Analyzing Political Polarization in Canada's Parliament

Connor Glossop, 1005228576

Supervisor: Professor Rohan Alexander

Table of Contents

List of Figures and Tables

Introduction

Political polarization has been shown to reduce the effectiveness of governments and increase the occurrences of bad policies [1]. Research has therefore been conducted to try to measure political polarization in various countries, often leveraging the transcripts from government proceedings and outcomes of votes. The Canadian Hansard dataset [2] contains the transcripts of all debates in the Canadian Parliament (both the House of Commons and the Senate). The full transcripts of each Member of Parliament and Senators' speeches, questions and responses are recorded every day that the parliament is in session, and have been collected from January 1994 to present. The Hansard Index, a list of each individual MP/Senator's speeches for a given session, was also maintained between January 1994 and November 2005. The results of parliamentary votes are recorded both within the Hansard document for a given day and a separate Votes database [3]. This database contains records of all votes from October 2004 onwards.

In 2021, Alsinet et al. [4] developed metrics for evaluating the political polarization of Reddit users as a substitute for the general population. Hanretty et al. [5] used statistical regression techniques to analyze the left-right split of UK Members of Parliament based on their voting results, while Goet [6] analyzed the UK's Hansard dataset to estimate dyadic representation in various historical periods. Research has been done into analyzing dyadic representation in Canada [7], however little analysis has been performed on the political polarization present within Canada's parliament and its change over time.

The goal of this thesis project is to analyze the transcript and voting data contained in the Hansard datasets to try to answer the following questions: Are Canadian politicians more polarized in the periods leading up to elections compared to immediately after elections? Furthermore, has Canada become more politically polarized over the period covered by the Hansard dataset? This report details the progress made thus far on the thesis project, as well as future steps that will be taken to complete the project.

The information collected when analyzing the first research question can then be applied to the second question. These results can be combined to obtain a “polarization score” for each session, which can then be compared across parliamentary sessions to determine if Canada has become more politically polarized over the past decades.

Canadian Political Landscape

According to The Canadian Encyclopedia, there are currently five major political parties in Canada. For most of recent Canadian history, its politics had been dominated by the Liberal and the Conservative parties [23]. The Liberal Party (the Liberals) is a centre-left party historically connected to the ideas of liberalism, neoliberalism, globalism and biculturalism. The Conservative Party (the Conservatives) is a centre-right party historically connected to conservative and neo-conservative ideologies, advocating tax breaks and small government [24]. After being founded in 1961, the New Democratic Party (the NDP) became a progressively more important political party as time went on. A staunchly left-wing party, the NDP rose to national prominence in the 2011 election where they placed second. Since then, they have become an influential party despite lacking the support needed to win an election. In the event of a minority government, the NDP often holds the balance of power. The two other major parties are the Bloc Québécois, a regionalist party advocating for Quebec-related issues, and the Green Party, a left-wing environmentalist party [23].

Despite more parties gaining in political influence in recent years, Canada currently operates under a “two party-plus” system according to journalist J.J. McCullough, with the Liberals and Conservatives still acting as dominant forces [25].

Literature Review

Political Polarization

According to Cook et al., polarization is the irrational occurrence of when "two people respond to the same evidence by updating their beliefs in opposite directions" [14]. Similarly, DiMaggio et al. more directly defines political polarization as the "measure" of how opposed opinions on a given issue are, and how that opposition increases over time [12]. Specifically, they state that polarization refers to the "extremity of and distance between responses" to an issue rather than the content of those responses.

Political polarization is commonly cited as either the root cause or a key compounding factor in many of modern society's problems [10], [13]. Despite this, the exact causes and effects of political polarization are often debated. In *Linking Conflict to Inequality and Polarization*, Esteban and Debraj analyze the causes of societal conflict, specifically income inequality and polarization, with the goal of creating a behavioural model for conflict [16]. Their final model was constructed using a linear combination of three metrics (income inequality, social fragmentation and polarization), and was able to correctly predict the prevalence and scale of societal conflict (defined as a broad range of conflicts not limited to civil wars). As the model weights were all positive, this shows that an increase in polarization is correlated to an increased level of societal conflict.

In "*Is polarization bad?*" Testa analyzes the effects of elections in polarized societies, and how they can affect the quality of the resulting government [1]. She states "more [polarized] societies… tend to be ruled by bad governments that choose poor policies." Despite this, some amounts of political polarization can be seen to increase accountability of elected officials, especially on policies unrelated to the government's ideology. Due to the ubiquity of political polarization within modern democracies (despite it being seen as a fatal flaw of those democracies due to it undermining the political power of the majority of the population), understanding its effects are vital.

Measures of Polarization

As polarization is a sociological concept rather than a mathematical one, many attempts have been made to formulate accurate measures for it. *Have American's Social Attitudes Become More Polarized?*, written by DiMaggio et al., proposes four measures which can be used to evaluate the polarization of responses to a given issue [12].

Dispersion of opinion measures the "far apart-ness" of opinions, and can be directly measured using the variance of opinions. Greater variance indicates a more polarized population. In the equation below s^2, N and $\bar{x}$ are the variance, total number and mean value of opinions, with x_i being the value of an opinion.

$$s^2 = \sum_{i=1}^{N} \frac{(x_i - \bar{x})^2}{N-1}$$

Bimodality measures the extent to which opinions are divided into two separate "camps," with most opinions normally distributed around one of the two centroids and very few occupying the space between them. While it is theoretically possible for more than two centroids to exist, they are rarely seen in practice. DiMaggio et al. propose kurtosis to measure bimodality, a metric often used to differentiate between "peaked," "flat," and bimodal distributions [15]. Kurtosis can be defined as

$$k = \frac{\Sigma(x - \bar{x})^4}{Ns^4} - 3$$

where s, x, $\bar{x}$ and N are as described above. The factor of -3 ensures that an unaltered normal distribution has a kurtosis value of 0. Large values (approaching 2) indicate a highly non-polarized population, while lower values (approaching -2) indicate a highly bimodal, polarized population.

Constraint refers to how correlated a given person's views on one issue are to another issue. DiMaggio et al. note that a society that is more polarized is more likely to follow different ideologies, where each side has a "narrative" that connects various issues. They propose Cronbach's alpha, a metric which measures association between items. As α increases (corresponding to a strengthening in constraint as defined above), the polarization in the society increases. Cronbach's alpha is defined as

$$\alpha = \frac{k}{k-1}[1 - \frac{\Sigma\sigma_i^2}{\sigma_{yi}}]$$

where k is the number of items in the scale, σ_i^2 is the diagonal covariance for the ith item, and σ_{yi} is the sum of the diagonal and off covariances for all items.

Finally, Consolidation is the amount of "in-population" bimodality. This measures the amount of polarization within a given group, with the idea that less-polarized groups likely lead to increased overall polarization than highly polarized groups. This can also be measured with kurtosis.

In their paper *Measuring Polarization in Online Debates*, Alsinet et al. explore various methods of measuring polarization via debates on online social networks [4]. While some social networks, such as Facebook, have actively implemented policies and algorithms to attempt to mitigate polarization on their platforms, these have largely been unsuccessful. Alsinet et al. attempt to create a model to measure political polarization in debates on Reddit, a social media website focused on social news, user discussions and the sharing of web-based content. Alsinet et al. use a graph-based method to identify connections between user comments as they exist on the website. Through constructing a graph, they partition the graph into two sets: those who agree with the root comment (R) and those who don't (L). Using this partition method, they were able to create a measure of group partition by finding the number of interactions where members of one group are in agreement with another member of that group and contrasting that to the number of interactions where members of the two groups interact and are in disagreement. The partition is generated by randomly assigning users to one of the two groups, and iterating until the polarization measured is maximized. This method was able to accurately estimate the polarization of Reddit threads, with threads about Halloween tending to have significantly lower polarization measures than those about World Politics.

In *Linking Conflict to Inequality and Polarization* [16], metrics were used to quantify the levels of income inequality, social fragmentation and polarization in a society. The Gini Index is a measure of distance between two groups, and is commonly used to measure income inequality. The Hirschman-Herfindahl Fractionalization Index estimates the probability that two individuals belong to different societal groups, making it an effective measure of societal fragmentation. It has been demonstrated to relate ethnolinguistic diversity to conflict and growth; however, no clear link between conflict and either income inequality or fragmentation has been shown. The third metric used is Polarization, expressed with the following formula:

$$\tilde{P} = \Sigma_{i=1}^{m} n_i^2 (1 - n_i)$$

where m is the total number of groups and n_i is the share of the total that is a member of that group. Groups are assumed to be uniformly "distant" from one another.

In their 2017 paper *Cross-national measurement of polarization in political discourse: Analyzing floor debate in the U.S. the Japanese legislatures*, Sakamoto and Takikawa use topic modelling to analyze the relative levels of polarization during debates in the Japanese Diet compared to the United States Congress [26]. Records from January 1994 to December 2016 were taken from both the US Congress and Japanese Diet. After constructing a bag-of-words model and using multinomial regression to identify topics, a stochastic estimate for each group was calculated by aggregating the probability distributions of each topic over a specified group (e.g. party, gender) and over a specified timeframe. Sakamoto and Takikawa propose using Jensen-Shannon divergence between the distributions' estimates of the aforementioned group distributions to measure the level of political polarization between the two groups. The equation for Jensen-Shannon divergence can be seen below, where p and q (Θ_1 and Θ_2 in the original paper) refer to the distributions for two groups:

$$D_{JS}(p||q) = \frac{1}{2} D_{KL}(p||\frac{p+q}{2}) + \frac{1}{2} D_{KL}(q||\frac{p+q}{2})$$

These measures were then used to compare the major political parties in both the Japanese Diet and the United States Congress over the specified period. Due to limitations in the methods, each party could only be compared to a single party at a time. Sakamoto and Takikawa found that Japan was significantly more polarized during the period than the United States.

Polarization in the USA

DiMaggio et al. applied the metrics they developed to the General Social Survey (GSS) and National Election Study (NES), two large scale surveys of American political opinion, between 1972 and 1994 [12]. They found that on the overall population, Americans had become less politically polarized over the selected timeframe. Variance decreased on GSS responses despite staying approximately the same on NES responses, while kurtosis increased towards a value of 0. A decrease in polarization was also found in most issues when comparing both inter-group and intra-group polarization.

In the modern day, social media is often seen as a catalyst for increasing the political polarization of groups through the creation of "echo chambers" where members of those groups are pushed further to the extremes via the lack of contact with contradictory viewpoints. C. A. Bail et al. investigated this concept by exposing a large group of politically diverse American social media users to a Twitter bot representing views opposite to their own (i.e. a Republican was exposed to a Democrat Twitter bot while a Democrat was exposed to a Republican Twitter bot) [17]. This Twitter bot used tweets from various sources, including elected officials, opinion leaders, media organizations, and nonprofit groups. Study participants had their political beliefs measured via a short survey about various policies, and were ranked on a seven-point scale based on their responses.

Despite many studies reporting that regularly exposing people to beliefs opposing their own reduces political polarization, C. A. Bail et al. observed that users became significantly more polarized following the experiment. Liberal study participants outside of the control group became slightly more liberal after exposure to the Twitter bot. In contrast, C. A. Bail et al. found that "Republicans exhibited substantially more conservative views posttreatment that increase in size with level of compliance, and these effects are highly significant." On average, conservatives saw an increase of 0.6 points on the seven-point scale in the direction of conservatism, compared to just 0.14 points for liberals in the direction of liberalism (although this was not statistically significant).

Dyadic Representation

Dyadic Representation refers to “how well the sitting legislator acts as an agent for the constituency on legislative decisions” [11]. While dyadic representation is not identical to political polarization, it is closely linked and many of the methods for measuring dyadic representation can be extended to political polarization [28].

In *Dyadic Representation in a Westminster System*, Hanretty et al. used statistical techniques to analyze the dyadic representation of the United Kingdom specifically by looking at the voting record of Members of Parliament and comparing that to the estimated political views of the ridings they represent [5]. Using “multilevel regression and post-stratification,” a method for estimating public opinion on issues with small amounts of data, in conjunction with national survey data, the public opinion of 632 constituencies is approximated. These results are then compared to the voting record of the MPs to establish a measure of dyadic representation. Following this, the intra-party voting compliance of MPs is analyzed. While this is intended for use in dyadic representation analysis, it is closely connected to the ideas put forward by DiMaggio et al. for analysis of political polarization [12].

Natural Language Processing

Sentiment Analysis

Sentiment analysis is commonly performed on parliamentary debates to perform various tasks, one of the most common of which is measuring polarization [27]. According to Abercrombie and Batista‑Navarro, there are five primary methods for analyzing the sentiment of parliamentary debates. These include statistical machine learning and rule-based systems, both of which will be used in this thesis. One of the most popular statistical machine learning methods for sentiment analysis is BERT [29], a large-language model designed for “language inference without substantial task-specific architecture modifications” [8].

The rise of social media has led to increasing challenges for textual sentiment analysis. With often limited amounts of text available, more efficient methods of sentiment analysis have been developed. VADER (Valence Aware Dictionary for sEntiment Reasoning) is a rule based method for sentiment analysis, focusing specifically on the analysis of text from social media websites such as Twitter [18]. VADER shows large advances over previous lexical baselines when performing sentiment analysis on text retrieved from social media websites, with an overall F1 score of 0.96 on a corpus of 4200 tweets (a value higher than humans). While VADER is primarily used for analyzing social media text, it also outperforms seven other lexical baselines on Amazon.com product reviews and NY Times Editorials, and performs only slightly worse than best on a corpus of Movie Reviews. In these cases however, it performs significantly worse than on social media data (F1 scores of 0.63, 0.55 and 0.61 respectively) and does not outperform humans.

Progress to Date

All data for this thesis was retrieved from the Canadian Hansard using the Lipad project [19], a record of speeches given on each day that parliament is in session, divided by parliament. All initial analysis was performed on the 42nd Canadian parliament, which lasted from December 2015 until September 2019. To begin, various statistics and measures were calculated on speeches from May 2019, the final month of the 42nd parliament where parliament was held uninterrupted. During the 42nd parliament, the Liberals held a majority of seats, with the Conservatives acting as the official opposition. Following this, these measures were extended to the entirety of the 42nd parliament. Analysis was performed using the Pandas library [20], with the Natural Language Toolkit [21] (NLTK) used for natural language processing tasks.

May 2019 Analysis

First, the number of speeches each party delivered was analyzed, along with the most prevalent topics discussed on each day.

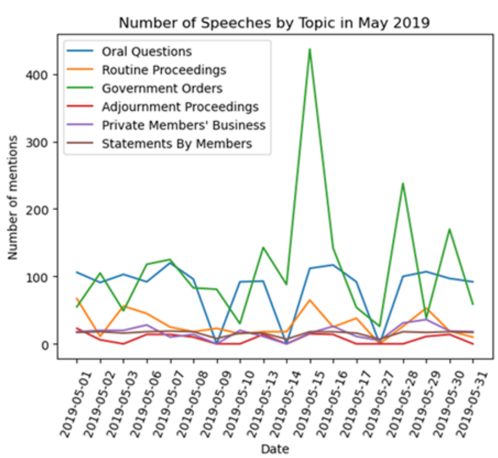


Figure 1: Number of Speeches by Topic in May 2019:
On most days, there were approximately 100 speeches concerning Government Orders and Oral Questions, with most other topics hovering at approximately 10 speeches per day. On May 15th (and to a lesser extent May 28th), there was a large increase in the number of speeches concerning Government Orders.

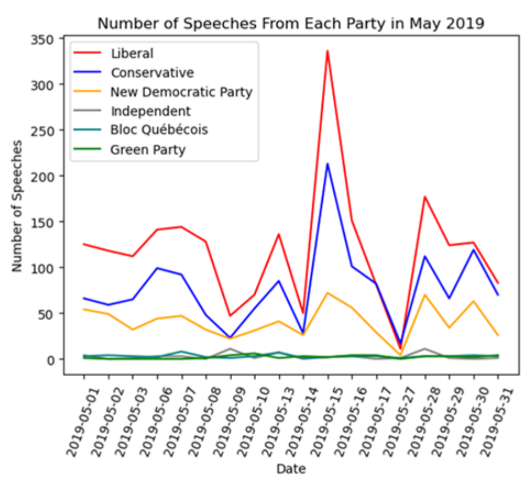


Figure 2: Number of Speeches From Each Party in May 2019:
The number of speeches from each party coincides generally with the number of MPs associated with those parties, with Liberal MPs making the most speeches followed by Conservative and NPD MPs.

Following this, further analysis was performed to identify the most prominent speakers by party, as well as the most commonly used words by each party once common stop words (e.g. “he”, “will”, “again”) had been removed. To remove repetitions of the same word in different forms (e.g. “a *community*”, “the *community’s*”, “many *communities*”), all words were stemmed using the Snowball stemmer, an inbuilt method in NLTK.

Table 1: Top 5 Most Prolific Speakers By Party in May 2019:

> It can be seen above that most speeches are performed by the "top 3" parties: the Liberals, Conservatives and the NDP. As Geoff Regan was Speaker of the House at this time, he delivered by far the most speeches. As the Liberals were the largest party in the House at this time, they made the most speeches.

	Liberal		Conservative		New Democratic Party		Independent		Bloc Québécois		Green Party	
	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches
0	Geoff Regan	260	Bruce Stanton	122	Carol Hughes	128	Celina Caesar-Chavannes	13	Monique Pauzé	13	Elizabeth May	36.0
1	Harjit S. Sajjan	171	Cheryl Gallant	55	Peter Julian	48	Hunter Tootoo	11	Michel Boudrias	8	Paul Manly	2.0
2	Kevin Lamoureux	168	James Bezan	48	Jagmeet Singh	34	Erin Weir	11	Luc Thériault	7		
3	Anthony Rota	130	Leona Alleslev	47	Rachel Blaney	28	Jane Philpott	5	Marilène Gill	7		
4	Justin Trudeau	100	Kelly McCauley	47	Pierre-Luc Dusseault	27	Darshan Singh Kang	4	Gabriel Ste-Marie	5		

Table 2: Top 10 Most Commonly Used Words by Party::

> It can be seen above that the majority of commonly used words (with stopwords removed) are standard words that do not provide much insight into the policies of the parties. Further analysis is needed where these words are removed.

	Liberal		Conservative		New Democratic Party		Independent		Bloc Québécois		Green Party	
	Word	Num Appearances	Word	Num Appearances	Word	Num Appearances	Word	Num Appearances	Word	Num Appearances	Word	Num Appearances
0	govern	1819	govern	2302	govern	1324	govern	93	quebec	142	govern	30
1	work	1355	minist	1736	peopl	775	legisl	71	languag	97	time	25
2	peopl	1063	liber	1506	liber	770	peopl	69	french	88	place	24
3	hous	980	peopl	1094	need	449	languag	67	govern	73	say	23
4	year	889	prime	1062	time	422	want	57	english	34	peopl	23
5	conserv	885	tax	989	year	417	indigen	56	want	33	hon	22
6	communiti	823	go	901	want	414	communiti	48	peopl	32	use	21
7	import	817	one	827	hous	414	question	46	offici	32	minist	19
8	make	804	time	803	one	409	one	41	year	30	countri	19
9	one	800	hous	720	work	378	hous	41	make	26	veri	18

Finally, various NLP techniques were applied to the data. Term Frequency-Inverse Document Frequency (TF.IDF) is a method that uses the number of occurrences of words across various preselected classes to determine how strongly correlated those words are to a given class. TF.IDF is widely used in NLP for sentiment classification and finding important class-specific words [22].

Table 3: Top 10 Most Party-Specific Words Found Via TF.IDF in May 2019:

	Bloc Québécois	Conservative	Green Party	Independent	Liberal	New Democratic Party
0	pointe	friendship	beauséjour	via	prosecute	ratify
1	assimilation	fishermen	68	inuktitut	deployed	fry
2	anglophone	disabled	gulf	guaidó	cyber	pharmacare
3	anglophones	virtue	greece	prairies	nato	cell
4	irvings	unifor	whale	itk	phasing	opioid
5	newcomers	litre	maori	abatement	doug	laundering
6	francophones	cta	seconded	stouffville	missions	automatic
7	bilingualism	huawei	nanaimo	whitby	pcrc	expunge
8	cheque	firearms	samos	kmaq	cse	havens
9	421	happiness	refinery	mi	ppsc	medication

In the above table, it can be seen that TF.IDF extracts words that are closely connected to what would be imagined given each party's mandate. For example, the top words for the NDP include "pharmacare," "medication," and "opioid," all words closely connected to their free healthcare platform.

Compound Vader Scores (described above) were calculated for each speech, with a separate box plot constructed for each political party. The exact values of the compound Vader score of each speech was then overlaid onto the box plot in the colour corresponding to the party's primary colour.

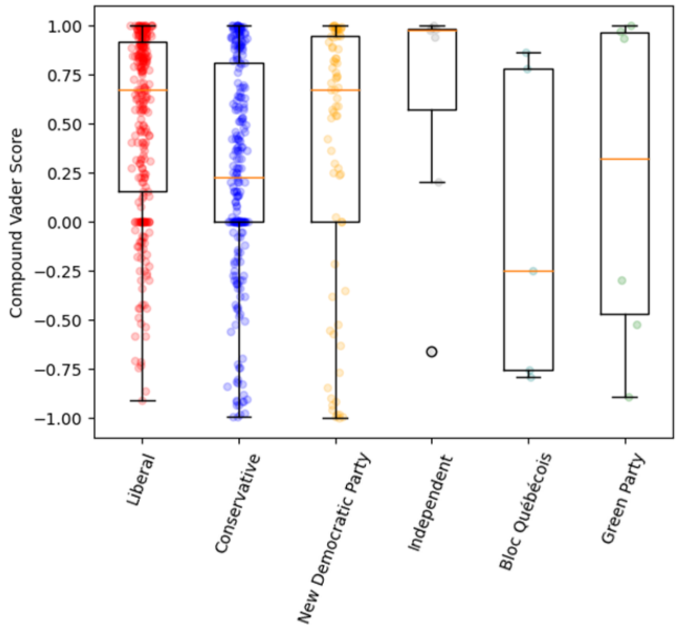


Figure 3: Average Compound Vader Scores of Speeches by Party in May 2019:

> The Liberal Party and New Democratic Party members are seen to make generally more positive speeches than either the Conservatives or Bloc Quebecios. As this was a Liberal majority government, this makes sense - the Liberals are in power and the NDP hold similar views to them. The Conservative Party, being the official opposition party, held far more negative speeches. The Bloc Quebecois, a Quebec nationalist/sovereignist party, held significantly more negative speeches than other parties.

The full breakdown of each subcomponent of the Compound Vader Scores can be seen below in Figure 4. It is important to note that very short speeches (such as those containing a single word e.g. “yes”) often return a Compound Vader Score of exactly 0, with the Positive and Negative Vader Scores equal to 0 and the Neutral Vader Score equal to 1.

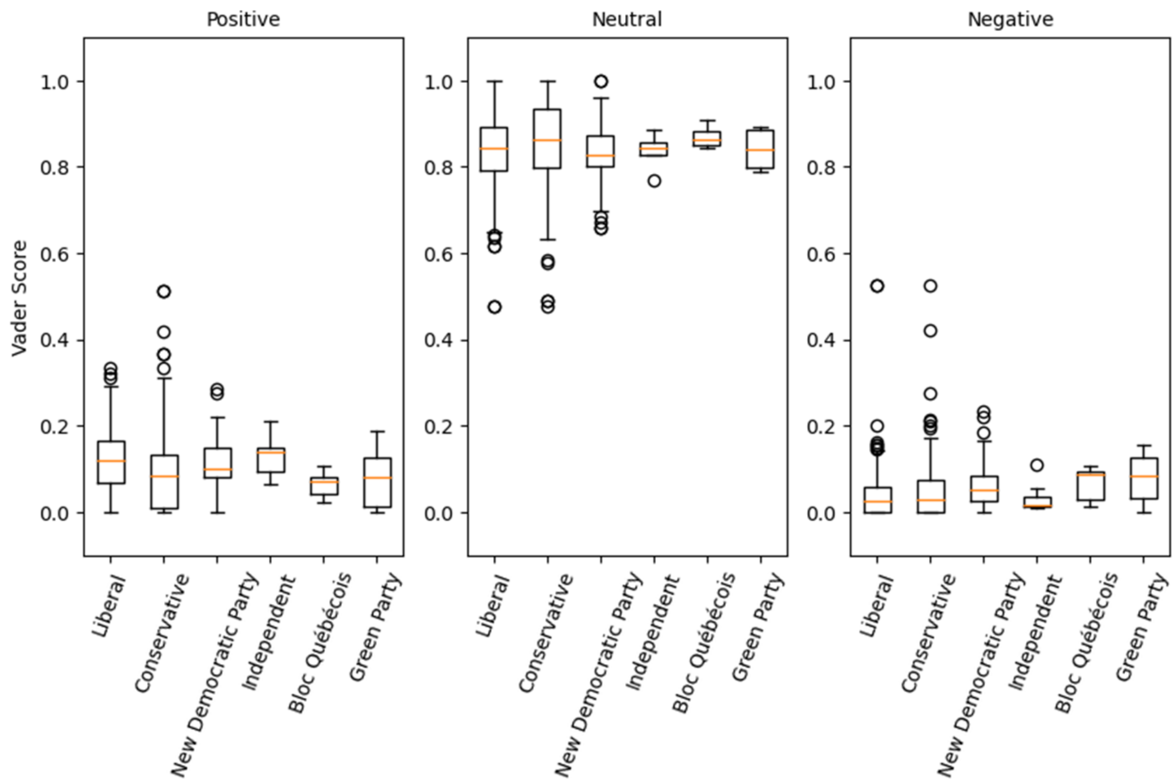


Figure 4: Vader Scores of Speeches by Party in May 2019:

> Here we can see the boxplots of vader scores for each party. This shows a better representation of the overall emotion of each party's speeches, and how individual speeches were inside or outside of the norm for those parties.

Figure 4 shows interesting insight into the relationships between the parties. As in Figure 3, the NDP had generally more positive speeches than the Conservatives, the official opposition. Interestingly however, the NDP also had on average a higher negative Vader score than conservatives, albeit with fewer extremely negative outliers. It is important to note that as this only covers a single month at the very end of the parliamentary cycle, caution should be taken when making statements about the 42nd parliament in general.

Full 42nd Parliament Analysis

Much of the analysis performed above was then redone on the 42nd parliament as a whole. First, the total number of speeches delivered by party, both over the entire parliament and over time, was calculated.

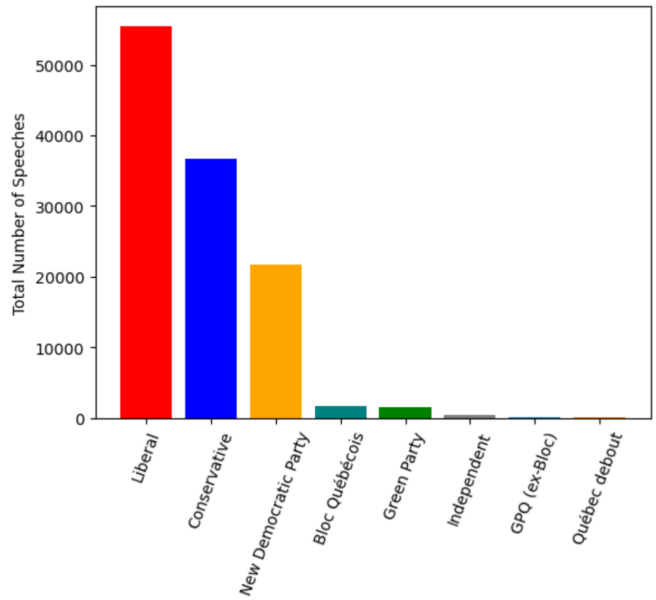


Figure 5: Total Number of Speeches by Party:

> The Liberal Party made the most speeches, with the Conservatives and New Democratic Party each making a substantial number of speeches throughout the 42nd Parliament. In total, 129,743 speeches were delivered.

Certain months experienced sharp drops in the number of speeches given, as seen in Figure 6. This can be explained by parliament going into or coming out of a recess. In these months the parliament is in session for only a few days (such as in January 2017 where parliament only met for two days). The share of speeches delivered by each party each month can be seen below in Figure 7, while Figure 8 shows the number of unique people from each party who gave a speech in a given month.

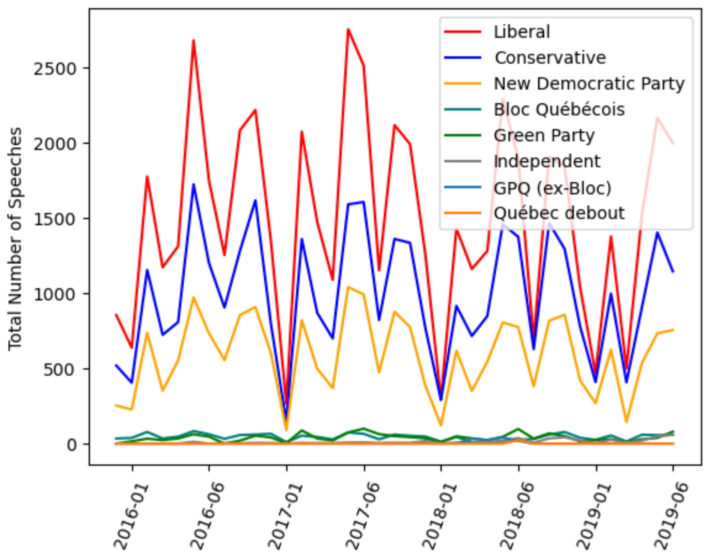


Figure 6: Number of Speeches Given by Each Party Over Time:

> In every month the Liberal party made the most overall speeches, followed closely by the Conservative party. May 2017 had the most speeches at 5529, closely followed by May 2016 with 5522 speeches.

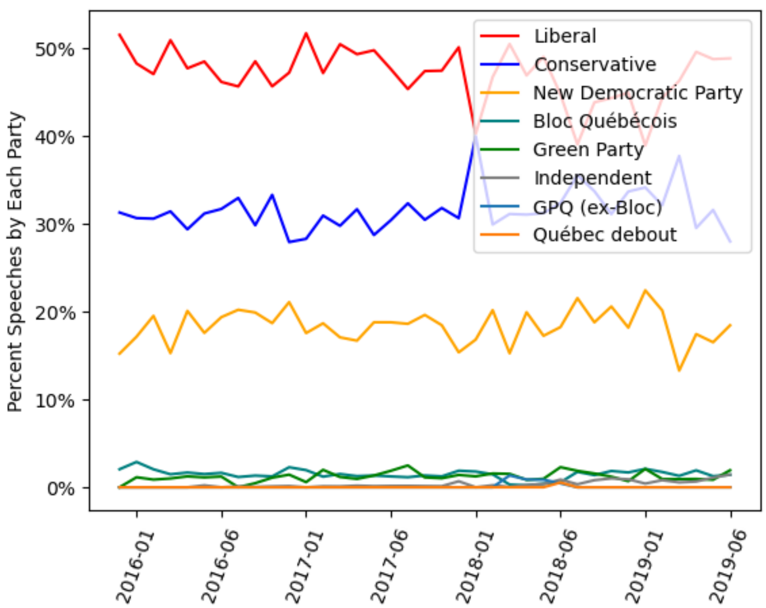


Figure 7: Relative Number of Speeches Given by Each Party Over Time:

> The Liberal Party made approximatly 50% of speeches in most months, with the Conservatives and New Democratic Party making approximatly 30% and 20% of speeches in most months respectivly. The jump in Conservative-given speeches in January 2018 corresponds to a month with an unusually low number of speeches given - Parliament was in recess from December 13, 2017 until January 29, 2018.

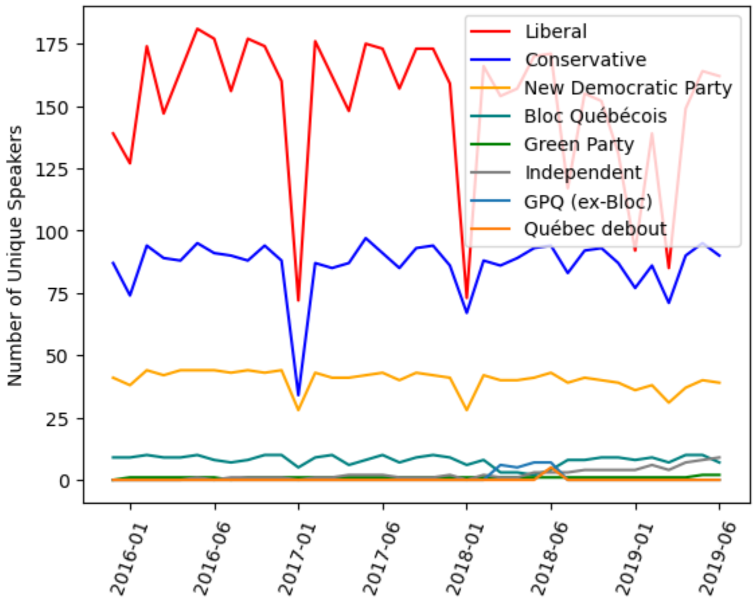


Figure 8: Number of Unqiue Speakers Given by Each Party Over Time:

> In every month the Liberal party consistently had the largest number of unique speakers, followed by the Conservatives and the New Democratic Party.

To conclude the general analysis, some analysis was performed on the most prolific speakers, as well as the main topics discussed on days where there were a large number of speeches. In Table 4 the five most prolific speakers by party are shown along with the cumulative number of speeches given by that person during the 42nd parliament. Table 5 shows the ten days with the largest number of speeches alongside the main topic discussed, as categorized by the LIPAD dataset.

Table 4: Top 5 Most Prolific Speakers By Party:

> As seen above, the majority of speeches were delivered by the Liberal Party. As Geoff Regan (Liberal) was Speaker of the House at this time, he delivered by far the most speeches. Other notable speakers include Carol Hughes (NDP) who was Assistant Deputy Speaker, and Bruce Stanton (Conservative) who was Deputy Speaker. Future Conservative Party leader Pierre Poilievre also ranked highly, as did Justin Trudeau and Andrew Scheer (leader of the Conservative party during the 42nd Parliament).

	Liberal		Conservative		New Democratic Party		Bloc Québécois		Independent		Green Party	
	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches	Speaker	Num Speeches
0	Geoff Regan	7701	Bruce Stanton	2953	Carol Hughes	3201	Gabriel Ste-Marie	317	Hunter Tootoo	154	Elizabeth May	1431.0
1	Kevin Lamoureux	3942	Garnett Genuis	1460	Pierre-Luc Dusseault	653	Monique Pauzé	224	Erin Weir	122	Paul Manly	39.0
2	Justin Trudeau	3060	Pierre Poilievre	1165	Sheila Malcolmson	648	Luc Thériault	223	Darshan Singh Kang	53		
3	Anthony Rota	2982	Gérard Deltell	1116	Jenny Kwan	646	Xavier Barsalou-Duval	188	Maxime Bernier	30		
4	Bardish Chagger	1606	Andrew Scheer	852	Alexandre Boulerice	630	Rhéal Fortin	158	Celina Caesar-Chavannes	21		

Table 5: Top Ten Days by Number of Speeches Delivered and Main Topic Discussed

	Date	Num Speeches	Main Topic Discussed
0	2016-06-14	701	Opposition Motion—Internal trade
1	2019-05-15	664	Department of National Defence—Main Estimates, 2019-20
2	2018-05-24	635	Department of Citizenship and Immigration—Main Estimates, 2018-19
3	2018-05-22	624	Department of Finance—Main Estimates, 2018-19
4	2019-06-18	622	Opposition Motion—The Environment
5	2016-09-19	621	Procedure and House Affairs
6	2017-05-29	600	National Defence—Main Estimates, 2017-18
7	2018-06-11	576	Opposition Motion—Iran
8	2016-05-30	564	Finance — Main Estimates 2016-17
9	2018-01-29	556	Finance

Of interesting note in Table 5 is the abundance of budgets included: the days with the 2nd, 3rd, 4th, 7th, 9th and 10th most speeches all primarily discussed a budget.

Following the general statistics, NLP analysis was run on the full 42nd parliament dataset. Vader Score analysis was performed on the speeches from each party as shown below in Figure 9.

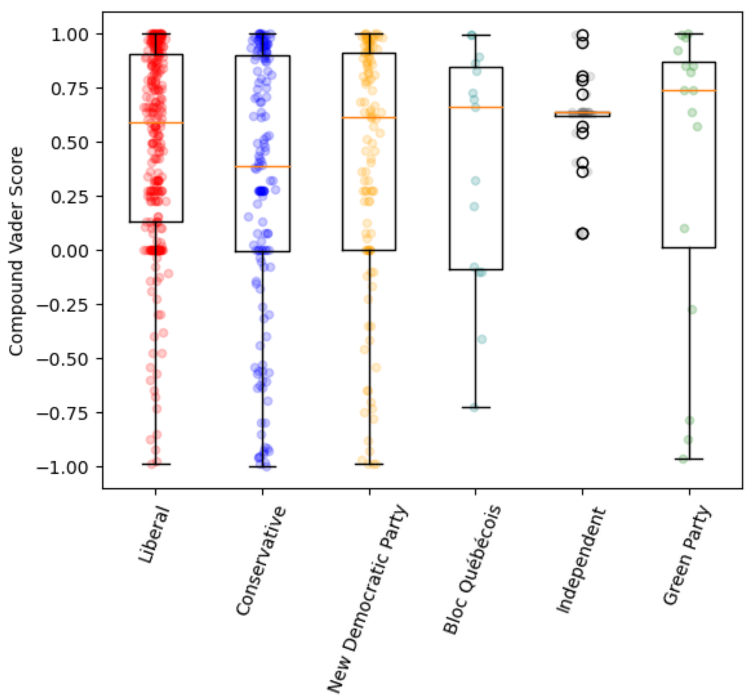


Figure 9: Average Compound Vader Scores of Speeches by Party:

> The Liberal Party and New Democratic Party members are seen to make generally more positive speeches than the Conservatives. As this was a Liberal majority government, this makes sense - the Liberals are in power and the NDP hold similar views to them. The Conservative Party, being the official opposition party, held far more negative speeches.

It is interesting to note the extreme number of outliers present for speeches given by Independents, especially given the extremely small interquartile range. As before, the exact Compound Vader Scores of speeches are overlaid onto the boxplots. Finally, TF.IDF analysis was run on the entire 42nd parliament, showing the words most associated with each party.

Table 6: Top 10 Most Information-Giving Words by Party:

> This table shows the ten most information giving words by party as found using tf.idf analysis

	Bloc Québécois	Conservative	GPQ (ex-Bloc)	Green Party	Independent	Liberal	New Democratic Party	Québec debout
0	kpmg	recession	recipient	dilbit	regina	scarborough	ladysmith	excuses
1	shipyard	hike	irving	dafa	itk	enhanced	postal	ouvrière
2	banks	killing	inclusively	temperature	kivalliq	pan	medication	goon
3	secularism	butts	davie	falun	expulsion	modernize	spill	outremont
4	barbados	abiding	netflix	investor	uqaqti	ambitious	drummond	pot
5	davie	admiral	triage	diluent	inuktitut	counterparts	privatization	entirety
6	havens	terrorists	maritimes	bitumen	qujannamiik	strengthening	essex	formidable
7	quebeckers	tape	groupe	41st	ccf	brampton	hyacinthe	triage
8	bloc	gateway	parlementaire	saanich	nunavummiut	oakville	ceos	québec
9	québécois	norman	budworm	gulf	nunavut	lifted	nanaimo	debout

As seen in Table 6, most words found via TF.IDF are clearly associated with their party—"recession" is the top word for the Conservative party while both "secularism" and "québécois" are top words for the Bloc Québécois, all of which are key issues for their respective parties.

Future Work

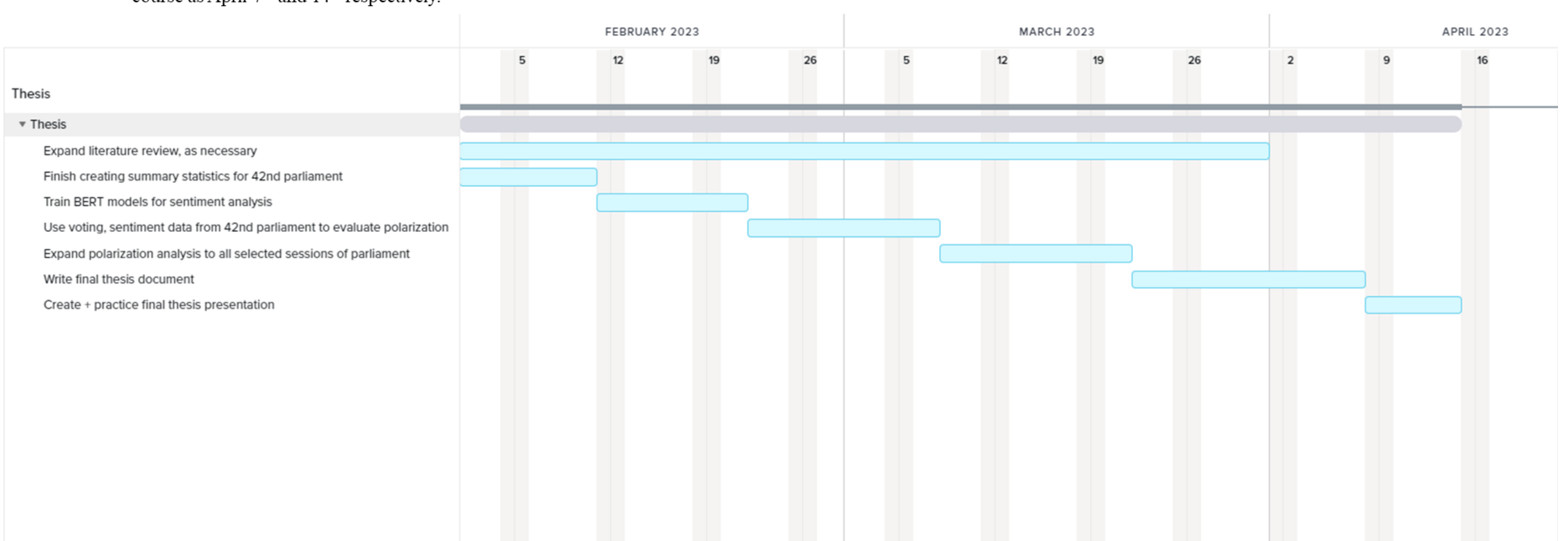
To continue work on this thesis some final summary statistics will be computed on the 42nd parliament, such as tracking interesting characters (e.g. Pierre Poilievre, the current leader of the Conservative Party who was also the 3rd most prominent Conservative speaker during the 42nd parliament). Next, work will be performed on creating AI models, especially BERT-like models for sentiment analysis. Following this, the methods described in Hanretty et al. [5] and DiMaggio et al.'s [12] paper will be applied to parliamentary voting data from each parliament to attempt to directly measure the polarization of the Canadian parliament. These methods will first be applied to the 42nd parliament, to measure the change in polarization over time within that single parliamentary session. Once this has been done, the methods will be expanded to all selected sessions, both to measure the change in polarization during these sessions and to attempt to measure the change in polarization between sessions. Using these results, the final thesis document will be written and the presentation detailing an overview of the findings from this thesis will be made.

A Gantt Chart showing an approximate timeline of the thesis project can be seen below in Figure 10.

Timeline: Gantt Chart

Figure 10: Gantt Chart of Future Work with Expected Dates of Completion:

As work will be continuing on the use of machine learning models for sentiment and polarization detection, it is likely that further academic research will be necessary to provide a more complete literature review and inform the final directions of this thesis. Primary sections of work (with the exception of expanding the literature review and finishing the summary statistics) are held to end on "every other Tuesday," the regular meeting time with my supervisor. Deadlines for the final thesis submission and presentation are set by the course as April 7th and 14th respectively.



References

[1] C. Testa, “Is polarization bad?,” European Economic Review, vol. 56, no. 6, pp. 1104–1118, Aug. 2012, doi: 10.1016/j.euroecorev.2012.04.005.

[2] “Welcome to the House of Commons of Canada,” Ourcommons.ca, 2018. https://www.ourcommons.ca/en

[3] “Votes - Members of Parliament - House of Commons of Canada,” www.ourcommons.ca. https://www.ourcommons.ca/members/en/votes

[4] T. Alsinet, J. Argelich, R. Béjar, S. Martínez. (2021). Measuring Polarization in Online Debates. Applied Sciences. Dec. 2021, doi: 10.3390/app112411879.

[5] C. Hanretty, B. E. Lauderdale, and N. Vivyan, “Dyadic Representation in a Westminster System,” Legislative Studies Quarterly, vol. 42, no. 2, pp. 235–267, Aug. 2016, doi: 10.1111/lsq.12148.

[6] N. D. Goet, “Measuring Polarization with Text Analysis: Evidence from the UK House of Commons, 1811–2015,” Political Analysis, vol. 27, no. 4, pp. 518–539, 2019.

[7] C. Moore, A. Rohan, “Dyadic Representation in Canadian Parliament,” April 2022

[8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Oct. 11, 2018.

[9] G. Abercrombie and R. Batista-Navarro, “Sentiment and position-taking analysis of parliamentary debates: a systematic literature review,” Journal of Computational Social Science, vol. 3, no. 1, pp. 245–270, Jan. 2020, doi: 10.1007/s42001-019-00060-w.

[10] K. Arceneaux, M. Johnson, and C. Murphy, "Polarized Political Communication, Oppositional Media Hostility, and Selective Exposure," The Journal of Politics, vol. 74, no. 1, pp. 174–186, Jan. 2012, doi: 10.1017/s002238161100123x.

[11] S. Ansolabehere and P. E. Jones, Dyadic Representation. Oxford University Press, 2011. doi: 10.1093/oxfordhb/9780199559947.003.0013.

[12] P. DiMaggio, J. Evans, and B. Bryson, "Have American's Social Attitudes Become More Polarized?," American Journal of Sociology, vol. 102, no. 3, pp. 690–755, Nov. 1996, doi: 10.1086/230995.

[13] W. Park, "How the views of a few can determine a country's fate," www.bbc.com. https://www.bbc.com/future/article/20190809-how-the-views-of-a-few-can-determine-the-fate-of-a-country

[14] J. Cook and S. Lewandowsky, "Rational Irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks," Topics in Cognitive Science, vol. 8, no. 1, pp. 160–179, Jan. 2016, doi: 10.1111/tops.12186.

[15] D. Ruppert, "What Is Kurtosis?: An Influence Function Approach," The American Statistician, vol. 41, no. 1, p. 1, Feb. 1987, doi: 10.2307/2684309.

[16] J. Esteban and D. Ray, "Linking Conflict to Inequality and Polarization," American Economic Review, vol. 101, no. 4, pp. 1345–1374, Jun. 2011, doi: 10.1257/aer.101.4.1345.

[17] C. A. Bail et al., "Exposure to opposing views on social media can increase political polarization," Proceedings of the National Academy of Sciences, vol. 115, no. 37, pp. 9216–9221, Aug. 2018, doi: 10.1073/pnas.1804840115.

[18] "Progress of generating sentiment lexicons from text in social media," SCIENTIA SINICA Informationis, 2014, doi: 10.1360/n112013-00167.

[19] “Lipad - Data - Browse the Canadian House of Commons,” www.lipad.ca. https://www.lipad.ca/data/ (accessed Jan. 10, 2023).

[20] Pandas, “Python Data Analysis Library — pandas: Python Data Analysis Library,” Pydata.org, 2018. https://pandas.pydata.org/

[21] NLTK, “Natural Language Toolkit — NLTK 3.4.4 documentation,” Nltk.org, 2009. https://www.nltk.org/

[22] B. Das and S. Chakraborty, “An Improved Text Sentiment Classification Model Using TF-IDF and Next Word Negation.” Accessed: Jun. 01, 2021. [Online]. Available: https://arxiv.org/ftp/arxiv/papers/1806/1806.06407.pdf

[23] W. Christian, H. Jansen, “Party System | The Canadian Encyclopedia,” thecanadianencyclopedia.ca, 2018. https://www.thecanadianencyclopedia.ca/en/article/party-system

[24] Alain-G Gagnon and B. Tanguay, Canadian parties in transition : recent trends and new paths for research. Toronto University Of Toronto Press, 2017.

[25] J.J. McCullough, “Canadian Political Parties,” The Canada Guide, 2013. https://thecanadaguide.com/government/political-parties/

[26] T. Sakamoto and H. Takikawa, "Cross-national measurement of polarization in political discourse: Analyzing floor debate in the U.S. the Japanese legislatures," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, USA, 2017, pp. 3104-3110, doi: 10.1109/BigData.2017.8258285.

[27] G. Abercrombie and R. Batista-Navarro, “Sentiment and position-taking analysis of parliamentary debates: a systematic literature review,” Journal of Computational Social Science, vol. 3, no. 1, pp. 245–270, Jan. 2020, doi: 10.1007/s42001-019-00060-w.

[28] “The Party Polarization and Issue Complexity Theory of Dyadic Representation,” Representation in Congress, pp. 23–46, Jul. 2015, doi: 10.1017/cbo9781316256626.003.

[29] H. Xu, B. Liu, L. Shu, and P. Yu, “BERT Post-Training for Review Reading Comprehension and Aspect-based Sentiment Analysis.” [Online]. Available: https://arxiv.org/pdf/1904.02232.pdf