weblog corpora. In *Proceedings of 45th annual meeting of association for computational linguistics (ACL).* (Poster paper)

Yu, B., Diermeier, D., & Kaufmann, S. (2008). Classifying party affiliation from political speech. *Journal of Information Technology in Politics*, *5*(1), 33–48.

Yu, B., Kaufmann, S., & Diermeier, D. (2008). Exploring the characteristics of opinion expressions for political opinion classification. In *Proceedings of the 2008 international conference on Digital government research (dg.o 2008)* (pp. 82–91). Digital Government Society of North America.

Yu, H., & Hatzivassiloglou, V. (2003). Towards answering opinion questions: Separating facts from opinions and identifying the polarity of opinion sentences. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP).*