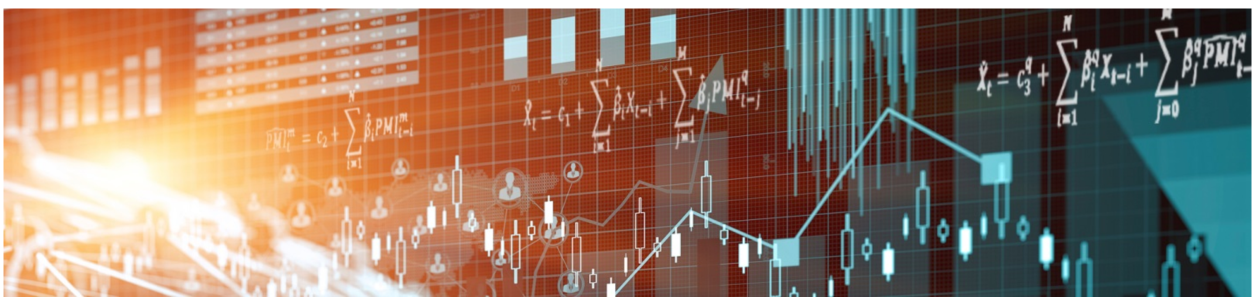
BANK OF CANADA
BANQUE DU CANADA

Staff Analytical Note/Note analytique du personnel 2019-5

Canada’s *Monetary Policy Report*: If Text Could Speak, What Would It Say?



by André Binette and Dmitri Tchebotarev

Canadian Economic Analysis Department
Bank of Canada
Ottawa, Ontario, Canada K1A 0G9
abinette@bankofcanada.ca

Acknowledgements

The authors are grateful to José Dorich, Gabriela Galassi and Eric Santor for helpful comments and suggestions. Special thank you to Boyan Bejanov for his technical help throughout the project and to Nicholas Curtis and Alexa Evans for excellent research assistance. We also thank Alison Arnot and Carole Hubbard for valuable editorial suggestions. All remaining errors are our own.

ISSN 2369-9639

Abstract

This note analyzes the evolution of the narrative in the Bank of Canada's *Monetary Policy Report* (MPR). It presents descriptive statistics on the core text, including length, most frequently used words and readability level—the three Ls. Although each Governor of the Bank of Canada focuses on the macroeconomic events of the day and the mandate of inflation targeting, we observe that the language used in the MPR varies somewhat from one Governor's tenure to the next. Our analysis also suggests that the MPR has been, on average, slightly more complicated than the average Canadian would be expected to understand. However, recent efforts to simplify the text have been successful. Using word embeddings and applying a well-established distance metric, we examine how the content of the MPR has changed over time. Increased levels of lexical innovation appear to coincide with important macroeconomic events. If substantial changes in economic conditions have been reflected in the MPR, quantifying changes in the text can help assess the perceived level of uncertainty regarding the outlook in the MPR. Lastly, we assess the sentiment (tone) in the MPR. We use a novel deep learning algorithm to measure sentiment (positive or negative) at the sentence level and aggregate the results for each MPR. The exceptionally large impacts of key events, such as 9/11, the global financial crisis and others, are easily recognizable by their significant effect on sentiment. The resulting measure can help assess the implicit balance of risks in the MPR. These measures (lexical innovations and sentiment) could then potentially serve to adjust the probability distributions around the Bank's outlook by making them more reflective of the current situation.

Bank topics: Central bank research; Monetary policy
JEL codes: E, E02, E52

Résumé

Dans cette note, nous analysons comment les explications données dans le *Rapport sur la politique monétaire* (RPM) de la Banque du Canada ont évolué. Nous présentons des statistiques descriptives sur le texte principal, y compris sa longueur, les mots les plus employés et son niveau de lisibilité. Même si chaque gouverneur se préoccupe avant tout des événements macroéconomiques du jour et du mandat de ciblage de l'inflation, nous constatons que les mots utilisés dans le RPM varient en fonction du gouverneur en poste. Il ressort aussi de notre analyse qu'en moyenne, le degré de complexité du RPM dépasse légèrement la capacité de compréhension du Canadien typique. Cependant, de récents efforts de simplification ont porté leurs fruits. À l'aide de la technique de plongement lexical et d’une mesure bien établie de la distance entre les mots, nous examinons l’évolution du RPM au fil du temps. Les innovations lexicales semblent coïncider avec d’importants événements macroéconomiques. Si les livraisons successives du RPM ont pris en compte les changements importants de la conjoncture économique, la quantification de l’évolution du texte peut nous aider à évaluer le niveau d’incertitude entourant les perspectives économiques que la publication laisse transparaître. Enfin, nous évaluons le degré de confiance manifesté dans le RPM, soit son ton. Nous nous servons ainsi d’un nouvel algorithme d’apprentissage profond pour mesurer le degré de confiance (positif ou négatif) exprimé dans chacune des phrases et, à partir de ces résultats, établir un degré de confiance global pour chaque livraison du RPM. L’impact exceptionnel de certains événements clés, comme les attentats du 11 septembre et la crise financière mondiale, ressort clairement des changements intervenus dans le degré de confiance. La mesure obtenue peut aussi nous aider à évaluer la résultante implicite des risques dans le RPM. Ces mesures (innovations lexicales et degré de confiance) permettraient ensuite d’ajuster les distributions de probabilité autour des perspectives de la Banque en les faisant mieux correspondre à la situation actuelle.

Sujets : Recherches menées par les banques centrales; Politique monétaire
Codes JEL : E, E02, E52

1 Introduction

Central banks put a lot of emphasis on communication because it is an integral part of effective monetary policy.[1] Clear communication helps economic agents understand current economic conditions and the reasoning behind monetary policy actions. In Canada, monetary policy decisions are explained through press releases eight times a year. At four of those eight times, the Bank of Canada further explains its monetary policy decisions in the *Monetary Policy Report* (MPR). The MPR discusses the pertinent developments in Canada and abroad, their implications for the Bank's projection and the principal risks associated with the inflation outlook.[2] Given the Bank's clear mandate of keeping inflation low, stable and predictable, the MPR is very attentive to the evolution of inflation. In addition to these written reports and associated press conferences,[3] Bank officials recently began to give speeches on the economic conditions since the previous MPR. These speeches, called economic progress reports, take place after any policy decision that is not accompanied by an MPR. Finally, at least twice a year, both the Governor and the Senior Deputy Governor appear before committees in the House of Commons and Senate.[4] These appearances serve to hold the Bank publicly accountable to Parliament; members of Parliament and the Senate ask questions about recent developments related to all the Bank's functions. Together, these communication strategies increase transparency around decisions that affect the economic and financial well-being of Canadians.

In this note, we examine the text of the MPR from the first issue of the report in 1995 to the October 2018 issue. Because the MPR is the most exhaustive document produced by the Bank about the economic outlook, it is a good vehicle to use to study the evolution of the Bank's views and to extract underlying signals about the economy. In particular, we study how the MPR has evolved, the influence of the different governors and key macroeconomic events, and what the MPR can possibly say about uncertainty and the balance of risks. This work is part of a growing literature on assessing and extracting information from documents written by policy-makers.[5] While some of the techniques we use are relatively simple (e.g., word cloud and readability statistics), a few are cutting-edge natural language processing (NLP) and deep learning methods, such as the word mover's distance (WMD) technique and transfer learning for detecting changing economic conditions and measuring sentiment (tone).

After processing all MPR PDFs to extract the core text, we create a stylized account of the last three decades of Canadian macroeconomic history to contextualize our analysis. We look at different metrics (the three Ls—language, readability level and length) that expose the stylistic choices made by each Governor, from Governor Carney's preference for more detailed reports to Governor Poloz's penchant for writing shorter documents. Although each Governor focuses on the macroeconomic events of the day and the mandate of inflation targeting, the language used in the MPR varies somewhat from one Governor's tenure to the next. The metrics also suggest that the readability level of the MPR has been, on average, slightly more complicated than readers with the average level of education in Canada would be expected to understand, although efforts have been made recently to simplify it.

We then assess how the MPRs differ from each other, using the WMD technique. Significant shifts in this metric identify changes in the Bank's narrative (lexical innovations), which often coincide with large economic shocks, and could potentially serve as a proxy for the Bank's implicit assessment of uncertainty. The premise

1. Blinder et al. (2008) present a review of the literature, and Poloz (2018) discusses the importance of communication at the Bank of Canada.

2. A primer on the monetary policy decision-making process at the Bank of Canada is available online. For more detail, see Murray (2013) and Macklem (2002).

3. Each MPR is discussed at a press conference held by the Governor and Senior Deputy Governor where they answer journalists' questions about the report, current events and any other topic.

4. These committees are the House of Commons Standing Committee on Finance and the Standing Senate Committee on Banking, Trade and Commerce. These two bodies review Canadian government work pertaining mainly to finance and economics.

5. See, for example, Correa et al. (2017), Hansen and McMahon (2015), Hendry (2012), and Hendry and Madeley (2010).

is that new situations, or situations that have not been seen in a while, are expected to generate higher uncertainty as policy-makers learn about their overall impact on the economy. For example, the proposed measure of changes spiked during the global financial crisis (2008-09). At the time, the ramifications of the crisis were unknown, and policy-makers faced significant uncertainty.

Finally, we look at the sentiment in the MPR. The use of transfer learning allows us to apply deep learning methods to a limited sample of text and train a neural net to identify positive and negative sentences. The resulting sentiment score for each MPR could suggest an economic outlook that is more or less favourable than a single number from the projection would indicate. Therefore, the sentiment score can be used to assess the Bank's underlying view about the balance of risks. Extensive use of positive or negative references could suggest a tilt in the balance of risks not explicitly acknowledged by policy-makers. In addition to tracking relatively well the different events since 1995, the proposed measure of sentiment also captures the instability following the global financial crisis with a sentiment score that is often negative.

The rest of the note is organized as follows: Section 2 briefly presents the text data we use in the project. Section 3 highlights multiple events that have occurred over the history of the MPR. Section 4 discusses simple metrics describing the nature of the MPR during the different Governors' tenures. Section 5 assesses the differences in the narrative between MPRs. Section 6 presents a measure of sentiment. Finally, Section 7 summarizes our findings.

2 Data

The MPR has been published since 1995. At first, it was published twice per year, in May and November. In 2000, the Bank began publishing a short update to the MPR in February and August. In 2002, the publication schedule was shifted back one month—MPRs were released in April and October with updates in January and July. Finally, in July 2009, the January and July updates were expanded to be full-featured MPRs, and the publication format and schedule took on their current form—a full MPR in January, April, July and October.

Before a text can be analyzed, all documents must be prepared, and choices regarding which parts to include must be made. To create the corpus for our study, we downloaded all MPRs from the Bank's website in PDF format. To be usable, these documents had to be converted from PDF to text format. While many tools are available to do this conversion, the process is imperfect.[6] Each tool introduces artifacts into the text caused by the PDF format and the way the information is encoded. Common issues include the omission of spaces between words; the use of ligatures; and the miscoding of non-standard text elements, such as quotation marks, currency symbols and hyphens. As well, conversion tools will indiscriminately transcribe footnotes, charts, tables, headers and footers. All this increases the difficulty of extracting meaningful information. We correct these issues so that they do not influence our results.[7] Also, elements and passages of the text that are deemed not part of the core text—technical boxes, annexes, footnotes, charts, tables and references—are removed manually. This helps us simplify the analysis and avoid any impact on the results from occasional additions to an MPR. These elements appear sporadically and typically do not contain information about the Bank's outlook that is not already elaborated on in the main text.[8]

Finally, we exclude MPR updates from the study. In 2000, to shorten the time between official communications, the Bank introduced short updates to the publication schedule. While these showed the evolution of economic conditions in Canada and abroad, they were supported by less staff analysis than the full-length MPRs. For example, they rarely included comprehensive projections derived from a macroeconomic model. By including only complete MPRs in our study, we ensure the comparability of the MPRs across time by

6. We used an open source utility called pdftotext to convert the PDF documents into plain text files.
7. While most of these issues were corrected programmatically using regular expression techniques, some were fixed manually.
8. The cleaned text files are released with this paper.

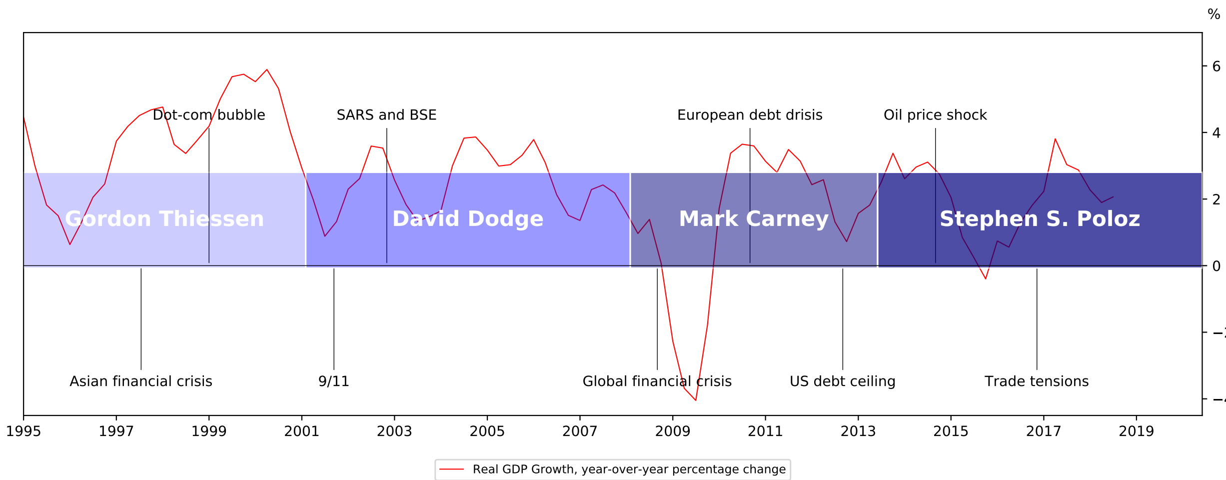
eliminating possible distortions due to the length, objectives or composition of the documents.

Our final corpus contains the textual information of 67 MPRs, close to 15,000 sentences and about 375,000 words (around 5,000 of these are unique words).

3 Canada's macroeconomic history since 1995

In this section, we provide a high-level summary of the last three decades of Canada's macroeconomic history. It is by no means an exhaustive account—it is meant to refresh the reader's memory of the major events discussed in the MPR from 1995 to 2018 and will be used as context for the analysis that follows. **Chart 1** illustrates the major events.

Chart 1: Canada's macroeconomic history through the lens of the *Monetary Policy Report*



Note: SARS is severe acute respiratory syndrome. BSE is bovine spongiform encephalopathy.
Source: Statitics Canada
Last observation: 2018Q3

When the MPR was first published in 1995, Gordon Thiessen was Governor of the Bank of Canada. The Canadian economy progressed steadily until 1997, when the Asian financial crisis hit the global economy. Several Asian economies experienced rapid currency devaluations, capital outflows and weak economic activity. Fears of contagion were extremely high, and the International Monetary Fund intervened.

After two years of turmoil, global economic growth resumed, driven in part by the boom in information and communications technology, which lasted until 2000. The end of the dot-com bubble and the terrorist attacks of 9/11 pushed the United States into a mild recession. In Canada, growth slowed significantly, dropping slightly below zero in the third quarter of 2001. That year, David Dodge took over from Gordon Thiessen as Governor of the Bank of Canada.

The following year, economic activity was affected by the SARS (severe acute respiratory syndrome) epidemic and the outbreak of spongiform encephalopathy (BSE) or mad cow disease. While the SARS epidemic, which provoked significant public health concerns, was a frightening prospect at the time, it had limited economic consequences beyond losses in the travel and tourism industry. Similarly, mad cow disease was devastating for the rural economy, especially in Alberta, but its overall effect on gross domestic product (GDP) was limited. After these outbreaks, Canada entered a period of economic stability. Indeed, the period 2001–08 had one of the lowest amounts of volatility in GDP growth in history.[9]

9. As measured by the standard deviation of a 10-year rolling window of quarterly real GDP growth.

In 2008, Mark Carney took over from David Dodge as Governor of the Bank of Canada. He did not have long to settle in before the brunt of the global financial crisis hit. The bursting of the US housing bubble—also referred to as the subprime mortgage crisis—shook the global financial system to its foundations. The use of sophisticated financial instruments such as mortgage-backed securities, collateralized debt obligations and credit default swaps exposed many large market participants to the bubble and subsequent crash. As credit dried up, large financial firms were under tremendous pressures, and authorities had to intervene. The result was a severe drop in economic activity across the world, triggering the first recession in Canada since 1992.

Throughout 2010, the level of economic activity struggled to return to its pre-crisis levels, and unemployment remained high. After that, projected growth failed to materialize year after year because of adverse shocks that buffeted the Canadian economy—from the European debt crisis starting in 2010 to the oil price shock in 2014–15.[10] In 2013, Mark Carney left the Bank of Canada, and Stephen S. Poloz assumed the role of Governor.

In 2016, Donald Trump was elected president of the United States, Canada's largest trading partner. The new president's position on NAFTA and other international policies introduced tensions into the macroeconomic landscape. Despite these tensions, in 2017 Canada posted its highest gain in real economic activity since the global financial crisis. This impressive gain followed two years of subpar growth affected by the oil price shock (2014–15).

4 The three Ls: Language, level and length

The text of the MPR has evolved significantly since 1995 as each Governor left his mark and Canada's economic landscape changed. In this section, the three Ls of the MPR are presented. An examination of the language used over time is followed by an assessment of the level of complexity of the text and statistics about the length of the documents.

4.1 Language: The most used terms in the *Monetary Policy Report*

One way of presenting the language used in the MPR is with word clouds. A word cloud is an intuitive visualization of the relative frequency of terms that occur in each text. The more frequently a term appears, the larger its visual representation in the cloud. The word cloud in **Figure 1** captures the text of the entire history of the MPR.[11]

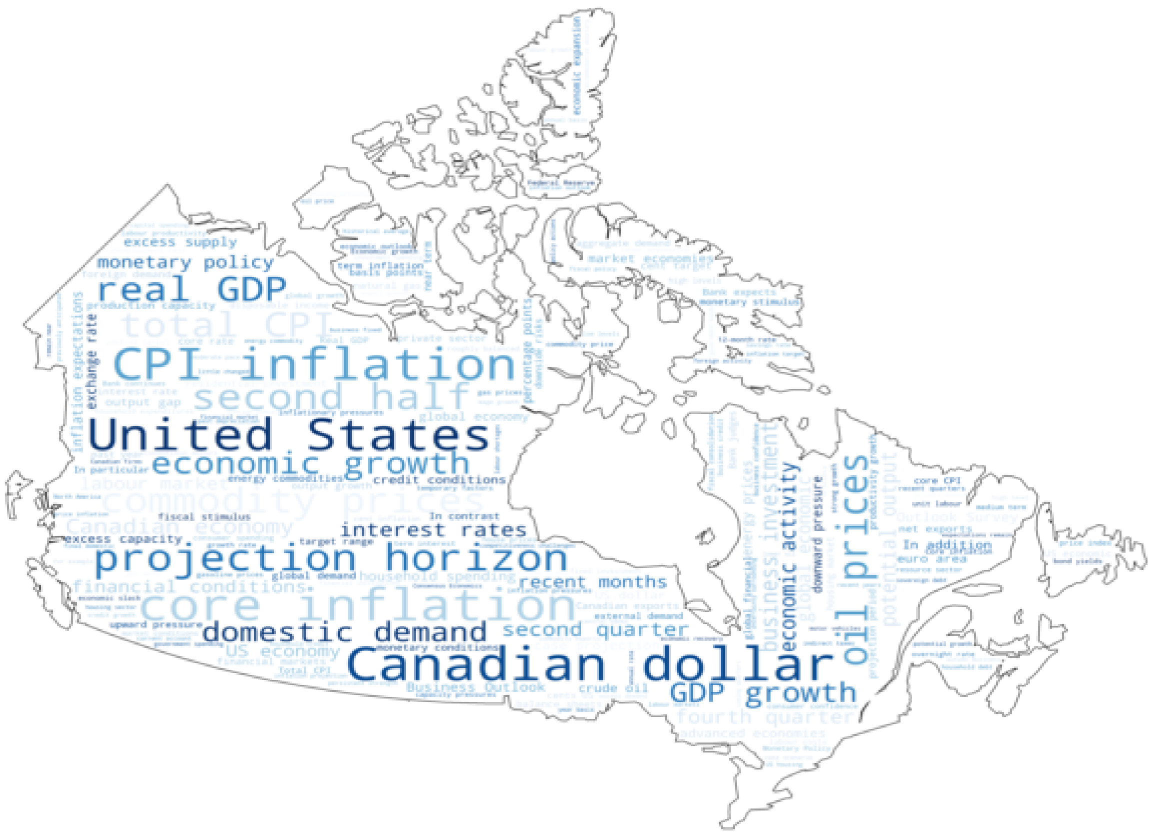
The terms that appear in the word cloud are the ones readers might expect, given the Bank's mandate and Canada's economic situation. Frequently occurring are "CPI inflation" and "core inflation"—terms related to the Bank's mandate; "Canadian dollar" and "United States"—terms related to international trade; "commodity price" and "oil price"—terms reflecting Canada's dependence on the resource sector; and "real GDP" and "GDP growth"—measures of economic activity.

While **Figure 1** is representative of the entire history of the MPR, different subperiods have additional themes. In the early days of the inflation-targeting regime, under Governor Thiessen, additional emphasis was put on the monetary conditions index (MCI). At the time, the MCI was built to summarize the effect that monetary policy has on the economy through both interest rates and the exchange rate.[12] Under Governor Dodge, the focus included production capacity, while during Governor Carney's tenure we observed the emergence of the terms "financial conditions" and "credit conditions" as the global financial crisis hit the world economy. In recent years, Governor Poloz's additional emphasis has been on endogenous growth,

10. This period is often described in the discussion as "serial disappointment." For more on this, see Guénette et al. (2016).
11. We used bigrams—a combination of two adjacent words—since they usually contain more meaning than single words.
12. For more on the origin of the MCI, see Freedman (1995).

with terms like “business investment” and “potential output” becoming more predominant.[13] The term “oil prices” moved front and centre as the Canadian economy had to adjust to lower prices in 2015. Lastly, the importance of the Bank’s *Business Outlook Survey* has grown significantly in recent years, under Governor Poloz.[14]

Figure 1: The language in the MPR clearly illustrates the Bank’s mandate and the economic landscape of Canada



4.2 Level: Readability of the *Monetary Policy Report*

As mentioned earlier, communication from the central bank is critical for the effectiveness of monetary policy. In a staff analytical note, Deslongchamps (2018) assessed the readability of multiple Bank documents—the MPR, the *Financial System Review*, speeches and others—over three years and concluded that they could be simplified to reach a broader audience. Deslongchamps used the Gunning Fog Index (GFI) as the sole measure of readability. While our note focuses only on the MPR, we extend our analysis to 1995. Given a lack of consensus regarding which measure is most appropriate, we use multiple measures and average them to get the final assessment. Six statistics are calculated from the MPR text to assess readability, each of which produces a score equivalent to the education grade level required to understand the text (e.g., a 13 corresponds to the first year of an undergraduate degree). These measures are the Flesch Kincaid Grade Level Formula, the Coleman Liau Index, the Gunning Fog Index, the Simple Measure of Gobbledygook Index, the Automated Readability Index and the Dale-Chall Score.[15] The methods all have similar approaches,

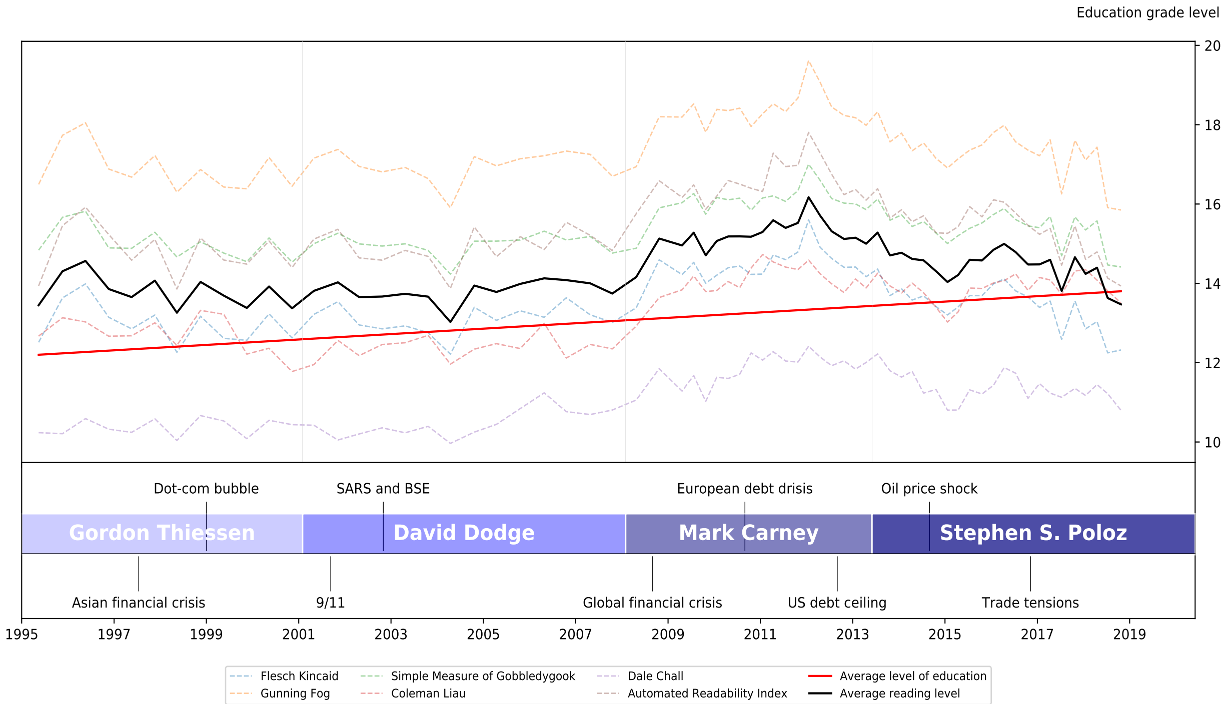
13. For a discussion on natural economic growth, see Poloz (2013).
14. For details about the Bank of Canada’s *Business Outlook Survey*, see Martin (2004).
15. For information about each measure, see the following Wikipedia page or the original papers: Kincaid et al. (1975), Coleman and Liau (1975), Gunning (1952), McLaughlin (1969), Senter and Smith (1967), and Dale and Chall (1948).

combining the average length of words or sentences, the number syllables per word or the number of complex words.

The final resulting overall readability assessment is compared with the average education level in Canada calculated using publicly available Census data. This comparison is natural since these readability measures have been designed for this purpose.[16] Over the past 25 years, the average education level has increased from about 12 to almost 14, indicating that the average Canadian now has some post-secondary education.

Chart 2 shows the average readability of the MPR over the years and compares the results of the individual measures with the average level of education reached in Canada. While we focus on the average grade, all six measures share a common dynamic: they increase during Governor Carney's tenure and fall back in recent years. Indeed, this is a common finding in the literature—measures of readability tend to be highly correlated.[17] That said, the grade levels could be, and are, very different across measures in our study. The GFI suggests the MPRs are reserved for highly educated people, while the Dale-Chall score indicates the MPRs are accessible for people in their final years of high school. As we said above, no real consensus has emerged in the literature as to which measure is most appropriate for analyzing a given text. For that reason, we remain agnostic about the right measure and discuss the average result.

Chart 2: The gap between the readability level of the MPR and the average level of education in Canada has reached its lowest level in history



Note: SARS is severe acute respiratory syndrome. BSE is bovine spongiform encephalopathy.

Last observation: October 2018 MPR

The average readability measure suggests the MPR has been written for an audience that is not representative of Canadians with the average level of education; there is a gap between the readability level of the MPR and the average level of education of about two to three years' worth of education. This gap reached

16. The educational attainment level in Canada is compiled using seven groups: grades 0 to 8 (level 5), some secondary (level 9.5), grades 11 to 13 (level 12), some post-secondary (level 14), post-secondary certificate or diploma (level 15), bachelor's degree (level 16), and graduate degree (level 19).

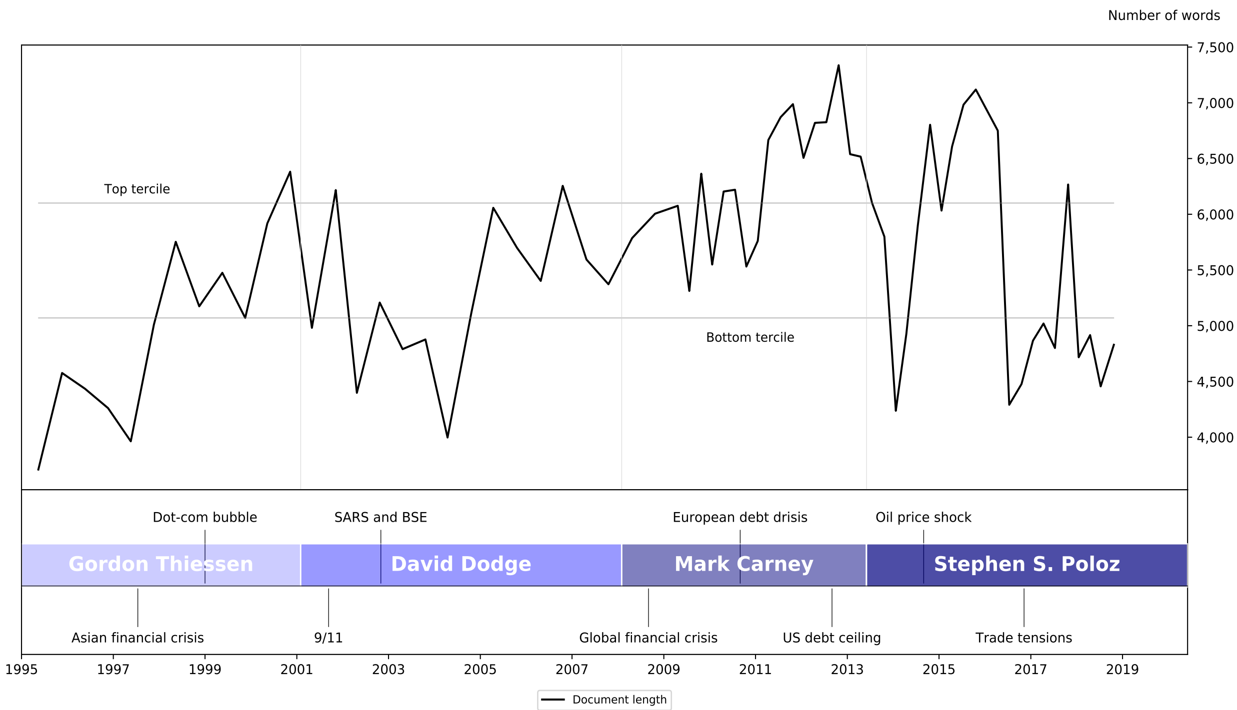
17. This is discussed in VanOosten, Tanghe and Hoste (2010).

its largest point during the global financial crisis under Governor Carney. Since then, under Governor Poloz, the gap has shrunk to its lowest level in history. This is due in part to increases in the average level of education reached in Canada but also, more importantly, to the significant improvement in the readability level of the MPR. Indeed, for the first time, the readability level of the October 2018 MPR was below the average education level in Canada, a welcome development for the central bank's communication strategy and mandate.[18]

4.3 Length: How long is the *Monetary Policy Report*?

We complete Section 4 by examining the length of the MPR over time, as measured by a simple word count. While it is difficult to determine an appropriate benchmark for this metric, *ceteris paribus*, a shorter MPR is easier to read. **Chart 3** shows that the average length of the MPR has increased since 2009 from about 5,200 to 6,000 words, an increase of more than 15 per cent. As the chart shows, Governor Carney and Governor Poloz have produced the longest MPRs. Additionally, spikes occur after significant events that warrant further explanation, like the Asian financial crisis, 9/11, the global financial crisis and the recent oil price shock. At the same time, also under Governor Poloz, some recent MPRs have been among the shortest in history, with about half of them in the lower tercile (below 5,070 words).

Chart 3: The MPR has gotten longer over time, but a reversal appears to be in play



Note: SARS is severe acute respiratory syndrome. BSE is bovine spongiform encephalopathy.

Last observation: October 2018 MPR

18. It is assumed that reaching a broader audience is beneficial because it enables more discussions among multiple agents.

5 Measuring lexical innovations between MPRs

There are many channels through which uncertainty could have a negative impact on business activity. The economic intuition is that when businesses and individuals make investment or spending decisions, they require some degree of certainty, especially since many business decisions require an upfront cost and pay off only later. When there is a lot of uncertainty, such as price volatility in financial markets or rapidly changing economic conditions, agents tend to put off spending and investment, which affects economic activity. This section focuses on building an index of lexical innovation, which can be used to identify changing economic conditions and serve as a proxy for the Bank's implicit assessment of uncertainty. The idea is that new conditions, or situations that have not been seen in a while, are expected to generate greater uncertainty as policy-makers learn about their impacts on the economy. In contrast, when situations are familiar (i.e., discussed previously in recent MPRs), uncertainty will be assessed as normal.

We assume that large shifts in the macroeconomic landscape will be reflected in the analysis and discussion in the MPR. These shifts could then be quantified by measuring how different one MPR is from those that precede it. To do this, we use the WMD measure (Kusner et al. 2015) to capture *semantic* innovation as opposed to only *syntactic* innovation. To illustrate this difference, consider two sentences: "I am going for a walk with my dog" and "We are taking our poodle to the park." Syntactically, these sentences contain no common words, but semantically they are quite similar—both refer to taking a dog for a walk. A naïve algorithm would score their similarity as 0, but the WMD approach more appropriately assigns them a much higher similarity score of about 0.6.[19]

Using word embeddings, the WMD approach overcomes a significant weakness of one-hot vectors, where the words "powerful," "strong," "forceful" and "legume" are all equidistant to each other.[20] We first put the documents in an embedding space. Specifically, we use GloVe word embeddings to map each word to a vector of 300 dimensions ($\mathbb{R}^{300}$).[21] The vectors are computed in such a way that each of the 300 dimensions encodes information about the meaning of words. This method allows us to compute distances in the meanings of words, which are aggregated to calculate a distance between MPRs. The aggregation algorithm moves a conceptual "mass" from the set of word vectors representing one MPR to a set of word vectors representing another MPR. It does this in a way that minimizes the work (mass $\times$ distance) required to get one document to overlap with another completely. The minimum amount of work required to do this transformation is precisely the WMD—the measure of similarity between an MPR and those preceding it.[22]

Chart 4 shows the WMD computed using lags of 1, 4, 8 and 16 previous MPRs. The distances are normalized to have zero mean and a variance of one because the WMD has no economic interpretation. One MPR is compared with previous MPRs[23] because we are interested in new economic conditions that did not appear previously and that might introduce greater uncertainty. Because we use a longer history than one MPR, our algorithm has a longer-term memory; therefore, conditions that emerged some time ago and then re-emerge are not perceived as entirely new conditions. As well, if a specific new condition (e.g., the oil price shock) is persistent over multiple MPRs, a WMD with a longer lag can correctly detect uncertainty as remaining elevated or fading gradually. An algorithm with a shorter lag structure might assume that a shock that affects the text of two consecutive MPRs is the new normal; it could misleadingly detect no change between MPRs and so report a decrease in uncertainty. However, the drawback is that using a longer history shortens the sample size accordingly. In practice, algorithms using different numbers of lags tend to tell the same broad story (**Chart 4**). Thus, the trade-off for losing a portion of history is not necessarily

19. The WMD similarity score that we implement is bound between 0 and 1 by construction. See Chartbeat Labs (2018) for details.

20. Word embeddings transform language into a numerical form (vectors), where semantically similar words have similar vectors. This is not the case with one-hot vectors, where each vector is assumed to be completely different from each other.

21. For more detail, see Pennington, Socher and Manning (2014).

22. This measure uses an existing measure called Earth Mover's Distance. For details, see Kusner et al. (2015).

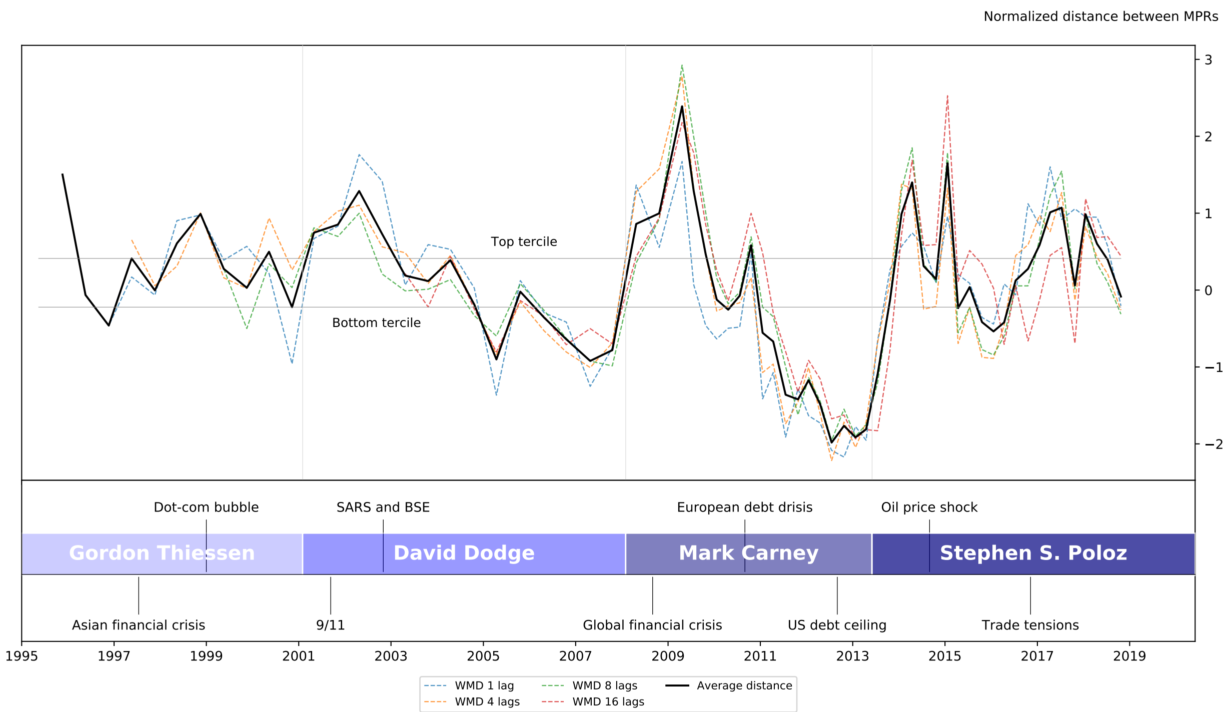
23. To compute the distance between one MPR and a set of MPRs, we concatenate the set of MPRs to be one long MPR.

worth the better theoretical characteristics of an algorithm with more lags. Furthermore, since uncertainty is not observable, it is difficult to judge which series has the best properties.[24]

The proposed measure of change shows the largest spikes are around the global financial crisis in 2008–09, the oil price shock in 2015 and, to a lesser extent, the events of 9/11 and SARS. It has also spiked on a few occasions since President Trump was elected and began to introduce new trade policies. During those times it is plausible to think that forecasting economic growth was less straightforward and more uncertain. Appropriately, the measure responds most strongly to the global financial crisis, given that it has had the biggest macroeconomic impact. After the SARS episode and the global financial crisis, the measure recedes to low levels, reflecting a relative dearth of new events.

In the end, by calculating a measure of lexical innovations from the MPR, we obtain a current, albeit implicit, evaluation of the potential level of uncertainty around the central bank's outlook, which can then be used to build a more representative probability distribution. The use of terciles (or any other percentiles) could help identify periods of high, normal or low uncertainty.

Chart 4: The text of the MPR changes significantly in response to large shocks



Note: SARS is severe acute respiratory syndrome. BSE is bovine spongiform encephalopathy.

Last observation: October 2018 MPR

6 Assessing sentiment in the MPR

Measuring the changes in the narrative of the MPR is a worthwhile goal and could provide valuable information for economic agents, such as the degree of uncertainty perceived by the central bank, but it does not paint a complete picture. It measures only the intensity of the change in economic conditions, not the direction. To determine the direction of the change in economic conditions, a measure of sentiment (positive or negative) is calculated.

Since we do not have a large enough sample to operate at the level of complete individual MPRs, we asses

24. For a discussion, see Baker, Bloom and Davis (2016) and Alexopoulos and Cohen (2015).

the MPRs at the level of sentences. We divide each MPR into sentences, assess each sentence independently, and then amalgamate those results back to the MPR level to create a sentiment index. The MPR is broken down into standalone sentences using the spaCy[25] sentence parser, which yields roughly 15,000 sentences.[26] We posit that each sentence on its own belongs to one of three classes: positive, negative or neutral. Positive sentences contain information that would lead the reader to believe the Canadian economy is doing well. An example of such a sentence is "Business investment is expected to expand at a solid pace." Naturally, negative sentences mean the opposite; for example, the following sentence is considered to be negative: "The resulting weaker-than-expected foreign demand would be an additional drag on Canadian exports and business investment." Neutral sentences are purely factual; they contain facts but nothing that suggests the economy is heading in any direction. For example, these sentences are neutral: "The Bank is continuing to monitor NAFTA renegotiations," and "The comparable figure for the United States is about 2 per cent."

From the set of 15,000 sentences, we select 300 of each type, for a total data set of 900 sentences labelled either "positive," "negative" or "neutral." Of these, 810 were picked randomly to be the training set, and 90 were picked to be the validation set. As well, 100 additional sentences were picked and labelled to be the test set; these are not used in the development phase and serve as a real-world assessment of the algorithm.

Next, to classify the remaining MPR sentences, we need an algorithm to distinguish between positive, negative and neutral sentences. Deep learning tends to perform significantly better at this task than other types of algorithms. While deep learning is an incredibly powerful technique, not long ago it could be applied only to large data sets because of the incredibly large number of parameters that require estimation—typically many millions. However, recently researchers have been able to reduce the amount of data necessary to train a neural net. This technique is called transfer learning. Transfer learning allows us to train a base neural net model on a much larger data set that is otherwise similar to the one we want to analyze, and then to fine-tune it to perform well on the more specialized data set. In our case, the base neural net predicts the next word in a body of English language text and is later trained to differentiate whether sentences from the MPR are positive, negative or neutral.[27]

The base neural net model operates on the powerful LSTM architecture. In an LSTM neural net, instead of having a single stateless function representing each neuron, the neural net has states, or, more colloquially, memory. With this memory, LSTM-based neural nets can learn long-term dependencies that are always present in language (Karpathy 2015). More specifically, our work is based on the averaged stochastic gradient (ASGD) weight-dropped long short-term memory (AWD-LSTM) architecture from Merity, Keskar and Socher (2017). They apply a set of regularization strategies to the standard LSTM to achieve a state-of-the-art lingual model that consists of predicting the next word in a sentence. Regularization is necessary to prevent overfitting,[28] which makes the neural net unable to generalize. Our estimation strategy is based on the work of Howard and Ruder (2018), who propose the Universal Language Model Fine-tuning (ULMFiT) transfer learning approach. Building on the success of Merity, Keskar and Socher (2017), Howard and Ruder (2018) use transfer learning to train a neural net to classify IMDb movie reviews as positive or negative. We employ the same approach, but instead of training the network to classify IMDb reviews, we substitute our labelled MPR sentences.

The specific ULMFiT transfer learning approach has three steps. First, a neural net is trained to predict the next word in a passage of text. For example, when the phrase "to avoid sunburn I put on" is input into the model, the model would predict "sunscreen."[29] The neural net is trained on WikiText-103, a data

25. See Explosion AI (2018) for details.

26. In a second step, the resulting text is carefully examined and programmatically corrected to yield clean sentences.

27. The fastai deep learning library is used in this note—Howard et al. (2018).

28. The symptom of overfitting is a much better performance on the training set than on the validation set, making the neural net less useful for prediction.

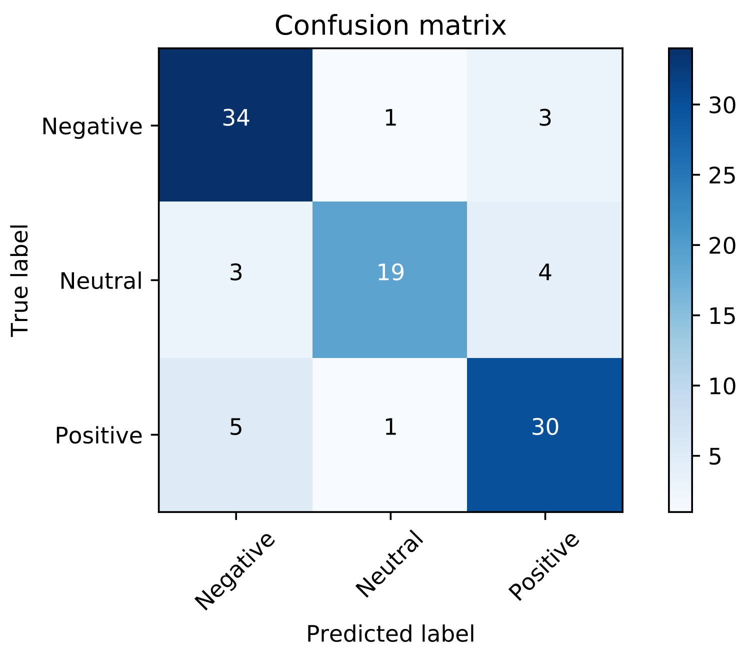
29. The model is a 3-layer LSTM, with 400-dimensional word embedding.

set of close to 104 million words from Wikipedia.[30] Once the neural net has learned the English language, it has encoded some form of meaning into its millions of parameters. Second, the same model is retrained or fine-tuned on the 15,000 MPR sentences. Now, instead of predicting the next words in sentences of a general English corpus, the neural net is trained to predict words in sentences from MPRs. The length of this training sample is only about 375,000 words. Traditionally, this would be nowhere near the amount of data required to train a neural net, but with transfer learning this limitation is overcome. When the neural net trains on the 375,000 MPR words, it does not start from scratch. Instead, it carries on from where the Wikipedia training left off.

At this point, the algorithm can predict only the next word in an MPR sentence, so the third step is to turn to the actual classification task. Two layers are added at the end of the neural net to predict the sentiment of each sentence. While this may sound complicated, these layers are simply a couple of affine transformations, which use the information contained in the LSTM hidden states to output probabilities of a given sentence belonging to each sentiment class.

As benchmarks, lexical methods[31] achieved about 60 per cent accuracy on our test set, and a support vector machine[32] with a linear kernel achieved roughly 75 per cent accuracy in classification. In comparison, the proposed neural net performs extremely well on the validation set with 95 per cent accuracy, and well on the test set, with accuracy of about 85 per cent (**Figure 2**). The proposed approach is very good at classifying positive and negative sentences, although it struggles a bit with the neutral sentences. These results are impressive, especially considering the subjectivity inherent in the task.

Figure 2: Neural net performs well across the board but is better at identifying positive and negative sentiment



Next, the individual sentence scores must be combined into a per-MPR figure. To do this, we divide the

30. In practice, the pre-trained weights are downloaded without running the training algorithm itself since our results with the same parametrization would be similar.

31. A simple version of SO-CAL is used as a benchmark; see Taboada et al. (2011) for a description.

32. See the original Cortes and Vapnik (1995) paper for a description.

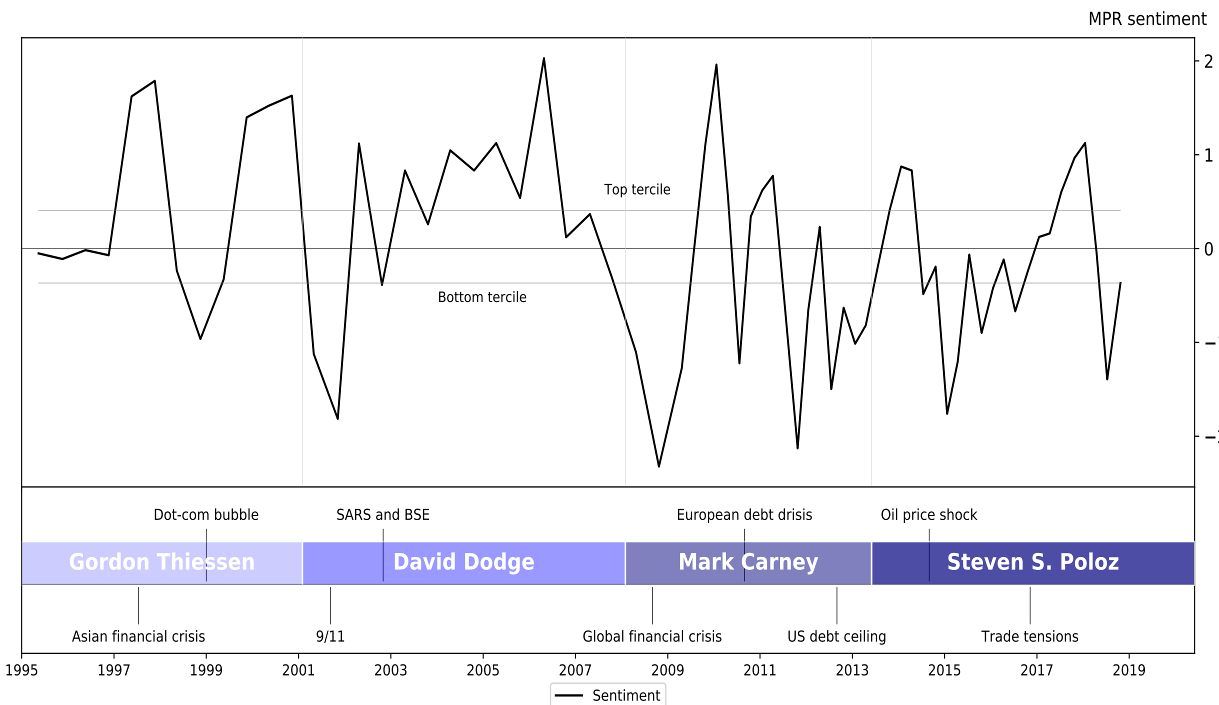
balance of positive to negative sentences by the total number of sentences.

$$\frac{\#positive - \#negative}{\#positive + \#neutral + \#negative}$$

The resultant series is presented in **Chart 5**, where macroeconomic events described in Section 3 are well captured. The Asian financial crisis, 9/11 and the global financial crisis have all had a significant impact on the tone in the MPR, which reached a low point on two occasions (after 9/11 and during the global financial crisis). The ups and downs after 2009 can be seen as well with repeated negative shocks hitting the global economy, though the overall tone since the recession has been pessimistic. Only recently, after the oil price shock began to fade, did the tone of the MPR improve in step with the recovery of the Canadian economy.

The resultant sentiment score for each MPR could be used to assess the Bank's underlying view about the balance of risks. An extensive use of positive or negative references could suggest a tilt in the balance of risks not explicitly acknowledged by policy-makers. If so, this could serve to adjust the probability distributions around the Bank's economic outlook. It would also have the advantage of making the probability distributions more reflective of the current situation. Again, here the use of terciles (or any other percentiles) could be useful. They can help identify optimistic times versus pessimistic ones.

Chart 5: Measure of MPR sentiment reflects macroeconomic events



Note: SARS is severe acute respiratory syndrome. BSE is bovine spongiform encephalopathy.

Last observation: October 2018 MPR

7 Conclusion

In this note, we apply multiple techniques to analyze the evolution of the narrative in the Bank of Canada's *Monetary Policy Report*. To help contextualize the analysis, we present a brief history of the main economic events since 1995. The note presents descriptive statistics, including length, the most frequently used words (language) and the reading level. These metrics show the stylistic choices made by each Governor, from Governor Carney's more detailed reports to Governor Poloz's more succinct style. While each Governor focuses on the macroeconomic events of the day and the Bank's monetary policy mandate, the language used in the MPR varies over time. The metrics also suggest that the MPR has been, on average, slightly more complicated than the average Canadian would understand. We highlight recent efforts to simplify it. Indeed, the October 2018 MPR is the first to have a readability level below the average level of education in Canada.

The note also assesses how MPRs differ from each other, which is a potential proxy for perceived uncertainty from the monetary policy authority around the economic outlook. The algorithm highlights large perturbations in the macroeconomic history of Canada when the themes changed drastically. Heightened levels of this measure of lexical innovation appear to coincide with important macroeconomic events where uncertainty was elevated.

Finally, we use a novel deep learning algorithm to measure sentiment (positive or negative) at the sentence level and aggregate the results for each MPR. The exceptionally large impacts of key events such as 9/11, the global financial crisis and others are easily recognizable and have moved sentiment significantly. This sentiment measure could help assess the central bank's view of the implicit balance of risks in the MPR.

Since the measures of lexical innovation and movement in sentiment coincide with important macroeconomic events, they could both potentially be used to build more representative probability distributions around the central bank's outlook.

References

Alexopoulos, M. and J. Cohen. 2015. "The Power of Print: Uncertainty Shocks, Markets, and the Economy." *International Review of Economics & Finance* 40 (C): 8–28. https://ideas.repec.org/a/eee/reveco/v40y2015icp8-28.html.

Baker, S. R., N. Bloom and S. J. Davis. 2016. "Measuring Economic Policy Uncertainty." *The Quarterly Journal of Economics* 131 (4): 1593–1636. https://ideas.repec.org/a/oup/qjecon/v131y2016i4p1593-1636..html.

Blinder, A. S., M. Ehrmann, M. Fratzscher, J. de Haan and D.-J. Jansen. 2008. "Central Bank Communication and Monetary Policy: A Survey of Theory and Evidence." Available at https://ideas.repec.org/a/aea/jeclit/v46y2008i4p910-45.html, *Journal of Economic Literature* 46 (December): 910–945.

Chartbeat Labs. 2018. *Textacy: NLP, Before and After spaCy.* Available at https://github.com/chartbeat-labs/textacy.

Coleman, M. and T. L. Liau. 1975. *"A Computer Readability Formula Designed for Machine Scoring".* Journal of Applied Psychology 50 (2): 283-284.

Correa, R., K. Garud, J. M. Londono and N. Mislang. 2017. *"Sentiment in Central Banks' Financial Stability Reports".* International Finance Discussion Papers No. 1203. Available at https://ideas.repec.org/p/fip/fedgif/1203.html. Board of Governors of the Federal Reserve System.

Cortes, C. and V. Vapnik. 1995. "Support-vector networks." *Machine Learning* 20, no. 3 (September): 273–297. ISSN: 1573-0565. https://doi.org/10.1007/BF00994018.

Dale, E. and J. Chall. 1948. *"A Formula for Predicting Readability".* Ohio State University Educational Research Bulletin 27 (January).

Deslongchamps, A. 2018. *"Readability and the Bank of Canada".* Bank of Canada Staff Analytical Note No. 2018-20. Available at https://www.bankofcanada.ca/2018/06/staff-analytical-note-2018-20/.

Explosion AI. 2018. *"spaCy: Industrial-Strength Natural Language Processing".* Available at https://spacy.io/.

Freedman, C. 1995. "The Role of Monetary Conditions and the Monetary Conditions Index in the Conduct of Policy [speech]." Available at https://ideas.repec.org/a/bca/bcarev/v1995y1995iautumn95p53-59.html, *Bank of Canada Review* 1995 (Autumn): 53–59.

Guénette, J.-D., N. Labelle, M. Leduc and L. Rennison. 2016. *"The Case of Serial Disappointment".* Bank of Canada Staff Analytical Note No. 2016-10.

Gunning, R. 1952. *"The Technique of Clear Writing".* New York: McGraw-Hill.

Hansen, S. and M. McMahon. 2015. *"Shocking Language: Understanding the Macroeconomic Effects of Central Bank Communication".* Discussion Papers 1537. Available at https://ideas.repec.org/p/cfm/wpaper/1537.html. Centre for Macroeconomics.

Hendry, S. 2012. *"Central Bank Communication or the Media's Interpretation: What Moves Markets?"* Bank of Canada Staff Working Paper No. 2012-9.

Hendry, S. and A. Madeley. 2010. *"Text Mining and the Information Content of Bank of Canada Communications".* Bank of Canada Staff Working Paper No. 2010-31.

Howard, J. et al. 2018. *fastai.* https://github.com/fastai/fastai.

Howard, J. and S. Ruder. 2018. "Universal Language Model Fine-tuning for Text Classification." In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Volume 1: Long Papers,* 328–339. Available at https://aclanthology.info/papers/P18-1031/p18-1031.

Kincaid, J. P., R. P. Fishburne, R. L. Rogers and B. S. Chissom. 1975. *"Derivation of New Readability Formulas for Navy Enlisted Personnel"*. Naval Technical Trainning Command Research Branch Report 8-75. Millington, Tennessee: Naval Air Station Memphis.

Kusner, M., Y. Sun, N. Kolkin and K. Weinberger. 2015. "From Word Embeddings to Document Distances." In *Proceeding of Machine Learning Research 37 (International Conference on Machine Learning): 957-966.*

Macklem, T. 2002. "Information and Analysis for Monetary Policy: Coming to a Decision." *Bank of Canada Review* (Summer): 11–18.

Martin, M. 2004. "The Bank of Canada's Business Outlook Survey." *Bank of Canada Review* (Spring): 3–18.

McLaughlin, G. H. 1969. *"SMOG Grading – a New Readability Formula"*. Journal of Reading (May): 639-646.

Merity, S., N. S. Keskar and R. Socher. 2017. "Regularizing and Optimizing LSTM Language Models." Available at http://arxiv.org/abs/1708.02182, *CoRR* abs/1708.02182.

Murray, J. 2013. "Monetary Policy Decision Making at the Bank of Canada." *Bank of Canada Review* (Autumn): 1–9.

Pennington, J., R. Socher and C. D. Manning. 2014. "GloVe: Global Vectors for Word Representation." In *Proceeding of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 1532–1543. Stroudsburg, Pennsylvania: The Association for Computational Linguistics. Available at http://www.aclweb.org/anthology/D14-1162.

Poloz, S. S. 2013. *"Returning to Natural Economic Growth"*. Speech at the Vancouver Board of Trade, Vancouver, British Columbia, September 18.

———. 2018. *"Let Me Be Clear: From Transparency to Trust and Understanding"*. Speech at the Greater Victoria Chamber of Commerce, Victoria, British Columbia, June 27.

Senter, R. J. and E. A. Smith. 1967. *"Automated Readability Index"*. Wright-Patterson Air Force Base.

Taboada, M., J. Brooke, M. Tofiloski, K. Voll and M. Stede. 2011. "Lexicon-Based Methods for Sentiment Analysis." Available at https://doi.org/10.1162/COLI_a_00049, *Computational Linguistics* 37 (2): 267–307.

VanOosten, P., D. Tanghe and V. Hoste. 2010. "Towards an Improved Methodology for Automated Readability Prediction." In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta.* Available at http://www.lrec-conf.org/proceedings/lrec2010/summaries/286.html.