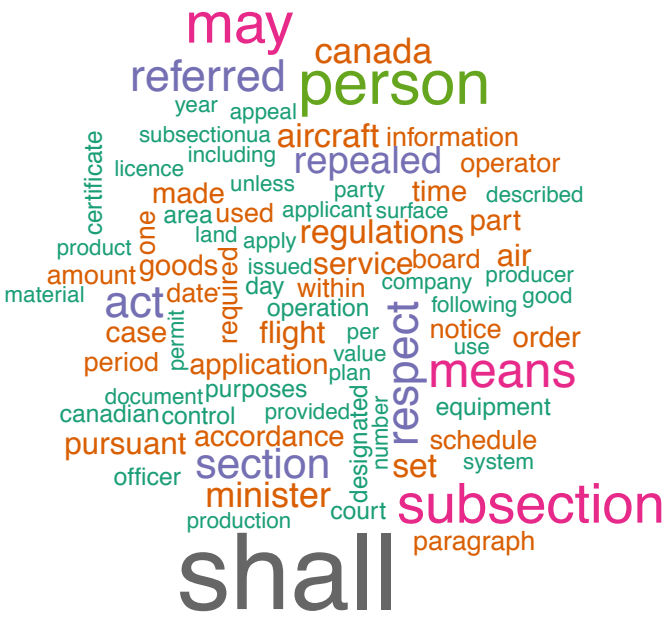
October 16, 2018
*University of Ottawa, Common Law Section*

# Semantic Analysis of Canadian Regulations

Report prepared for the Canada School of Public Service (CSPS)




Authors:

Professor Wolfgang Alschner
Professor John Mark Keyes
Alexander Geddes
Elina Korchagina
Marco Pontello
Rex Yeung
Peter C. Zachar

## Table of Contents

## Executive Summary

This project combines expertise in legislative drafting and legal data science to conduct an automated semantic analysis of more than 2700 Canadian federal regulations. It investigates four legal characteristics of regulations: **(1) Prescriptivity, (2) Flexibility, (3) Complexity, and (4) Age.** Drawing on the literature in relation to the assessment of regulatory burden, automated analysis of legal text, and plain language drafting, we devise a conceptualization and operationalization of these four characteristics. In order to scale the legal analysis of these characteristics, we rely primarily on an easily extendable and transparent dictionary approach that searches through the corpus of regulations for the presence of signaling terms and phrases. Our main insights and findings include:

**Prescriptivity:** How binding are regulations?

- We define prescriptivity as a *relative* measure of prescriptiveness ("shall", "must", ect.) in comparison to permissiveness ("may", "entitled to", ect.). Prescriptivity of Canadian regulations is increasing. Although prescriptive terms themselves are decreasing, this is offset by a larger drop in permissive terms rendering regulations more rigid.

- Prescriptivity tends to vary by regulatory subject matter. Consumer product regulations tend to be very prescriptive (see e.g. *Carriages and Strollers Regulations* scoring amongst the highest on our count) whereas other regulatory fields such as procedural court regulations tend to display low prescriptivity. Additional grouping will be necessary to determine whether, within a group of similar regulations, a document is *overly* prescriptive. *Overly* prescriptive regulations can signify either poor drafting (e.g. unnecessary repetition of prescriptions) or burdensome regulation.

**Flexibility:** How responsive are regulations to changing circumstances?

- We investigate three aspects of flexibility: (1) exceptions (on regulatees), (2) discretion (on regulators) and (3) incorporation by reference. We find that dimensions have remained constant over time, but vary strongly across regulations.

- Similarly to precriptiveness, flexibility varies systematically across the type of regulations. Whereas financial or transport regulations, for instance, have more opportunity to incorporate external industry standards, other types of regulation contain few such references.

**Complexity:** How easily understandable are regulations?

- Complexity, in our definition, captures how accessible the regulatory text is to its readers. This is achieved through plain, clear and well-structured drafting. Unfortunately, off-the-shelves readability measures (e.g. *Flesch Kincaid*) do not accurately capture the accessibility of regulatory texts since they regulations are differently formatted and structured as compared to the natural language texts for which these readability measures were developed.

- Using a word list approach instead, we find that regulations become more complex in their structure through more cross-references, but rely less on legal jargon.

**Age:** What is the average age of regulations?

- We calculate how much time has passed since a regulation was last amended, but also look for technology-related outdatedness specifically. We find that some regulations specify permitted means of communications, which can lead to the omission of newer forms of communciation such as EMail in older regulations that have not yet been recently amended.

Future work includes validating and refining our measures, unresolving outstanding issues such as regulations' readability, and deploying our measures to derive best practices within regulatory areas.

# 1. Project Scope and Overview

This project treats federal Canadian regulations as data in order to investigate their semantic content automatically. Specifically, this project empirically assesses four aspects of Canadian regulations by leveraging legal data science:

1. **Prescriptivity:** How binding are regulations?
2. **Flexibility:** How responsive are regulations to changing circumstances?
3. **Complexity:** How easily understandable are regulations?
4. **Age:** What is the average age of regulations?

In this report, we summarize our methodology and findings. An Excel dataset with index scores along the four categories for each regulation as well as an interactive website allowing users to interactively engage with the research findings accompanies this report.

The project work was conducted between May to October 2018 by researchers at the Common Law Section of the University of Ottawa for the Canada School of Public Service (GSPS). The group of researchers was led by Professor Wolfgang Alschner, an expert in legal data science, and Professor John Mark Keyes, a legal drafting expert and former Chief Legislative Counsel of the Canadian Department of Justice. Throughout that period, researchers were in close contact with GSPS receiving guidance and sharing preliminary results. The results provided in this report are subject to further validation and refinement through follow-up research.

This report is structured as follows. First, we identify the academic literature pertinent to this project. Second, we develop a conceptual framework building on that literature in order to clearly delineate the four characteristics of regulations we are interested in. Third, we introduce a methodology to empirically measure these characteristics using the regulations' text data provided to us by the GSPS. Fourth, we implement the analysis and describe our findings on the semantic content of the stock of federal Canadian regulations. Finally, we reflect on our results, including their limitations, and suggest avenues for future work.

## 2. Literature Review

Automatically investigating the semantic content of legal documents raises unique challenges for several reasons. First, legal documents have traditionally been analyzed by hand. While this allows analysts to deploy their legal training and engage in a context-specific interpretation of texts, this approach is not scalable and hence not suitable for investigating the corpus of Canadian regulations containing thousands of legal texts. The first challenge of the project thus consists of automating the process a lawyer or regulatory official would go through when analyzing these texts by hand. There is a growing literature that seeks to render legal analysis scalable through such automation.

Second, concepts such as prescriptivity, flexibility and complexity have no inherent meaning across fields and domains. Using the existing literature on these concepts in the legal domain, this project thus needs to clearly define them in order to make them measurable empirically.

Third, legal documents, especially those using formulaic, legislative language, are different from standard natural language texts. Whereas computational linguists have developed measures to assess the semantic characteristics of natural language texts, these tools cannot be blindly applied to legal documents. Using existing literature, we thus discuss the applicability of off-the-shelf analytical tools and potential adaptations and alternatives in the legal domain.

Below we discuss these three challenges in the context of the literature on (1) prescriptivity and flexibility as well as (2) linguistic complexity.

### 2.1. Prescriptivity and Flexibility

Prescriptivity and flexibility are two characteristics of a regulation that are difficult to quantify because a regulation's basic statistical details do not identify them as such. Therefore, an analysis of these attributes requires determining proxies that assist in quantifying the degree to which a regulation is prescriptive or flexible. The search for proxies has been ongoing in multiple academic disciplines. First, economists have been interested in this issue as they search for statistical information to support their critiques of extensive government regulation. Second, computer engineers have tried to locate proxies so they can facilitate regulatory compliance in regulation-heavy industries. Third, legislative drafting experts can provide insight into the prescriptivity and flexibility implications of words and phrases that they use in legal documents.

#### *Economics Measurement of Constraints in Regulations*

Economists have been at the forefront of the search for proxies that measure the extent of the obligations imposed by regulations. Some economists are concerned that state regulation imposes too many obligations and have negative impacts on the economy. Their research has sought out ways to quantify regulation so they can provide data to support their arguments.[1]

The primary way in which economists have measured obligations in regulations is to count the number and size of regulations. Although researchers agree that a regulation's

[1] Cary Coglianese has called for "empirical evidence for a retreat from rulemaking" (Cary Coglianese, "Empirical Analysis and Administrative Law" (2002) 1:4 U Ill L Rev 1111 at 1127-1128).

intrusiveness is countable, their methods of counting vary. Casey Mulligan and Andrei Shleifer use the size in kilobytes of digital versions of the USA's Federal Register.[2] Others count the number of pages in the Federal Register to determine the amount of regulation.[3] Clyde Wayne Crews Jr counts the number of "rules" created compared to the number of laws passed in a year, and he argues that it is more accurate than page counts.[4] Despite differences between the methods used for counting, these economists agree that the size and number of regulations is a accurate proxy for the amount of obligations imposed on industry. Their argument assumes that because one of the purposes of regulation is to impose obligations more regulations or longer regulations equal more obligations on industry.

In addition to page counting, economists at the Mercatus Centre have developed a new way of calculating the amount of obligations in a regulation. Led by Patrick McLaughlin, the Mercatus Centre has developed an open-source tool called 'RegData' that counts the number of "restrictions" in a regulation.[5] These researchers use computer programs to search regulatory text for five terms indicating restrictions: "shall", "must", "may not", "prohibited", and "required."[6] McLaughlin and his colleagues do not explain why they use these terms and not others, but they do discuss the advantages of this method over the page-counting approach.[7] 'RegData' has had success with its innovative approach as researchers within and outside Mercatus have referenced its data.[8] Reviews of 'RegData' have been generally positive, but doubts linger about its utility and critics highlight the fact that its data is merely a proxy for determining the impact of regulations.[9] Despite these concerns, 'RegData' is a popular tool for measuring the prescriptivity of regulations in the USA. It offers little insight into the flexibility of regulations because it assumes that restrictions are the foci of regulations. This model needs refining as the terms its measures are rudimentary, but it seems ripe for refinement and application in other jurisdictions.

---

[2] Casey Mulligan & Andrei Shleifer, "The Extent of the Market and the Supply of Regulation" (2005) 120:4 The Quarterly Journal of Economics 1445 at 1451-1452, 1456.

[3] Edward Stiglitz, "Unaccountable Midnight Rulemaking? A Normatively Informative Assessment" (2014) 17:1 NYU J Leg & Pub Pol'y 137 at 154; Bentley Coffey, Patrick McLaughlin & Robert Tollison, "Regulators and Redskins" (2012) 153:1-2 Public Choice 191 at 192; John Dawson and John Seater, "Federal Regulation and Aggregate Economic Growth" (2013) 18:1 J Econ Growth 137 at 140.

[4] Clyde Wayne Crews Jr, "Ten Thousand Commandments: Toward Accountability and Disclosure in the Regulatory State" (1998) 27:2 The Public Manager: The New Bureaucrat. Crews Jr is part of the Competitive Enterprise Institute that publishes an annual report on the USA's federal regulatory process.

[5] Omar Al-Ubaydli & Patrick McLaughlin, "RegData: A Numerical Database on Industry-Specific Regulations for all United States Industries and Federal Regulations, 1997-2012" (2017) 11:1 Regulations & Governance 109.

[6] Ibid at 112.

[7] The primary advantage is that its permits a greater level of "granularity" of a regulation (Ibid). However, researcher that use 'RegData' continue to use page numbers as well (McLaughlin, "The Consequences of Midnight Regulations and Other Surges in Regulatory Activity" (2011) 147:3-4 Public Choice 395 at 397-398).

[8] For research from within and outside the Mercatus Centre, see the papers listed at https://quantgov.org/research/.

[9] Christopher DeMuth Sr & Michael Greve, "Agency Finance in the Age of Executive Government" (2017) 24:1 Geo Mason L Rev 555 at 558; Michael Livermore, Vladimir Eidelman & Brian Grom, "Computationally Assisted Regulatory Participation" (2018) 93:1 Notre Dame L Rev 977 at 1013.

***Computer Tools for Identifying Rights and Obligations in Regulations***

Regulation is complex and industry actors often face a web of regulations from various levels of government. Computer scientists have taken up the task of extracting obligations, restrictions and rights in regulations to design systems that make regulatory compliance easier.[10]

Computer scientists use word-recognition techniques to extract details from regulations that impose obligations and create rights. Some, led by Travis Breaux, use the method of extracting rights and obligations from regulations and translating them into natural language statements.[11] Breaux searches regulations for constraints, obligations, refrainments, rights, and exceptions.[12] The phrases used as proxies in that search are the following: "does not have a right to", "has a right to", "is not required to", "may", "may deny", "may not", "may not require", "may require", "must", "must deny", "must permit", "must request", and "retains the right to."[13] Although the purpose of extracting this information is to develop tools for regulatory compliance, it can measure the prescriptivity and flexibility of a regulation because the details being extracted about rights and obligations correspond to prescriptivity and the information about exceptions corresponds to flexibility.[14] The Breaux approach is similar to 'RegData', but there are differences. First, Breaux casts a wide net by searching for more phrases. Second, Breaux identifies rights, obligations and exceptions as well as anti-rights and anti-obligations.

The work of Breaux has inspired other computer scientists to refine the tools used for extracting rights, obligations, and exceptions from regulatory texts. The most contemporary tool is 'GaiusT', which is a tool that extracts rights and obligations from legal documents using semantic annotation.[15] The indicators of prescriptiveness and flexibility are the following:

- *Right*: may, can, could, permit to have a right to, should be able to.
- *Anti-Right*: does not have a right to.
- *Obligation*: must, requires, should, will, would, which is charged with, may not, can not, must not.

---

[10] Ira Rubinstein identifies the importance of this task as she believes computer scientists are ideally positioned to bridge the gap between the rules regulators are creating and the need for companies to improve their compliance practices (Ira Rubinstein, "Regulating Privacy by Design" (2011) 26:1 Berkeley L J 1409 at 1425-1426).

[11] Travis Breaux, Matthew Vail & Annie Anton, "Towards Regulatory Compliance: Extracting Rights and Obligations to Align Requirements with Regulations"(2006) Requirements Engineering, 14th IEEE International Conference 49; Breaux & David Gordon, "Regulatory Requirements Traceability and Analysis Using Semi-formal Specifications" in Joerg Doerr et al, eds, Requirements Engineering: Foundation for Software Quality 19th International Working Conference (e-book: Springer, 2013) 141; Breaux & AI Anton, "Analyzing Regulatory Rules for Privacy and Security Requirements" (2008) 34:1 IEEE Transactions on Software Engineering 5.

[12] Breaux & Anton, supra note XX at 7.

[13] Breaux, Vail & Anton, supra note XX.

[14] Another aspect of the computer engineering research that is informative for measuring prescriptivity and flexibility is the analysis of the legislative drafting technique of incorporating by reference. The solutions developed by researchers for extracting information incorporated in this way could be of use in assessing prescriptivity and flexibility (see Daniel Martin Katz & M J Bommarito II, "Measuring the Complexity of the Law: the United States Code" (2014) 22:1 Artificial Intelligence & L 337 at 360).

[15] Nicola Zeni et al, "GaiusT: Supporting the Extraction of Rights and Obligations for Regulatory Compliance" (2015) 20:1 Requirements Engineering 1.

- *Anti-Obligation*: is not required, does not restrict, does not require.
- *Exception*: only, but, except, without, limit, restrict, exclusion, other than, unless, release, not include, deny, eliminate, limited to, prevent, exempt, not apply to, to object, refuse, not in, objected, objection, outside of, omit, reject, omission, prelude, rule out, discharge, save, bar, relieve, exclude from, besides, not valid, aside from, leave out, restrict to, not listed, exonerate.[16]

The 'GaiusT' tool is the culmination of research into refining Breaux's techniques.[17] The tool searches regulations for these terms and creates compliance systems using them. It also counts the number of times those terms appear in a regulation, and it creates possibilities for using that data to examine prescriptivity and flexibility.

***Legislative Drafting Terminology***

While economists and computer scientists have tools for extracting information from regulatory documents, legislative drafters select the terminology used in legislative and regulatory texts. The analysis of regulatory language should be informed by the use of this terminology.

The literature on legislative drafting comments on three aspects that assist in analyzing drafting terminology.

First, drafters explain the kinds of language used to create the key elements of regulatory texts: obligations, prohibitions, and rights. In the drafting of federal legislation, the most common way to create obligations is to use the terms: "shall", "must", "is + [verb infinitive]" and "is required to". Prohibition are generally created by using these terms with negative modifiers ("no"' or "not"). The most common way to create rights (and consequently less prescriptivity) is to use "may", "is entitled to", and "has the right to". The literature also highlights poor drafting practices, like using the phrase "no person may" to create prohibitions.[18]

Second, legislative drafting literature explains how drafters can create exceptions and carve-outs. These indicators of flexibility usually drafted using terms like "except", "unless", "provided", "if …, then …", and "subject to".[19]

Finally, the literature discusses the use of abstract or subjective qualifiers and adjectives. It cautions against using subjective qualifiers and adjectives (like "reasonable") because they create uncertainty, but these terms nevertheless have a role to create room for flexible interpretation.[20]

---

[16] Ibid at 6.

[17] Nicola Zeni, Luisa Mich & John Mylopoulos, "GaiusT 2.0: Evolution of a Framework for Annotating Legal Documents" in Emmanouel Garoufallou et al, eds, Metadata and Semantics Research (e-book: Springer, 2016) 43.

[18] Paul Salembier, Legal and Legislative Drafting (Markham: LexisNexis, 2009) at 34-38. However, drafters also note that the words "may" and "must" can be used in confusing ways (e.g. "may not" creates an obligation not a right) (Salembier, ibid at 160; GC Thornton, Legislative Drafting, 4th Ed (London: Butterworths, 1996) at 103; Louis-Philippe Pigeon, Drafting and Interpreting Legislation (Agincourt: Carswell, 1988) at 35.

[19] Salembier, supra note XX at 40-52, 261-262; Thornton, supra note XX at 101-102.

[20] Salembier, supra note XX at 40-52, 261-262; Thornton, supra note XX at 101-102.

### 2.2. Linguistic Complexity

Legislation is most useful when it is accessible, relevant and applicable. Efforts to make legislative documents more accessible provide helpful clues to understand some of the factors contributing to complexity in legislation. These efforts are most evident in the research on *Plain Language* and legislative drafting guides that incorporate Plain Language approaches in Commonwealth countries such as Canada.[21] This research and drafting guidance provides a structural framework to assess legislative texts in terms of complexity and the reduction of complexity by improving readability. Plain language approaches are not a panacea for reducing complexity,[22] but they have been adopted in legislative drafting in Canada[23] and other Commonwealth countries.[24]

The factors that define complexity in legislative documents are not definitive or clear-cut, but rather exist on a dynamic spectrum of linguistic elements. This makes measuring and assessing complexity in legislation challenging. Complexity factors range from the stylistic structure of regulations (including paragraphing and numbering) to the use of legal jargon. The goal of reducing complexity is to help the reader understand the requirements, prohibitions and rights created by the legislation. Therefore, one aspect of reducing complexity involves drafting with the reader in mind to improve the accessibility of the law.

The following is an overview of the linguistic factors that might serve as indicators of complexity and are considered most frequently among scholars and legislative drafters. It is important to note that some of them overlap and may be combined with each other; they are not separate concepts, but rather guides to assessing linguistic complexity.

***Major linguistic structures***

The *Guide to Legislation and Legislative Process* from the province of British Columbia provides a structure to examine how plain language techniques are being implemented by legislative drafting guides in many Commonwealth countries.[25] This guide sets out the challenges and constraints of clear drafting. The following are some examples of linguistic structures it considers:

- *Words*: technical terms will only be used if necessary for precision or if they are familiar to the general public,
- *Sentence Length*: shorter sentences preferred, and clauses used sparingly,

---

[21] See, for example, *Legistics*, published by the Department of Justice (Canada), Principles of Clear Drafting, published by New Zealand, Parliamentary Counsel Office, Ministry of Justice and *Guide to Legislation and Legislative Process in British Columbia- Part 2: Principles of Legislative Drafting*, published by the Office of Legislative Counsel, British Columbia: 2013.

[22] See Ruth Sullivan, "The Promise of Plain Language Drafting" (2001) 47 McGill LJ 97 for a critical assessment of plain language approaches and their limitations.

[23] See the Uniform Acts Drafting Conventions of the Uniform Law Conference of Canada, notably s. 2 which says: "2. An Act should be written simply, clearly and concisely, with the required degree of precision, and as much as possible in ordinary language.".

[24] Australian Office of Parliamentary Counsel Plain Language webpage and United Kingdom, Office of the Parliamentary Counsel, "When Laws Become Too Complex", Cabinet Office (16 April 2013).

[25] Supra n. 21.

- *Structure*: condition and qualification should not be embedded in the middle of a sentence and positive sentences rather than the less understandable negative form.[26]

***Unnecessary words***

In her work on analyzing the ambiguity in legislative drafting, Ann Nowak suggests a list of "unnecessary words" for drafters to eliminate in their writing.[27] This list may prove to be a useful building in creating a vocabulary of measurable attributes that lead to complexity in legislation. Nowak suggests that using unnecessary extra words bloats writing and detracts from its power.[28] She suggest the following list of unnecessary words:

- due to the fact that (because),
- in spite of the fact that (although),
- in the event that (if),
- in the course of (during),
- in numerous instances (often),
- for the purpose of (for).

Nowak is not the only source that provides words that are presumed to cause ambiguity and complexity in legislation. Drafting guidelines often provide some examples of terms to be avoided. For example, the Department of Justice Canada *Legistics* provides a helpful list of terms that are most likely to cause unnecessarily complex legislation. These are some examples that would be least ambiguous based on their context:

- notwithstanding,
- pursuant to,
- hereby.[29]

***Indefinite adjectives***

Maurizio Gotti suggests that some of the more measurable factors that cause ambiguity in legislation are indefinite adjectives such as "reasonable", "substantial", "satisfactory", "negligent", and "unconscionable".[30] These terms confer wide discretion on judges. As a result, what is "unjust" or "negligent" in a specific circumstance is difficult to anticipate *ex ante*, which, in turn lowers the predictability of a regulation and thereby increases its complexity.

---

[26] *Guide to Fostering Readability in Legislative Texts*, Department of Justice (Canada), *Guide to Legislation and Legislative Process in British Columbia- Part 2: Principles of Legislative Drafting*, published by the Office of Legislative Counsel, British Columbia: 2013 at 7. See also Nick Horn, "Legislative Section Headings: Drafting Techniques, Plain Language, and Redundancy" (2011) 32:3 Statute Law Rev 186 and New Zealand, Parliamentary Counsel Office, Principles of Clear Drafting.

[27] Ann L Nowak, "Demystifying Ambiguity in Legislative Writing" (2016) 37:2 *Statute Law Review* at 170.

[28] Ibid. at 170.

[29] Canada, Department of Justice, *Legistics*.

[30] Maurizio Gotti, " Linguistic Insights into Legislative Drafting" (2014) 2:2 The Theory and Practice of Legislation 123 at 132.

## 3. Conceptual Framework

Building on the above literature, we develop a conceptual framework in this section to delineate the four characteristics of regulations we are tasked to assesses. This step is crucial in order to clarify what we aspect of a regulation we investigate (and what we do not) and to prepare the subsequent empirical measuring of these concepts.

### 3.1. Prescriptivity

#### *Comparing Prescriptions and Permissions*

Prescriptivity is an essential characteristic of regulations because the purpose of regulations is primarily to create rules for legal subjects. Those rules are usually in the form of obligations, prohibitions, or restrictions. The prescriptivity of a regulation is the degree to which it imposes obligations, prohibitions, or restrictions on the legal subjects it targets. The most common words used to create obligations in Canadian regulations are "shall" and "must", but there are other words and phrases that signal prescriptivity in a regulation. Thus, we start this analysis of prescriptivity by searching a regulation for words and phrases that signal 'prescriptions' because they indicate the number of obligations, prohibitions, and restrictions in a regulation.

All regulations are prescriptive. However, the degree to which they are prescriptive lends itself to analysis because a regulation can have variable levels of prescriptivity.

Counting obligations, prohibitions, and restrictions alone does not provide significant information about prescriptivity. Instead, determining prescriptivity requires a comparison of the number of obligations, prohibitions, or restrictions in a regulation with the number of permissions in that regulation. Permissions are antithetical to prescriptions. These are words or phrases that permit legal actors to take certain actions or create rights for those actors. The most common word used to create a permission in Canadian regulations is "may", but other words and phrases can create permissions. So, the second step in our prescriptivity analysis is to search a regulation for words and phrases that signal 'permissions'.

A regulation with many mandatory requirements and few permissive features likely has high prescriptivity, while a regulation with few mandatory requirements and many permissive features likely has low prescriptivity. Thus, in our definition the Prescriptivity of a regulation is not tied to its length or the number of 'rules' in it (as the scholars discussed in the literature review suggest). Instead, it is a characteristic that exists on a spectrum from high to low based on the balance between mandatory and permissive elements in the regulation. Differently put, (in contrast to the Mercatus group) we conceive of prescrivity as a *relative measure*. Therefore, the final step in our analysis of prescriptivity is to compare the number of prescriptions to the number of permissions in a regulation and use that comparison to assess a regulation's level of prescriptivity.

#### *Measuring Prescriptivity Poses Conceptual Challenges*

There are potential issues with counting and comparing the number of prescriptions and permissions to measure prescriptivity. The most significant issue is that this method does not distinguish between regulators and regulatees. Regulations target many legal subjects, but those

subjects usually fall into two categories: the entities (like citizens or businesses) who conduct the regulated activity (regulatees) and the public actors and agencies responsible for regulating the activity (regulators). Determining a regulation's degree of prescriptivity depends on whether the focus is on the regulator or the regulatee. Take, for example, a regulation that says "[regulatee] must ensure its products are safe" and also says "[regulator] may create guidelines for determining the safety of a product." This regulation is highly prescriptive for the regulatee but it also creates a permission for the regulator. Our approach does a holistic assessment that does not distinguish between regulators and regulatees; however, future research could use our methods and analyze the prescriptivity of a regulation for each distinct category of legal subject.

A second issue springs from the distinction between obligations or restrictions and rights or permissions. While these appear to be indicators of high or low prescriptivity, the choice by the regulatory drafter to use one or the other may be unprincipled. The following example illustrates this problem. The expression "a person may grow up to three cannabis plants" creates a permission, whereas the expression "a person shall not grow more than three cannabis plants" appears to be a restriction. These expressions appear have the same effect, but one creates that effect using a permission while the other does so by using a prescription.[31] This issue highlights why previous approaches that only counted the number of times "shall" and "may" appeared in a regulation are inadequate. Our approach reduces the impact of this issue by searching for more words and phrases that are indicative of prescriptions and permissions. The more exhaustive the lists of words and phrases are for these two characteristics, the greater the likelihood that we can mitigate the effects of divergent drafting practices.

### 3.2. Flexibility

#### *Measuring Flexibility*

A regulation's flexibility is the extent to which it can accommodate and respond to varying circumstances within its subject-matter. Regulations create rules that legal subjects must follow, and those legal subjects often raise concerns about the rigidity of those rules. A one-size-fits-all approach to regulation can fail to account for the nuances of the regulated activities and actors. In response to this issue, drafters of regulations often strive to strike a balance between the need for certainty through clear and universal rules and the desire for flexibility that accommodates the complexities within a regulated space. Like prescriptivity, flexibility is not a binary characteristic. Instead, a regulation can have a low, medium, or high degree of flexibility.

A regulation's flexibility, however, is also multi-faceted. We thus identify three categories of flexibilities. First, *exceptions* understood as deviations from otherwise prescribed conduct can generate flexibility. Second, a regulation may confer *discretion* and may thereby induce flexibility in implementation. Third, a regulation may not regulate conduct directly but *incorporate by reference external standards* which can be flexibly updated over time. Each of

[31] Although the examples provided appear to have the same effect, there is an essential difference between them. The permission version requires a prior prohibition against growing cannabis plants while the restriction version could stand alone. Thus, the permission version would be accompanied by a prohibition and the outcome would be a neutral prescriptivity value. However, the restriction version would need no other provision and would lead to a higher prescriptivity value.

these aspects constitutes a distinct dimension of flexibility. We therefore decided not to aggregate them into a single measure.

Exceptions are carve-outs or exemptions from general rules. For example, a general rule may prohibit the possession of a firearm on a construction site, but there may be a carve-out permitting the possession of a firearm provided the legal subject has written approval from the relevant authority. Exceptions signify that the drafters of a regulation have turned their minds to the fact that the regulatory rules must be adapted to different situations.

Discretions are expressions that use subjective language and permit variability (like requiring a regulatee to take "reasonable" actions).[32] This variability highlights the drafters' acknowledgement that they must leave space for the legal actors involved to determine an appropriate standard.

Finally, incorporation by reference ties a prescribed conduct in a regulation to an external standard. This may be an international standard created by the International Standardization Organization or a domestic industry standard. Incorporating these standards by reference makes a regulation more flexible as these standards are updated on a regular basis to reflect evolving safety or industry norms without thereby creating a need to update a regulation by itself. Furthermore, such incorporation of industry standards creates flexibility insofar it allows the industries to self-regulate by incorporating their standards by reference.

Once we have determined the number of exceptions, discretions, and incorporation by reference, we can compare those details to determine a regulation's flexibility. The greater the number of exceptions, discretions, and incorporation by reference in a regulation, the more flexible the regulation is. The fewer the exceptions, discretions, and incorporations by reference, the less flexible the regulation is.

***Conceptual Issues***

The characteristic of flexibility poses conceptual issues. As with prescriptivity, one of them involves taking account of whether a provision focuses on the regulator or the regulatee. If a regulation includes flexibility for one of these actors, that does not necessarily lead to flexibility for the other. For example, a regulation may require the regulator to create rules of procedure for administrative purposes while at the same time imposing a strict requirement that the regulatee must follow those rules of procedure. In this case, the regulator benefits from significant flexibility to create rules as it sees fit, but the regulatee faces a rigid requirement to comply with the rules. As with prescriptivity, our analysis is a holistic assessment that does not distinguish between regulators and regulatees; however, future research could use our methods and analyze the flexibility of a regulation from the perspectives of different legal subjects.

[32] In the context of this analysis, we do not use the term 'subjective' its legal sense (as in the difference between an objective 'reasonable person' standard in law versus a subjective standard based on a person's perspective). Instead, we use "subjective" and 'objective' in their ordinary senses. In their ordinary senses, "subjective' language is open to interpretation (like 'reasonable distance') whereas 'objective' language is quantifiable and measurable (like 'two metres').

Another conceptual issue with flexibility relates to the use of subjective language as a signifier of flexibility.[33] Regulatory drafters can use discretionary language to permit the legal actors involved (or an adjudicative body like a court) to determine an appropriate standard. However, the discretionary terms they use often acquire fixed and rigid meanings over time. For example, a regulation may require that a regulator provide a regulatee with "reasonable notice" before conducting an investigation; over time, customary practice or judicial decisions may determine that "reasonable notice" requires at least 48 hours notice. Thus, a seemingly-flexible standard can become inflexible over time. However, that standard can also change in the future and it is open to legal actors (or judges) to adopt a new understanding of the meaning of subjective language. Therefore, discretionary language remains an indicator of flexibility.

Finally, flexibility has a complex relationship with other characteristics of regulations. First, prescriptivity and flexibility have a close relationship with one another. It is possible to view them as two ends of a spectrum; but it is also possible to view them as independent characteristics. The fact that permissions appear in our analysis of both these characteristics highlights their relationship. We separate these characteristics because separating them provides more information about a regulation while leaving the door open to analyzing these characteristics together.[34] Second, the characteristics of flexibility and complexity are often in competition with one another. A regulation that imposes uniform standards is likely be less complex as it avoids confusing readers with exceptions or variable standards. However, those same exceptions or variable standards lead to a more flexible regulation. These similarities and differences between flexibility and other characteristics highlight the value of measuring them independent of one another.

### 3.3. Complexity

Reducing complexity is one of the core considerations for legislative drafters. Efforts to make legislative documents more accessible provide helpful clues to understand the factors that define complexity in legislation. The factors that define complexity in legislative documents are not definitive or clear-cut, but rather exist on a dynamic spectrum of elements. This is what makes measuring and assessing complexity in legislation so challenging. Complexity factors could be anything from the stylistic structure of regulations, including paragraphs, lettering and numbering, to excessive legalese jargon. The goal of reducing complexity is to ensure that the end user understands the rights and impacts enforced by the legislation. Therefore, reducing complexity could mean writing with the reader in mind to improve clarity and accessibility of the law. This is an overview of the linguistic factors that serve as potential indicators of complexity and are considered most frequently among scholars and legislative drafters. These factors will serve as the conceptual subcomponents of complexity in legislation. They will often overlap and

[33] In the context of this analysis, "subjective" is not used in its legal sense (as in the difference between an objective 'reasonable person' standard versus a 'subjective' standard based on a person's perspective). Instead, we use "subjective" in its ordinary sense. In their ordinary senses, "subjective" language is open to interpretation (like 'reasonable distance') whereas "objective" language is quantifiable and measurable (like 'two metres').

[34] For example, it may be desirable in certain contexts to have a regulation that is highly prescriptive but also highly flexible (like an occupational health and safety regulation that applies to many different industries). Providing separate scores for prescriptivity and flexibility would permit a drafter of such regulations to target changes to each characteristic.

be combined with one another, and are not separate concepts, but rather a guideline of non-exhaustive elements of complexity.

It is also important to note that we define complexity in *formal rather substantive* terms. Some regulatory matters are more complex than others and may involve more or less sophisticated compliance mechanisms. For the purpose of this analysis, we disregard such consideration of substantive complexity and instead focus on formal complexity, i.e. the accessibility of a text for its readers. From a normative point of view, such formal complexity should be low regardless of the substantive complexity of the subject matter to make a text accessible.

### *Potential Indicators of Formal Complexity*

#### *Clarity*

Clarity is perhaps the broadest and most important indicator. The concept of clarity touches on several overarching questions that affect complexity: (1) does the law achieve its intended purpose? (2) Are the provisions unambiguous? (3) Are the provisions relevant for the intended audience? If, when looking at any given legislation, the answer to any of these questions is no, this may indicate high complexity. The concept of clarity is also the foundation of plain language drafting, one of the main efforts to tackle complexity in legislation. However, as scholars behind plain language drafting indicate, many components of legislation can affect clarity,[35] therefore it is best assessed when taken into consideration as a theoretical factor.

#### *Readability*

Readability is another broad indicator that closely affects the complexity of legislation. The readability of a text is generally enhanced by writing for the reader, organizing the information well and making sure that the ideas and sentences have a logical flow. The structural format of the text as well as making its intended purpose clear are other important components of readability. The organization of the document is perhaps one of the most important elements of readability. In order to reduce complexity, the writer should structure the document in such a way that every part of the text conveys a useful element of the legislation that flows logically. The writer must employ well-chosen headings, points of reference and clear language to improve the overall readability of the text.

#### *Excessive Legalese*

Legalese is generally recognized as a barrier to comprehension by readers who do not have legal training. Modern legislative drafters have made significant efforts to avoid unnecessary, complex legalese. Most legislative drafting guidelines across Commonwealth countries include a section that reminds drafters to avoid the use of unnecessary Latin phrases and complicated expressions and words.[36] It is logical to conclude that any legislation that uses

---

[35] Noor A Hashim, "Plain language: give it a try!"(2013) 39:3 Commonwealth Law Bulletin 423, Ruth Sullivan, "The Promise of Plain Language Drafting" (2001) 47 McGill LJ 97, see also Alain Songa Gashabizi, "In pursuit of clarity: how far should the drafter go?" (2013) 39:3 Commonwealth Law Bulletin 415-432.

[36] Canada, Department of Justice, *Legistics* (http://www.justice.gc.ca/eng/rp-pr/csj-sjc/legis-redact/legistics/p1p18.html), New Zealand, Parliamentary Counsel Office, Principles of Clear Drafting, BC Office of Legislative Counsel, *Guide to Legislation and Legislative Process in British Columbia- Part 2: Principles of Legislative Drafting*.

many of these terms is unnecessarily complex. This is one of the more easily measurable elements of legislation and does not depend as heavily on context as some of the other factors.

*Sentence Structure*

The structure of sentences, or syntax, is one of the most important factors affecting the readability of a sentence. A clear and logical relationship between the words in a sentence helps readers understand how they are connected and often also helps them understand unfamiliar words or terms. Some ways to assess the complexity of sentences are its length and its general structure. Many scholars suggest that one of the best ways to improve legislation is to reduce sentence length and add more sections and subsections.[37] However, it would be a mistake to presume all long sentences to be complex. A long sentence that makes logical links between ideas and avoids unnecessary repetition may often be less complex for the reader than several shorter sentences that convey the same meaning. A more useful measure of complexity is the use of syntax, however this is more challenging to measure, as syntax may change based on context, audience or purpose.

*Format of the Document*

A clear and well-structured format facilitates understanding complex legislation. The format can include many elements such as the efficient use of titles and subtitles, consistency of these titles, well-spaced paragraphs and clear numbering and lettering. Many of these elements are closely tied to the general readability of a document. If the format is clear and consistent, there is a logical flow for the reader to follow and the document becomes less complex. The reader must be able to know exactly where they are within the document and be able to follow along with the structure rather than be distracted by the layout. A useful element that improves an otherwise complex document is a properly used definition section. This reduces repetition, promotes consistency of meaning and reduces ambiguity.

*Length and Volume*

Lengthy legislation is intimidating for most readers and is complex because of its sheer volume. Shorter legislation is less intimidating and more accessible. However, this is not necessarily a strong indicator of complexity. Modern drafting styles promote more white space and paragraphs, therefore a long piece of legislation that is well formatted and clearly spaced, could be less complex than a short document with poor formatting. Shorter documents may also be indicative of high complexity as they do not include the necessary detail required to make the provisions of the legislation clear and unambiguous.

### *How Indicators of Complexity Can Be Measured*

*Flesch-Kincaid readability tests*

The Flesch-Kincaid readability scores are some of the most widely used measures of readability and may provide a useful foundation in assessing the complexity of legislation. The Flesch-Kincaid uses the core measures of word and sentence length to indicate the level of reading difficulty. Although this test is used extensively in many fields, it may not be adequate as a standalone, comprehensive measure of complexity in legislation. Legislation is typically

[37] Elmer Driedger, *A Manual of Instructions for Legislative and Legal Writing* (Ottawa: Justice Canada 1982) at 556-557.

formatted in hierarchically arranged units of text (paragraphing). Although they constitute very long sentences, they are much easier to read because they expose the syntactic structure of the sentence. Traditional readability formulas do not take account of this formatting feature and result in deceptively poor readability scores. These tests also take no account of other factors that contribute to complexity such as the more subjective elements of discourse processing.[38]

*Generating lists of "unnecessary words"*

In her work on analyzing the ambiguity in legislative drafting, Ann Nowak suggests a list of "unnecessary words" for drafters to eliminate in their writing.[39] This list may prove to be a useful building in creating a vocabulary of measurable attributes that lead to complexity in legislation. Nowak suggests that using unnecessary extra words bloats writing and detracts from its power.[40] Nowak is not the only source of words presumed to cause ambiguity and complexity in legislation. Drafting guidelines also provide examples of terms to be avoided. The Department of Justice Canada *Legistics*, for example, provides a helpful list of terms that unnecessarily complicate legislation. It will be necessary to start with words that will almost always cause ambiguity or complexity rather than those that depend on linguistic context.

*Identifying parts of speech that cause ambiguity*

In his work on legislative drafting, Maurizio Gotti suggests indefinite adjectives that contribute to complexity, such as 'reasonable', 'substantial', 'satisfactory', 'negligent', and 'unconscionable'.[41]They increase complexity because of their 'borderline indefiniteness'.[42] Identifying these words would be another way of assessing unnecessary ambiguity. However, one challenge is that some parts of speech may generate false positives and be limited to certain contexts.

*Identifying complex sentence structures*

There are many different ways to approach measuring sentence structures. It will be particularly difficult to identify which structures cause complexity regardless of the context. Many legislative drafting guidelines suggest limiting the use of clauses, however this can be challenging as their degree of complexity may depend on the context and therefore not be easily measurable. Another way of measuring sentences could be based on readability formulas such as Flesch-Kincaid, however it is the interrelationship of words in sentences rather than their length that more often causes complexity. One possible measure of complexity could be the distance between a subjects and verbs in a sentence. The greater the distance, the more the reader must work to make the connection. However, splitting subjects and verbs is often done in regulations to achieve greater precision and avoid ambiguity. Reducing these splits must not be done at the expense of creating ambiguity. Sentence structure is critical to both readability and precision and merits further analytical research.

---

[38] Scott A Crossley, David B Allen & Danielle S McNamara, "Text readability and intuitive simplification: A comparison of readability formulas" (2011) 23:1 Reading in a Foreign Language at 84.

[39] Ann L Nowak, "Demystifying Ambiguity in Legislative Writing" (2016) 37:2 Statute Law Review at 170.

[40] Ibid at 170.

[41] Maurizio Gotti, " Linguistic Insights into Legislative Drafting" (2014) 2:2 The Theory and Practice of Legislation 123 at 132.

[42] *Ibid.*

### 3.4. Age

Of all our concepts, the age of a regulation appears to be most straightforward. Age indicates how up-to-date a regulation is. A regulation may have been enacted several decades ago, but if it was subsequently amended to reflect changes in society, its text has been brought up to date. We thus use the last amended date of a regulation, or alternatively, its date of original registration to calculate the age of a given regulation.

A second more substantive approach to measuring whether a regulation is up to date is to look for specific instances of outdatedness. While outdatedness may have different aspects, the CSPS suggested that this research should focus on *technology-related outdatedness*. The idea is that regulations should not make conduct, processes, and procedures contingent on certain technology, but rather remain technology-neutral thereby making them "future proof", i.e. able to accommodate technological advancements. One area where regulations may not be technology-neutral is in mandating the use of certain communication technology. A regulation enacted in the pre-Internet era may, for instance, list letter or fax as permissible means of communication, but not Email.

## 4. Methodology

In this section, we present the data used, describe the general approach taken and explain in greater detail how we operationalized the measurement of the four concepts.

### 4.1. Data Description

The project used the data provided by Justice Canada on the website ftp://205.193.86.89/ in the folder "Consolidation_Regs_1.2.0". Only English language regulations were considered in this iteration of the project.

The folder currently contains three sets of documents only one of which was excluded from the analysis after consultations with the CSPS:

1. Statutory Orders and Regulations ("SOR") – included
2. Consolidated Regulations of Canada ("CRC") – included
3. Statutory Instruments ("SI") – excluded

Each regulation was stored in an XML format, which contained (1) meta-data describing each regulation (such as its title, year of registration, or enabling statute) and (2) structured full text data, which further distinguished between a) the main body of a regulation, b) an order contained in a regulation or c) a schedule annexed to the regulation. Only a subset of the regulations contained order and schedule elements, which furthermore differed in style and content from ordinary regulations' main texts. Schedules, for instance, can contain tables, forms to be filled out for applicants as well as further specifications. To ensure that our measures were comparable and interpretable, we only used the meta-data and main body full text data in our analysis.

We also considered using the point in time xml (PITXML) data provided by the CSPS for our analysis. That data contained different versions of the same regulation if an amendment had

occurred and was stored in nested folders, one for each regulation. In the end, however, our approach did not require that data.

To access the meta and full text data of Canadian regulations, we scraped the data and stored it as a comma-separated values (csv) file. We used the programming language *R* for all scraping and analysis. Using a series of formulas and functions in R, we extracted the meta-data from the data's XML tags into a master data set. In addition to the full text, the resulting master data set included the following meta-data elements:

- Unique ID
- Short Title
- Long Title
- Registration Year
- Consolidation Year
- Last Modified Date
- Last Amended Date
- Enabling Authority
- Repealed (yes or no)

### 4.2. General approach: Dictionary of signaling words

In order to measure our four concepts, we considered two potential solutions. The first consisted of an inductive, machine learning approach where legal drafting experts would mark-up a subset of regulations and identify text specific passages that correspond to prescriptivity, flexibility, complexity or age. A supervised machine learning would then be trained on these manually annotated texts to learn how to label the remaining unlabeled documents.

A second approach, which we ultimately favored, consisted of deductively identifying terms that signal the presence of the characteristics we were interested in. Following that approach, we created dedicated word lists of prescriptive, permissive, complex or outdated terms. Each of these word lists constitutes a "dictionary" of signaling words for a specific characteristic. We would then calculate how often these signaling words appear in a document. For instance, a regulation that contains many outdated words tends to be more outdated than a regulation with few outdated words. To account for varying length, we normalized these words counts per 100 words.

We ultimately preferred the second approach for three reasons. First, it is intuitive to understand and fully transparent. We are essentially counting words with the help of a computer. Second, our dictionaries can be easily updated if a word needs to be added or eliminated making the approach flexible. A machine-learning approach in contrast would require new manually annotation of documents. Third, while expert annotation would introduce a level of discretion as to what different experts understand as "complexity", signaling words impose a clarity of concept and focus the debate on whether a certain term operates as proxy for the underlying proxy.

At the same time, we acknowledge that the approach also has shortcomings. First, some concepts may be more easily associated with signaling words than others. More interpretive, context-specific semantic information cannot be easily captured through such a dictionary

approach. Second, words can have different meaning in different contexts. We try to mitigate the risk of false positives, i.e. instances where our signaling words do not in fact signal the presence of the underlying concept, by choosing unambiguous words and embedding them in regular expressions. Our approach, given its simplicity, is thus only an imperfect, albeit useful means to get at the underlying semantic concepts.

### 4.3. Prescriptivity

Our analysis of Prescriptivity relied on generating counts of prescriptive and permissive words, normalized as counts of words of interest per 100 words of regulation text. These indices were then used to calculate a Prescriptivity Score based on the formula:

(Prescriptive Words) / (Permissive Words)

This formula was designed to reflect the moderating influence of permissive words on the rules imposed by prescriptive language. A higher score indicates a more prescriptive regulation.

The words used to generate a normalized count of *prescriptive words* in each regulation were:

- must
- shall
- required
- prohibited
- is to be
- cannot be
- may not
- may only

It was necessary to distinguish instances where the words "required" and "prohibited" appeared on their own, rather than being preceded by "not". This was achieved using a negative look-behind function, counting only instances where the word "not" did not appear immediately before "required" or "prohibited", respectively.

The words used to generate a normalized count of *permissive words* in each regulation were:

- may
- not required
- not be required
- not prohibited
- entitled to
- has the right to

It was again necessary to distinguish instances where the word "may" appeared on its own from those where it was followed by "not" or "only" – indicating a prescription rather than a permission. This was achieved using a positive look-ahead function, counting only instances where the word "may" was not immediately followed by "not" or "only".

### 4.4. Flexibility

Flexibility was similarly analyzed by first generating counts of words indicating exceptions and discretions, normalized as counts of words of interest per 100 words of regulation text.

The words used to generate a normalized count of words indicating *exceptions* were:

- subject to
- unless
- except
- provided that
- in a case where
- do not apply
- does not apply

The words used to generate a normalized count of words indicating *discretion* were:

- includes
- through no fault
- beyond the control
- if necessary
- possible
- practicable
- appropriate
- frequently
- regular
- reasonabl
- take into account
- taking into account
- without delay
- at the first opportunity

The truncated form “reasonabl” was used to capture instances of both “reasonable” and “reasonably”.

Finally, to identify incorporation by reference of external standards, we did not need to look for specific signaling words. Instead, we could use existing meta-data information embedded in the XML of each regulation to calculate how many times a regulation refers to an external “standard” or an “other” standard.

### 4.5. Complexity

Complexity was more difficult to operationalize. We tried a set of off-the-shelf readability measures, but results were not useful. We thus resorted to a dictionary-based approach instead.

### *a) Flesch Kincaid*

Of the readability measures we experimented with, the Flesch Kincaid score is perhaps the best known. A higher Flesch Kincaid score indicates material that is easier to read; lower numbers mark passages that are more difficult to read. The formula for the Flesch Kincaid score is:

206.835 - 1.015 (total words / total sentences) + 84.6 (total syllables / total words)

Hence, the Flesch Kincaid score critically depends on sentence length. Unlike in natural language texts where sentences are the primary structural unit, legal texts of a legislative kind use semi-colons and paragraphing to structure text. Off-the-shelf readability measures, like Flesch Kincaid can thus not be used in the legal context without further adaptation.

We thus sought to pre-process our text prior to running the Flesch Kincaid algorithm. First, the data had to be processed to make the data work with the Quanteda package in R. The need for processing the data is illustrated by the following example. Bulleted lists appear in many formats in regulations, often without periods between each paragraph level. Instead, they are separated by semi-colons. Since the Flesch Kincaid algorithm will not recognize the sentence breaks due to the absence of periods, we needed to process the text by converting semi-colons into periods. Considering the example above, several functions were performed to ensure accurate calculations by the Quanteda package. Brackets (ex. "[" and "]") were removed, as text containing the words "[Repealed…]" were likely skewing the results. Parentheses were also processed to attempt to control bulleted lists without thereby also including cross-references like "Section 11(b)".

Unfortunately, the preprocessing steps were not enough to generate meaningful results. After manually reading through high- and low-scoring regulations returned by the Flesch Kincaid algorithm, the ordering made little sense to the human reader. Texts with high scores often receive such a score because it has numerous paragraphed lists, which is being treated as a long ongoing sentence rather than individual sentences.

We thus concluded that the underlying algorithm would have to be adapted to create meaningful results. We do not know how exactly the Quanteda package in R analyzes the data to determine the Flesch Kincaid score. While we processed the text to account for shortcomings, we are still left with skewed results. It appears that the function does not simply rely on a period denoting the end of a sentence. Additionally, it appears as though line breaks are not present in the text, or they are simply unrecognized by the function, because those with high mean sentence lengths are very paragraph-list heavy. Without knowing how the algorithm operates, we could not find an appropriate solution. Since the creation of a new algorithm was beyond the scope of the project, we shifted to a dictionary approach instead.

### *b) Dictionary approach*

Using word lists, we gained more reliable insights into a regulation's complexity by counting occurrences of particular words. We used word lists containing the Ann Nowak list of words indicating complexity and a wordlist for cross-referencing. The idea behind each wordlist and the method of analysis is explained in each method's following section.

*Ann Nowak Method*

The Ann Nowak word list contains Ann Nowak's list of words that unnecessarily increase the complexity of a document, explained earlier in this report. The list of unnecessarily words contains the following:

- Due to the fact that
- In spite of the fact that
- In the event that
- In the course of
- In numerous instances
- For the purpose of
- Notwithstanding
- Pursuant to
- Hereby

*Cross-Referencing Method*

The wordlist for the cross-referencing method contains word that indicate a given regulation's reference to sections, subsections, clauses, etc. within that regulation. A high word count of these words indicates poor organization in a given document. The wordlist contains the following words:

- Section
- Subsection
- Paragraph
- Subparagraph
- Subclause
- Clause

### 4.6. Age

The Age analysis consists of two elements.

First, a numerical age is calculated based on the meta-data. It takes today's date and subtracts the last amended date of a given regulation. This calculation provides us with insight into which regulations are the oldest, or in other words, have not been amended in a long period of time.

Second, we use a dictionary to list outdated words. The presence of these words in each regulation is counted and normalized as counts of words of interest per 100 words of regulation text.

The wordlist contains the following words:

- Bank draft
- Bond
- Carbon copy
- Card
- Cash

- Certified copy
- Cheque
- Courier
- Stamp
- Facsimile
- Fax
- Hard copy
- In writing
- Ink
- Mail
- Microfiche
- Newspaper
- Original copy
- Paper
- Pen
- Pencil
- Post
- Print
- Printout
- Scan
- Seal
- Telex
- Written

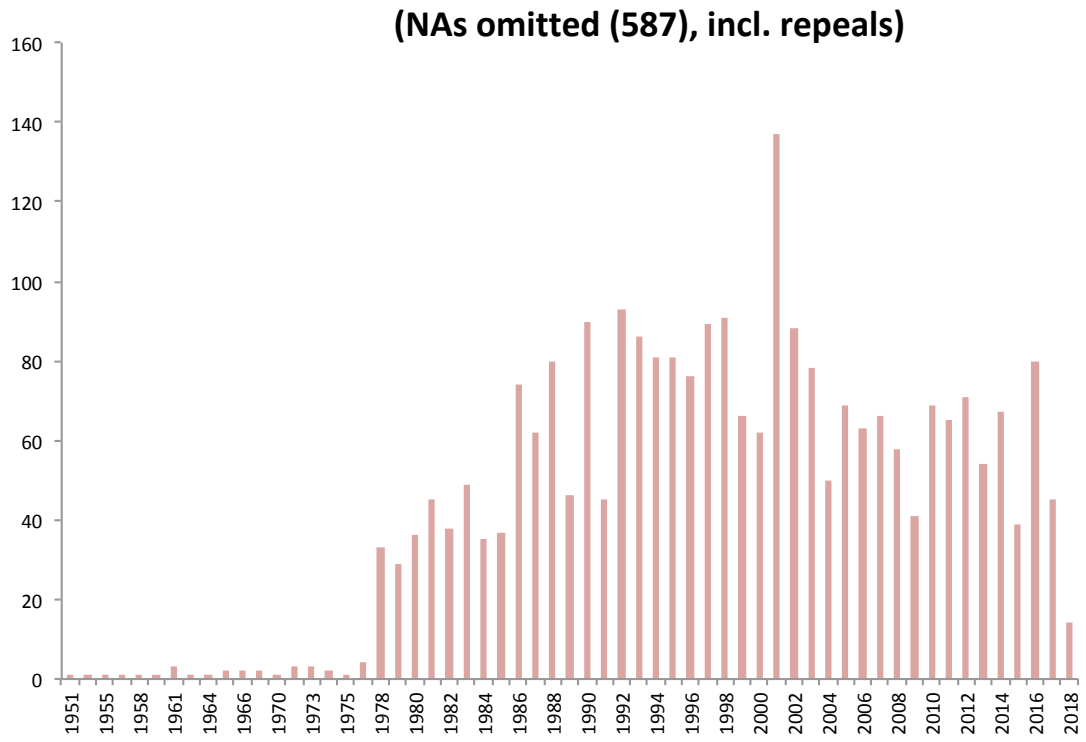
# 5. Findings

In this section, we briefly set out our findings. Top ranking regulations and scores for individual regulations can be found in the accompanying dataset and website.

## 5.1. General regulation corpus analysis

The creation of new regulations in Canada has been fairly even as Figure 1 suggests. Not included in that figure are Consolidated Regulations of Canada (CRCs), which were consolidated on 31 December 1977, but which do not have registration dates. With an average of 2206 words per regulation and 2769 regulations in force, the corpus of Canadian regulations requires scalable tools for legal semantic analysis.

Figure 1: Count of Regulations per registration year

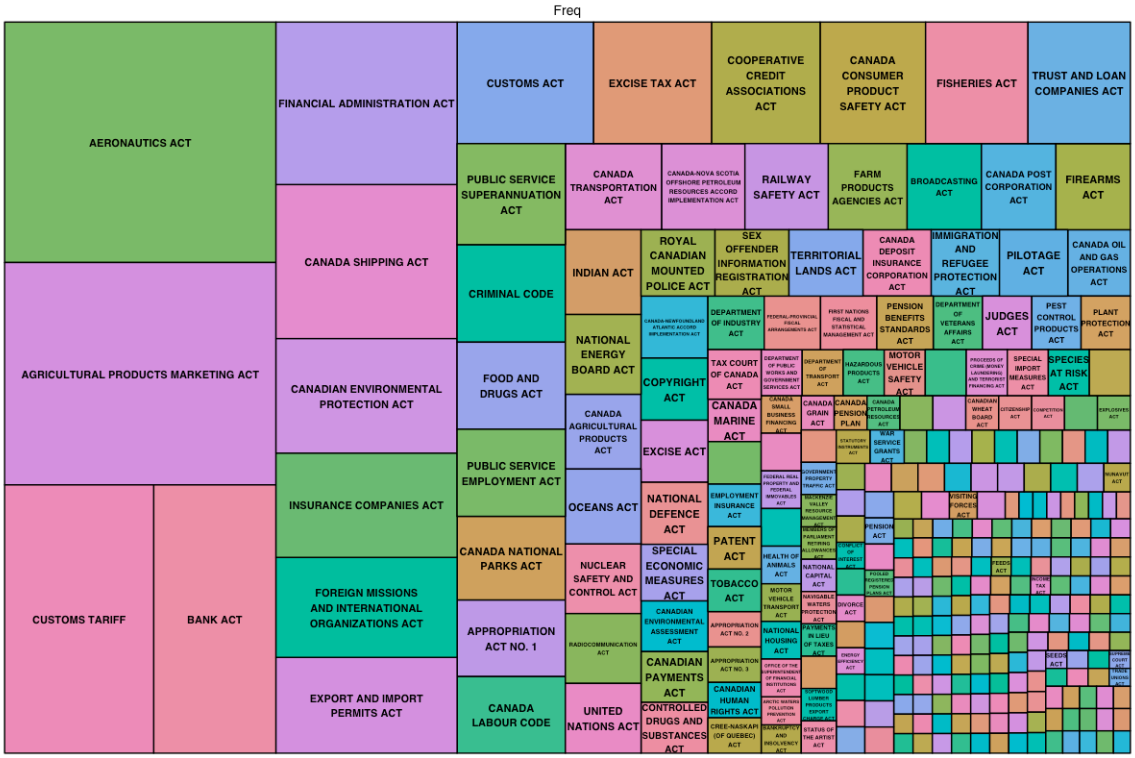



Count of regulations in corpus:
Total Count: 3196
CRCs: 571
SORs: 2625
In force: 2769 (repealed: 427)

Corpus:
Total words: 7'049'156
Average length: 2206 words

Part of the meta-data for each regulation includes the enabling statute, which authorizes a regulation. As Figure 2 shows in form of a tree map, where the area represents the number of regulations enabled under an act as share of all regulations, the Act with the most regulations is the *Aeronautics Act*.

Figure 2: Share of Regulations per enabling authority




At the same time, there is also considerable textual overlap across regulations. The just mentioned *Aeronautics Act*, for instance, has given rise to over 100 airport zoning regulations, which are – apart from the airport's name – almost exact copies of each other. Indeed, 1116 regulations are at least 70% or more textually similar to one or more other regulation(s). Differently put, 2 out of every 5 regulations belong to groups of similarly worded regulations. Hence, the number of uniquely worded regulations only comprise around 60% of all the regulations in force.

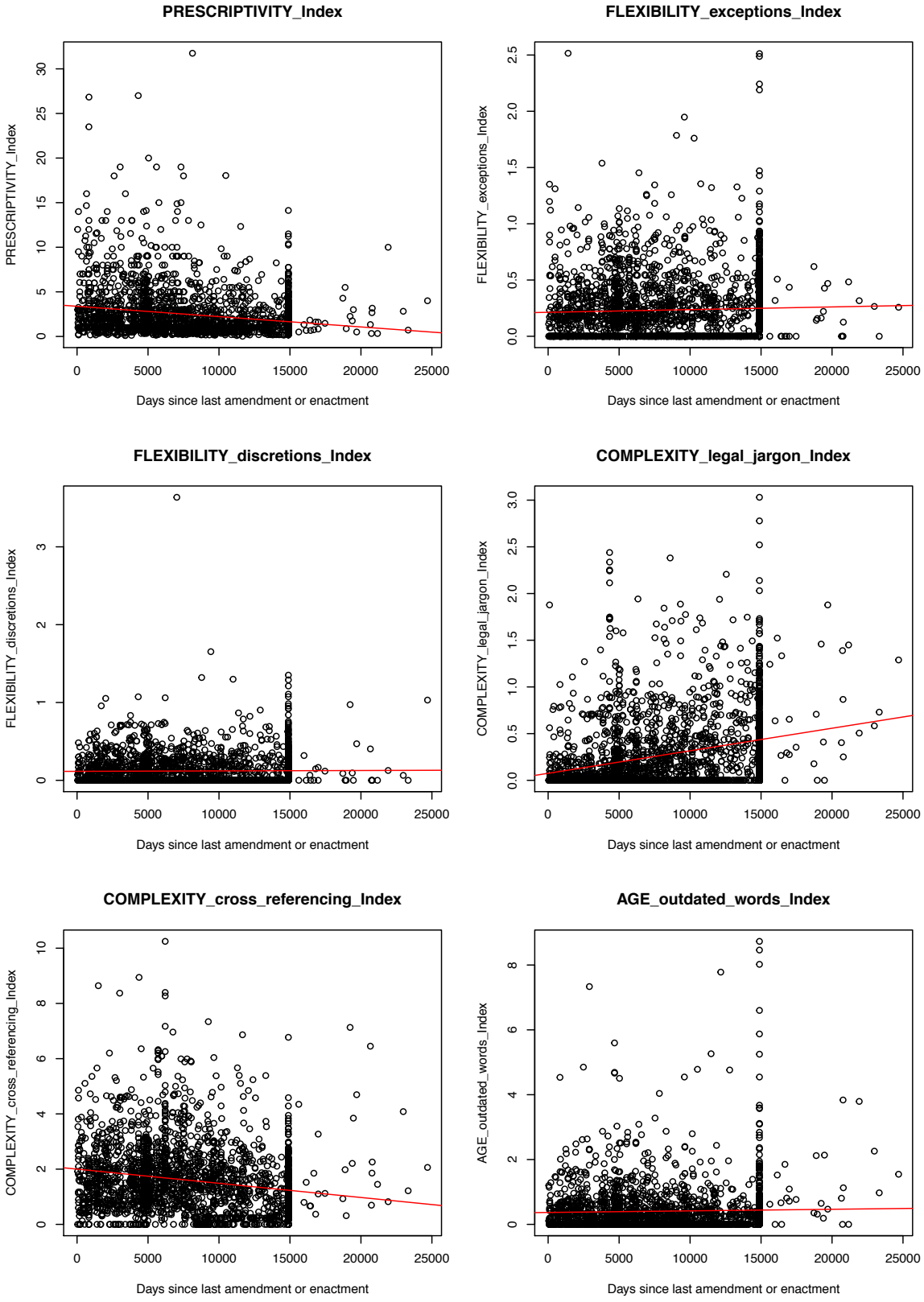
### 5.2. Overall Trends

Our analysis reveals several trends once we compare our indeces to the age of regulations, i.e. the amount of time since they were first enacted or last amended. Figure 3 shows the following trends:

- Prescriptivity of regulations and their complexity in terms of cross-references has *increased* over time.
- Complexity in terms of employing legal jargon has *decreased*.
- Flexibility through exception and discretion words as well as technology-related outdatedness has remained *constant*.

The following sections will explore each dimension in more detail.

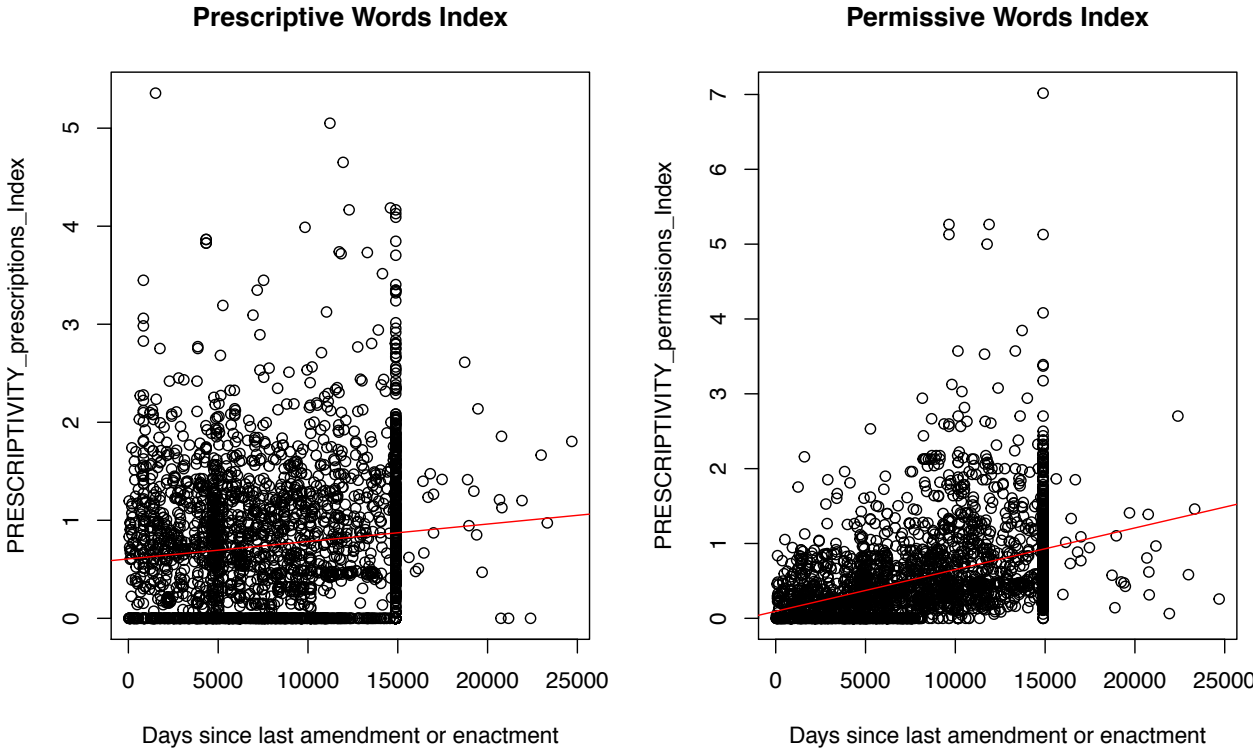
Figure 3: Indeces over time (trend line in red)



### 5.3. Prescriptivity

The prescriptivity of Canadian regulations has increased. Whereas Canadian regulations today tend to contain fewer prescriptive terms than in the past, they contain even fewer flexibility terms (see Figure 4). This renders Canadian regulations overall more rigid and prescriptive. This finding also underscores the importance of conceiving of prescriptivity as a relative measures. Focusing only on one aspect, such as prescriptive words, would obscure larger trends in the data.

Figure 4: Decrease of permissive words outweighs that of prescriptions (trendlinein red)



The mean prescriptivity value of all analyzed texts was 2.24, with a median value of 1.26.

*Most prescriptive*

| | Regulation | ID | Score | Words | Year |
|---|---|---|---|---|---|
| 1. | Transportation Information Regulations | SOR/96-334 | 31.75 | 11457 | 1996 |
| 2. | 2-Butoxyethanol Regulations | SOR/2006-347 | 27 | 1323 | 2006 |
| 3. | Cribs, Cradles and Bassinets Regulations | SOR/2016-152 | 26.83 | 8819 | 2016 |
| 4. | Carriages and Strollers Regulations | SOR/2016-167 | 23.5 | 1503 | 2016 |
| 5. | Conditions of Transferring Firearms and Other Weapons Regulations | SOR/98-202 | 20 | 867 | 2004 |

*Least prescriptive*

| | Regulation | ID | Score | Words | Year |
|---|---|---|---|---|---|
| 1. | Officers Authorized to Exercise Powers or Perform Duties of the Minister of National Revenue Regulations | SOR/86-1066 | 0.02 | 3860 | 1986 |
| 2. | CEFTA Rules of Origin for Casual Goods Regulations | SOR/2009-199 | 0.1 | 561 | 2009 |
| 3. | Non-residents' Temporary Importation of Baggage and Conveyances Regulations | SOR/87-720 | 0.1 | 1442 | 1987 |
| 4. | Administrative Monetary Penalties and Notices (CSA 2001) Regulations | SOR/2008-97 | 0.125 | 1115 | 2008 |
| 5. | CEFTA Tariff Preference Regulations | SOR/2009-200 | 0.125 | 789 | 2009 |

*Prescriptivity by Enabling Authority*

| Row Labels | Prescriptivity/Flexibility | Number of Regulations |
|---|---|---|
| PUBLIC SERVICE SUPERANNUATION ACT,FINANCIAL ADMINISTRATIOI | 0.30434783 | 25 |
| CUSTOMS TARIFF | 0.63593772 | 186 |
| TERRITORIAL LANDS ACT | 0.75686276 | 34 |
| PUBLIC SERVICE EMPLOYMENT ACT | 0.91346674 | 49 |
| FINANCIAL ADMINISTRATION ACT | 1.07710233 | 154 |
| EXCISE ACT, 2001 | 1.13333333 | 13 |
| AERONAUTICS ACT | 1.1974783 | 176 |
| AGRICULTURAL PRODUCTS MARKETING ACT | 1.22516077 | 164 |
| INSURANCE COMPANIES ACT | 1.23921435 | 61 |
| COPYRIGHT ACT | 1.36031746 | 18 |
| ROYAL CANADIAN MOUNTED POLICE ACT | 1.47224084 | 12 |
| SEX OFFENDER INFORMATION REGISTRATION ACT | 1.5 | 13 |
| NATIONAL DEFENCE ACT | 1.51974407 | 12 |
| EXCISE TAX ACT | 1.54712919 | 53 |
| COOPERATIVE CREDIT ASSOCIATIONS ACT | 1.6797619 | 42 |
| TRUST AND LOAN COMPANIES ACT | 1.72273663 | 43 |
| EXPORT AND IMPORT PERMITS ACT | 1.72953216 | 65 |
| CUSTOMS ACT | 1.90279895 | 49 |
| BANK ACT | 2.13416896 | 85 |
| CRIMINAL CODE | 2.23689627 | 42 |
| Average | 2.50993875 | 2770 |
| OCEANS ACT | 2.58253968 | 17 |
| CANADA NATIONAL PARKS ACT | 2.73489641 | 28 |
| INDIAN ACT | 2.755693 | 26 |
| FOREIGN MISSIONS AND INTERNATIONAL ORGANIZATIONS ACT | 2.90542636 | 64 |
| FIRST NATIONS FISCAL AND STATISTICAL MANAGEMENT ACT | 2.92020934 | 13 |
| PILOTAGE ACT | 3.26233766 | 12 |
| FISHERIES ACT | 3.32266356 | 37 |
| BROADCASTING ACT | 3.32973373 | 16 |
| EMPLOYMENT INSURANCE ACT | 3.40515927 | 10 |
| FARM PRODUCTS AGENCIES ACT | 3.41972789 | 28 |
| CANADA POST CORPORATION ACT | 3.51549242 | 16 |
| CANADA-NOVA SCOTIA OFFSHORE PETROLEUM RESOURCES ACCORD | 4.05432673 | 15 |
| CANADA OIL AND GAS OPERATIONS ACT | 4.07631934 | 10 |
| FOOD AND DRUGS ACT | 4.178809 | 25 |
| NATIONAL ENERGY BOARD ACT | 4.28148148 | 18 |
| CANADA AGRICULTURAL PRODUCTS ACT | 4.32910875 | 13 |
| FIREARMS ACT | 4.3308658 | 18 |
| UNITED NATIONS ACT | 4.61634199 | 13 |
| SPECIAL ECONOMIC MEASURES ACT | 4.625 | 18 |
| CANADA TRANSPORTATION ACT | 4.83418803 | 32 |
| PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMEN | 5 | 11 |
| SPECIES AT RISK ACT | 5 | 22 |
| CANADA DEPOSIT INSURANCE CORPORATION ACT | 5.03518519 | 19 |
| CANADA LABOUR CODE | 5.29857257 | 21 |
| CANADIAN ENVIRONMENTAL PROTECTION ACT, 1999 | 5.53370534 | 56 |
| CANADA SHIPPING ACT, 2001 | 5.60276798 | 53 |
| RAILWAY SAFETY ACT | 5.96547564 | 20 |
| NUCLEAR SAFETY AND CONTROL ACT | 6.28670912 | 14 |
| CANADIAN ENVIRONMENTAL ASSESSMENT ACT | 6.29411765 | 10 |
| RADIOCOMMUNICATION ACT | 7 | 15 |
| CANADA CONSUMER PRODUCT SAFETY ACT | 11.0040404 | 38 |

*Examples*

## Carriages and Strollers Regulations
SOR/2016-167 (Score very high)

…

**Child restraint system — stroller**

(6) Every seating unit of a stroller must be equipped with a child restraint system that meets the following requirements:

(a) it must be permanently attached to the frame or upholstery of the stroller;

(b) its anchorages must not break or become detached from their attachment point when tested in accordance with Schedule 4;

(c) it must consist of a lap belt and an additional restraint to prevent the child from sliding downward; and

(d) the fastener of the lap belt must not loosen more than 15 mm or break when tested in accordance with Schedule 4.

## National Parks of Canada Cottages Regulations
SOR/79-398 (Score high)

**Accessory Building Requirements**

**6** Every accessory building erected, altered, reconstructed, added to or enlarged after the coming into force of these Regulations shall comply with the following requirements:

(a) the total floor area in the aggregate of all accessory buildings on a cottage lot shall not exceed 37 square metres;

(b) the accessory building shall be located

(i) at least five metres from the cottage, clear of all projections,

(ii) clear of all projections, at least

(A) one metre from the lot lines, if it is located in that portion of a side or rear yard that does not abut on a street, or

(B) five metres from the lot lines if it is located in that portion of a side or rear yard that abuts on a street;

(c) the accessory building shall not be located in a front yard;

(d) the height shall not exceed three metres or one storey, whichever is the lesser;

(e) the building materials used for the exterior finish of the accessory building shall be of the same quality as those used for the cottage;

(f) the appearance of the accessory building shall be compatible with the cottage and the natural characteristics of the park in which it is located; and

(g) the development shall not adversely affect the characteristics of the surrounding area.

## Yukon Territory Supreme Court Rules for Pre-Hearing Conferences in Criminal Matters
SOR/88-427 (Score low)

**Pre-Hearing Conference**

**3** A pre-hearing conference shall be held at such time and date and in such place and manner as a judge may direct, or at such further dates, times and places as the judge who presides at the pre-hearing conference may direct.

**4** Unless otherwise ordered by the judge who presides at the pre-hearing conference, the pre-hearing conference shall be an informal meeting conducted in chambers at which a full and free discussion of the issues raised may occur without prejudice to the rights of the parties in any proceedings thereafter taking place.

**5** Unless otherwise ordered by a judge, counsel for the Crown and counsel for the accused, each fully briefed in respect of the issues to be discussed at the pre-hearing conference, or, in the case of an accused who is not represented by counsel, the accused, shall be present at the pre-hearing conference.

**6** A judge may require that an accused, represented by counsel, be available for consultation with counsel in respect of matters to be considered at the pre-hearing conference.

*Prescriptivity depends on the policy context*

Two of the five highest-scoring regulations for Prescriptivity deal with consumer items for the transportation and care of people. The third and fourth most prescriptive regulations deal with children specifically. Childcare is important to our society, and we would expect that regulations surrounding these consumer products should be rigid. The highest scoring regulation for prescriptivity deals with regulation of air, land, and sea carriers. Again, a rigid emphasis on rules makes sense in the context of safely transporting passengers and goods.

One interesting anomaly among these top-scoring regulations are the 2-Butoxyethanol Regulations. These regulations came into force in 2006, following a Health Canada report about 2-Butoxyethanol's harmfulness to human health. 2-Butoxyethanol has many industrial uses, as well as being found in many household cleaners. The regulations specify maximum allowable concentration, and labeling requirements. Given the harmful nature of 2-Butoxyethanol to humans, and the exposure of this fact to the public by Health Canada, it is not surprising that regulations around its use in commercial cleaners are strict.

At the opposite end of the spectrum, two of the five regulations are concerned primarily with defining "casual goods" for the purposes of assigning tariffs to the property authority. These regulations are minimally prescriptive, as they are concerned simply with definitions and allocations of payable tariffs. The least prescriptive regulation deals with the delegation of powers or duties by the Minister of National Revenue, specifically who is permitted to assume these responsibilities. As with the regulations dealing with casual goods, this regulation is, for the most part, a simple list. The third least-prescriptive regulation regulates the baggage and mode of transportation non-residents bring with them during their stay. Finally, the fourth least prescriptive regulation outlines the ranges of penalties to be paid by individuals that violate various provisions of regulations dealing with import of goods. A commonality in all these regulations is their nature akin to lists of criteria, rather than imperatives.

In sum, prescriptivity appears to be driven at least in part by the policy context.

*Too much prescriptivity may be a drafting or a regulatory issue*

High prescriptivity scores may be desirable where they efficiently protect important interests, but they may also indicate drafting or regulatory issues. The *Regulation on Carriages and Strollers*, for instance, may protect a vital interest, the safety of infants, but is also drafted poorly as the repetition of "must" in each sub-paragraph is redundant given the clause's chapeau.

In other instance, high prescriptivity may indicate over-regulation. The *National Parks of Canada Cottages Regulations*, for instance, specifies in minute detail Accessory Building Requirements. Whether this high prescriptivity is indeed necessary to achieve the regulation's purpose may be questioned. Hence, high prescriptivity could indicate different shortcomings of a regulation.

*Additional analysis needed to find best practices*

High prescriptivity is not necessarily bad. Nor is low prescriptivity necessarily good. Our analysis suggest that further analysis is needed to evaluate prescriptivity scores in the context of regulatory reform. Through benchmarking and appriopriate comparator groups, it should be possible to identify best practices among regulations and to reform those that fall below that standards.

For instance, it would be beneficial to categorize regulations by type, rather than simply enabling authority. Recalling the regulations with the first, third, and fourth highest prescriptivity scores (*Transportation Information Regulations*, *Cribs, Cradles and Bassinets Regulations* and *Carriages and Strollers Regulations*), all three deal at least in part with the safe transportation of people. Generating a category that encompasses these and regulations like them (e.g. "transporting humans") would provide a more coherent basis for comparison of Prescriptivity and flexibility. It would allow drafters and legislators to look at a category and question why certain regulations regarding the transport of humans are more prescriptive or flexible whereas others are not, and whether one extreme or the other should be reined in or expanded. These categories could be refined from lists generated by analyzing all available regulations for key words and phrases using machine learning algorithms.

Another approach would be to analyze Prescriptivity and flexibility of regulations by department. For example, evaluating trends over time within *Canada Consumer Product Safety Act*-enabled regulations risks casting too wide a net, and leading to conclusions that may be overly broad and lacking nuance. Narrowing the analysis to the departments drafting the regulations would help to avoid this, as well as allow drafters to evaluate their regulations against a more comparable standard – apples to apples. Beyond this, an intra-departmental approach would likely lead to valuable policy discussion if wide discrepancies in regulations drafted by the same department become apparent.

### 5.4. Flexibility

Flexibility has different dimensions, including the (1) incorporation of outside standards by reference, (2) exceptions for regulatees and (3) discretion for regulators. All three measures have remained constant over time. But they all display significant variation.

*Incorporation by reference frequent when dealing with vehicle safety*

The following tables present the top and bottom five regulations, with respect to flexibility as determined by number of incorporations by reference.

| | Regulation | ID | Number | Words | Year |
|---|---|---|---|---|---|
| 1. | Canadian Aviation Regulations | SOR/96-433 | 672 | 225149 | 2005 |
| 2. | Transportation Information Regulations | SOR/96-334 | 221 | 11457 | 1996 |
| 3. | Motor Vehicle Safety Regulations | C.R.C., c. 1038 | 197 | 20762 | 2005 |
| 4. | Canada Oil and Gas Installations Regulations | SOR/96-118 | 116 | 25102 | 1996 |
| 5. | Nova Scotia Offshore Petroleum Installations Regulations | SOR/95-191 | 112 | 24291 | 1995 |

The top three most flexible regulations, as indicated by number of incorporations by reference, all deal with types of vehicles. It is perhaps intuitive that regulations applying similar provisions to different locations or types of vehicle would have a high number of references to external, vehicle- or location-specific regulations in the interest of efficiency. Additionally, incorporating other, more detailed regulations dealing with the specific items referred to in the referencing regulation likely reduces the chances of contradictions. Indeed, this appears to be the case in the *Canadian Aviation Regulations*, which often refer to, for example, the *Safety Criteria for Approval of Extended Range Twin-engine Operations Manual* in its provisions dealing with Flight Operations.

The fourth and fifth most flexible regulations share the feature of dealing primarily with equipment, as well as construction standards. For example, in the "Design of Offshore Platforms" provisions of the *Canada Oil and Gas Installations Regulations*, each sub-section refers to external regulations such as *Steel Structures, Offshore Structures* and *Foundations, Offshore Structures*.

In general, flexibility analyzed through incorporation by reference points to situations where a "blanket" regulation seeks to impose guidelines on an industry, while allowing for changing standards for its various specific provisions.

*Flexibility (Incorporation by Reference) by Enabling Authority*

| Enabling Regulations | External References | Count of Regulations |
|---|---|---|
| FIRST NATIONS FISCAL AND STA | 1 | 13 |
| FOREIGN MISSIONS AND INTERN | 1 | 64 |
| OCEANS ACT | 1 | 17 |
| PERSONAL INFORMATION PROT | 1 | 11 |
| SPECIES AT RISK ACT | 1 | 22 |
| TERRITORIAL LANDS ACT | 1 | 34 |
| AGRICULTURAL PRODUCTS MAR | 1.170212766 | 164 |
| COOPERATIVE CREDIT ASSOCIAT | 1.25 | 42 |
| INSURANCE COMPANIES ACT | 1.25 | 61 |
| TRUST AND LOAN COMPANIES A | 1.25 | 43 |
| PUBLIC SERVICE EMPLOYMENT A | 1.333333333 | 49 |
| BANK ACT | 1.384615385 | 85 |
| CANADA NATIONAL PARKS ACT | 1.5 | 28 |
| CUSTOMS TARIFF | 1.5 | 186 |
| INDIAN ACT | 1.5 | 26 |
| NATIONAL DEFENCE ACT | 1.5 | 12 |
| FINANCIAL ADMINISTRATION AC | 1.525 | 154 |
| CANADA POST CORPORATION A | 1.555555556 | 16 |
| FARM PRODUCTS AGENCIES ACT | 1.888888889 | 28 |
| UNITED NATIONS ACT | 2 | 13 |
| NATIONAL ENERGY BOARD ACT | 2.111111111 | 18 |
| CRIMINAL CODE | 2.333333333 | 42 |
| CUSTOMS ACT | 2.333333333 | 49 |
| PILOTAGE ACT | 2.555555556 | 12 |
| EXPORT AND IMPORT PERMITS A | 2.619047619 | 65 |
| FIREARMS ACT | 3 | 18 |
| NUCLEAR SAFETY AND CONTROL | 3.285714286 | 14 |
| FOOD AND DRUGS ACT | 3.47826087 | 25 |
| CANADA CONSUMER PRODUCT | 3.625 | 38 |
| CANADA AGRICULTURAL PRODU | 3.9 | 13 |
| COPYRIGHT ACT | 4 | 18 |
| RAILWAY SAFETY ACT | 4.111111111 | 20 |
| FISHERIES ACT | 4.227272727 | 37 |
| EXCISE TAX ACT | 4.571428571 | 53 |
| CANADIAN ENVIRONMENTAL PR | 4.962962963 | 56 |
| SPECIAL ECONOMIC MEASURES | 5.777777778 | 18 |
| BROADCASTING ACT | 5.8 | 16 |
| Average | 5.908333333 | 2770 |
| AERONAUTICS ACT | 8.153061224 | 176 |
| CANADA SHIPPING ACT, 2001 | 11.67647059 | 53 |
| CANADA-NOVA SCOTIA OFFSHO | 17.28571429 | 15 |
| CANADA DEPOSIT INSURANCE C | 18.75 | 19 |
| CANADA LABOUR CODE | 36.57142857 | 21 |
| CANADA TRANSPORTATION ACT | 38.33333333 | 32 |
| EXCISE ACT, 2001 | 45 | 13 |

Not all regulations seem to equally incorporate external standards by reference. Particularly in transportation and finance, regulations refer frequently to external standards. In other sectors, including administrative affairs and agriculture, the figures are much lower.

*Exception scores are highest for regulations that exempt otherwise prohibted practices*

Turning to flexibility for regulatees through exceptions, scores have remained constant over time, but vary significantly across types of regulations. High scoring regulations are typically those that exempt certain practices or that allow the use of substances otherwise prohibited. In that vein, the short, highest ranking regulation by exception score is the *Order Declaring that the Wastewater Systems Effluent Regulations Do Not Apply in Yukon* (159 words). The highest ranking regulation over 1000 words is the 2012 *Prohibition of Certain Toxic Substances Regulations*, which allows the use of otherwise restricted substances. In short, the index of exception words is a good indicator to identify regulations that carve-out specific practices or usages.

*Regulations admistering public properties or processes have high discretion scores*

Finally, flexibility for regulators through discretion is highest when regulation deal with public property or processes that primarily affect the internal workings of government. A good example is the *Regulations Respecting the Writing Off of any Debt or Obligation Due to Her Majesty or any Claim by Her Majesty,* which gives the relevant ministry the discretion to cancel uncollectable debt to the government.

### 5.5. Complexity

Complexity scores are marked by two opposing trends. On the one hand, regulations have become more complex as they contain more cross-references to other parts of the same regulation or outside acts making texts more difficult to navigate. On the other hand, regulations relied less on legal jargon which makes texts more accessible to lay readers.

Regulations vary by both complexity scores. First, by cross-references: longer and more extensive cross-references indicate greater complexity. The mean complexity value of all analyzed texts, as determined by internal cross-reference, was 1.59, with a median value of 1.38.

*Most complex by cross-reference*

| | **Regulation** | **ID** | **Score** | **Words** | **Year** |
|---|---|---|---|---|---|
| 1. | Exemption from Restrictions on Investments (Insurance Companies, Insurance Holding Companies and Societies) Regulations | SOR/2001-385 | 10.25 | 283 | 2001 |
| 2. | Period for Entering into an Agreement for the Purpose of Jointly Establishing a Review Panel Regulations | SOR/2006-252 | 8.94 | 123 | 2006 |
| 3. | Order Designating Public Office Holders and Reporting Public Office Holders under Section 62.2 of the Conflict of Interest Act | SOR/2014-200 | 8.64 | 243 | 2014 |
| 4. | Exemption from Restrictions on Investments (Cooperative Credit Associations) Regulations | SOR/2001-384 | 8.4 | 131 | 2001 |
| 5. | Regulations Prescribing Certain Offences to be Serious Offences | SOR/2010-161 | 8.33 | 215 | 2010 |

*Least complex by cross-reference*

| | Regulation | ID | Score | Words | Year |
|---|---|---|---|---|---|
| 1. | By-laws Nos. 6 and 8 of VIA Rail Canada Inc. | SOR/79-817 | 0.08 | 1231 | 1979 |
| 2. | Railway Passenger Services Contract Regulations | SOR/78-286 | 0.22 | 893 | 1978 |
| 3. | Denatured and Specially Denatured Alcohol Regulations | SOR/2005-22 | 0.13 | 2374 | 2005 |
| 4. | National Security and Intelligence Committee of Parliamentarians Regulations | SOR/2017-222 | 0.13 | 779 | 2017 |
| 5. | Regulations Respecting the Administration and Distribution of the Estates of Deceased Members of the Canadian Forces | C.R.C., c. 1048 | 0.13 | 760 | 1977 |

*Complexity (Internal Cross-References) by Enabling Authority*

| Enabling Authorities | Complexity (internal cross | Count of regu |
|---|---|---|
| AERONAUTICS ACT | 0.41139972 | 176 |
| TERRITORIAL LANDS ACT | 0.58948733 | 34 |
| AGRICULTURAL PRODUCTS MARKETING ACT | 0.7697369 | 164 |
| CANADA CONSUMER PRODUCT SAFETY ACT | 0.79323226 | 38 |
| OCEANS ACT | 0.82256877 | 17 |
| FINANCIAL ADMINISTRATION ACT | 0.86424397 | 154 |
| NATIONAL DEFENCE ACT | 0.91479459 | 12 |
| FOREIGN MISSIONS AND INTERNATIONAL ORGANIZATIONS ACT | 0.92094415 | 64 |
| EXPORT AND IMPORT PERMITS ACT | 0.933912 | 65 |
| PILOTAGE ACT | 0.94941449 | 12 |
| FARM PRODUCTS AGENCIES ACT | 0.94960911 | 28 |
| CANADA POST CORPORATION ACT | 0.97996394 | 16 |
| BROADCASTING ACT | 1.01604874 | 16 |
| CUSTOMS TARIFF | 1.03688021 | 186 |
| INDIAN ACT | 1.18927779 | 26 |
| PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS | 1.19406474 | 11 |
| RAILWAY SAFETY ACT | 1.20579593 | 20 |
| CANADA AGRICULTURAL PRODUCTS ACT | 1.21203538 | 13 |
| NUCLEAR SAFETY AND CONTROL ACT | 1.22704673 | 14 |
| CANADA SHIPPING ACT, 2001 | 1.26475504 | 53 |
| CANADA TRANSPORTATION ACT | 1.29584452 | 32 |
| NATIONAL ENERGY BOARD ACT | 1.48479664 | 18 |
| UNITED NATIONS ACT | 1.52097995 | 13 |
| FOOD AND DRUGS ACT | 1.5426827 | 25 |
| SPECIAL ECONOMIC MEASURES ACT | 1.5594796 | 18 |
| ROYAL CANADIAN MOUNTED POLICE ACT | 1.56591019 | 12 |
| CANADA NATIONAL PARKS ACT | 1.60749382 | 28 |
| Average | 1.66126893 | 2770 |
| CANADIAN ENVIRONMENTAL PROTECTION ACT, 1999 | 1.74756398 | 56 |
| CANADA LABOUR CODE | 1.76429997 | 21 |
| PUBLIC SERVICE EMPLOYMENT ACT | 1.77436243 | 49 |
| FISHERIES ACT | 1.80968149 | 37 |
| TRUST AND LOAN COMPANIES ACT | 2.15379663 | 43 |
| COOPERATIVE CREDIT ASSOCIATIONS ACT | 2.16633955 | 42 |
| FIRST NATIONS FISCAL AND STATISTICAL MANAGEMENT ACT | 2.18207649 | 13 |
| FIREARMS ACT | 2.19323208 | 18 |
| CANADA-NOVA SCOTIA OFFSHORE PETROLEUM RESOURCES ACCORD IN | 2.2389924 | 15 |
| COPYRIGHT ACT | 2.26414282 | 18 |
| BANK ACT | 2.27810371 | 85 |
| EXCISE ACT, 2001 | 2.31530259 | 13 |
| CANADA DEPOSIT INSURANCE CORPORATION ACT | 2.43471843 | 19 |
| CUSTOMS ACT | 2.45638675 | 49 |
| CRIMINAL CODE | 2.48202452 | 42 |
| EXCISE TAX ACT | 2.54853743 | 53 |
| INSURANCE COMPANIES ACT | 2.57989944 | 61 |
| SPECIES AT RISK ACT | 2.92363094 | 22 |
| RADIOCOMMUNICATION ACT | 3.89237852 | 15 |
| PUBLIC SERVICE SUPERANNUATION ACT,FINANCIAL ADMINISTRATION A | 4.05177039 | 25 |

Second, complexity by legal jargon varies over time as well. It had a mean complexity value of all analyzed texts was 0.32, with a median value of 0.08.

*Most complex by legal jagon*

| | Regulation | ID | Score | Words | Year |
|---|---|---|---|---|---|
| 1. | Withdrawal of Certain Lands (South Nahanni River N.W.T.) from Disposal Order, 1971 | C.R.C., c. 1540 | 3.03 | 132 | 1977 |
| 2. | Withdrawal of Certain Lands (Lockhart River N.W.T.) from Disposal Order | C.R.C., c. 1537 | 2.78 | 108 | 1977 |
| 3. | Withdrawal of Certain Lands (Dubawnt Lake N.W.T.) from Disposal Order | C.R.C., c. 1533 | 2.52 | 119 | 1977 |
| 4. | Civil Remedies (Banks and Bank Holding Companies) Regulations | SOR/2006-299 | 2.44 | 246 | 2006 |
| 5. | Statistics Canada 1996 Census of Population Terms Exclusion Approval Order | SOR/95-153a | 2.38 | 126 | 1995 |

*Least complex by legal jargon*

| | Regulation | ID | Score | Words | Year |
|---|---|---|---|---|---|
| 1. | Motor Vehicle Restraint Systems and Booster Seats Safety Regulations | SOR/2010-90 | 0.005 | 20037 | 2010 |
| 2. | Blood Regulations | SOR/2013-178 | 0.007 | 13386 | 2013 |
| 3. | Fishing Vessel Safety Regulations | C.R.C., c. 1486 | 0.011 | 18197 | 1977 |
| 4. | Railway Safety Management System Regulations, 2015 | SOR/2015-26 | 0.012 | 8094 | 2015 |
| 5. | Vessel Fire Safety Regulations | SOR/2017-14 | 0.013 | 23792 | 2017 |

*Older regulations use more legal jargon while newer ones have more cross-references*

One of the most complex regulations with heavy legal jargon was drafted over twenty years ago, and three of the top five nearly forty years ago. In contrast, four of the five least complex regulations were drafted within the last eight years. This may indicate a trend towards more comprehensible regulations, or a move away from impenetrable traditional drafting language.

At the same time, there is a trend of higher complexity, as determined by internal cross-references, in more recent regulations. One reason for the increasing structural complexity may be that regulators are working more with existing regulations rather than creating new ones from scratch. As more norms are added to the existing text stock, new cross-references become necessary. Another driver of cross-references is brevity. Short regulations, such as the *Exemption from Restrictions on Investments Regulations* which offer specific exceptions to subsections within the *Insurance Companies Act*, for instance, make extensive use of cross-references in their provisions. At the other end of the spectrum, the least complex regulations as determined by cross-referencing share the characteristic of standing on their own. For example, the *National Security and Intelligence Committee of Parliamentarians Regulations* serve to completely outline topics such as security clearances and practices for dealing with sensitive information. It is again intuitive that these kinds of regulations, such as those dealing with handling of sensitive documents, should be comprehensive and immediately understandable to the intended audience.

A further refinement of this approach might be to modify the target of the wordlist to include only the regulations themselves, rather than a more expansive approach to regulations under the same enabling authority.

### 5.6. Age

We approach age from two angels. First, we calculate the time (in days) since a regulation has been last amended. If a regulation has not been amended, the time is calculated based on the date of enactment. The intuition is that older regulations (based on their enactment date) may have been updated over time. Hence their enactment date does not anymore reflect their true age. Second, we also count the number of outdated words.

*Oldest*

| | **Regulation** | **ID** | **Age** (Days) | **Words** | **Year** |
|---|---|---|---|---|---|
| 1. | Application of Defence Services Pension Act to Special Force | SOR/51-84 | 24690 | 388 | 1951 |
| 2. | War Claims Regulations | SOR/54-578 | 23327 | 411 | 1954 |
| 3. | Defence Services Pension Part V Regulations | SOR/55-416 | 22982 | 9430 | 1955 |
| 4. | Pension Increase Regulations, 1958 | SOR/58-417 | 21918 | 1582 | 1958 |
| 5. | Quebec Apple Growers' Marketing Order | SOR/60-481 | 21169 | 207 | 1960 |

*Most recent*

| | **Regulation** | **ID** | **Age** (Days) | **Words** | **Year** |
|---|---|---|---|---|---|
| 1. | Cannabis Tracking System Order | SOR/2018-178 | 42 | 2763 | 2018 |
| 2. | Order Respecting the Calculation of the Moisture Shrinkage for Grain | SOR/2018-176 | 43 | 167 | 2018 |
| 3. | Critical Habitat of the Western Chorus Frog (Pseudacris triseriata) Order | SOR/2018-169 | 66 | 113 | 2018 |
| 4. | Regulations for the Monitoring of Medical Assistance in Dying | SOR/2018-166 | 70 | 2619 | 2018 |
| 5. | Critical Habitat of the Lake Chubsucker (Erimyzon sucetta) Order | SOR/2018-156 | 91 | 443 | 2018 |

Second, we also count the number of technologically outdated words. The results are only partially meaningful since specific technologies, such as the mailing of letters, are specifically regulated through regulations. Unsurprisingly then, those regulations dealing with postage, mail and letters score very high on our index.

*Highest incidence of outdated words*

| | **Regulation** | **ID** | **Score** | **Words** | **Year** |
|---|---|---|---|---|---|
| 1. | Regulations Respecting the Reproduction of Postage Stamps | C.R.C., c. 1292 | 8.73015873 | 252 | 1977 |
| 2. | Regulations Respecting the Posting Abroad of Letter-Post Items | C.R.C., c. 1288 | 8.461538462 | 390 | 1977 |
| 3. | Regulations Made Pursuant to the Public Service Pension Adjustment Act | C.R.C., c. 1352 | 8.021390374 | 187 | 1977 |
| 4. | Regulations Respecting Mailable Matter Bearing No Postage or Insufficient Postage | SOR/85-567 | 7.780979827 | 347 | 1985 |
| 5. | Postage Meter Regulations, 2010 | SOR/2010-220 | 7.334963325 | 818 | 2010 |

*Age (outdated words) by Enabling Authority*

| Enabling Authority | Outdated Words Index | Count of Reg |
|---|---|---|
| RADIOCOMMUNICATION ACT | 0 | 15 |
| SPECIES AT RISK ACT | 0.01782531 | 22 |
| AGRICULTURAL PRODUCTS MARKETING ACT | 0.01990159 | 164 |
| AERONAUTICS ACT | 0.04106201 | 176 |
| CUSTOMS TARIFF | 0.05861775 | 186 |
| FOOD AND DRUGS ACT | 0.06601489 | 25 |
| FARM PRODUCTS AGENCIES ACT | 0.07073285 | 28 |
| OCEANS ACT | 0.08518647 | 17 |
| UNITED NATIONS ACT | 0.08624379 | 13 |
| PILOTAGE ACT | 0.09718377 | 12 |
| TERRITORIAL LANDS ACT | 0.1033721 | 34 |
| SEX OFFENDER INFORMATION REGISTRATION ACT | 0.10544027 | 13 |
| PUBLIC SERVICE EMPLOYMENT ACT | 0.1142876 | 49 |
| BROADCASTING ACT | 0.11987918 | 16 |
| EXPORT AND IMPORT PERMITS ACT | 0.12200406 | 65 |
| RAILWAY SAFETY ACT | 0.12586669 | 20 |
| CANADA SHIPPING ACT, 2001 | 0.13475735 | 53 |
| CRIMINAL CODE | 0.15546217 | 42 |
| INSURANCE COMPANIES ACT | 0.15555557 | 61 |
| CANADA NATIONAL PARKS ACT | 0.16579044 | 28 |
| FISHERIES ACT | 0.16829068 | 37 |
| FOREIGN MISSIONS AND INTERNATIONAL ORGANIZATIONS ACT | 0.17189577 | 64 |
| CANADA TRANSPORTATION ACT | 0.17210378 | 32 |
| CANADA-NOVA SCOTIA OFFSHORE PETROLEUM RESOURCES ACCORD IMPL | 0.17559608 | 15 |
| CANADA CONSUMER PRODUCT SAFETY ACT | 0.17891049 | 38 |
| CANADA LABOUR CODE | 0.18880687 | 21 |
| COPYRIGHT ACT | 0.20260375 | 18 |
| EXCISE TAX ACT | 0.20924297 | 53 |
| PERSONAL INFORMATION PROTECTION AND ELECTRONIC DOCUMENTS AC | 0.21848073 | 11 |
| CANADIAN ENVIRONMENTAL PROTECTION ACT, 1999 | 0.23634096 | 56 |
| Average | 0.24211739 | 2770 |
| EXCISE ACT, 2001 | 0.24258563 | 13 |
| CANADA DEPOSIT INSURANCE CORPORATION ACT | 0.24374791 | 19 |
| INDIAN ACT | 0.24812507 | 26 |
| NATIONAL ENERGY BOARD ACT | 0.26029762 | 18 |
| FINANCIAL ADMINISTRATION ACT | 0.27740668 | 154 |
| BANK ACT | 0.27796231 | 85 |
| FIRST NATIONS FISCAL AND STATISTICAL MANAGEMENT ACT | 0.30567624 | 13 |
| TRUST AND LOAN COMPANIES ACT | 0.32795789 | 43 |
| COOPERATIVE CREDIT ASSOCIATIONS ACT | 0.33123306 | 42 |
| FIREARMS ACT | 0.35953469 | 18 |
| SPECIAL ECONOMIC MEASURES ACT | 0.39911518 | 18 |
| NUCLEAR SAFETY AND CONTROL ACT | 0.41207402 | 14 |
| PUBLIC SERVICE SUPERANNUATION ACT,FINANCIAL ADMINISTRATION ACT | 0.42705664 | 25 |
| CANADA AGRICULTURAL PRODUCTS ACT | 0.44607374 | 13 |
| NATIONAL DEFENCE ACT | 0.44665022 | 12 |
| CUSTOMS ACT | 0.53087841 | 49 |
| ROYAL CANADIAN MOUNTED POLICE ACT | 0.59195378 | 12 |
| CANADA POST CORPORATION ACT | 4.11868827 | 16 |

*A regulation's age is influenced by novel developments in that area of law*

As seen in the results of "oldest" and "newest" regulations, regulations pertaining to areas of law that are subject to sociological or technological change are likely to undergo amendments. In looking at the "oldest" regulations, they cover behaviour that has been widely accepted in society and thus are not subject to multiple amendments over time. For example. In Canadian society it has been decided, and for the most part accepted, what the authorizations for transporting restricted firearms should be. On the other hand, regulations that are the "newest" are those covering by-laws, marketing, and broadcasting. With changing social behaviour and norms, and novel technologies changing how media is distributed to the masses, it makes sense that such regulations would undergo amendments to remain relevant.

At the same time, outdatedness remains. Some of the most outdated regulations identified are in fact linked to in our data generation. Naturally, regulating Canada Post will include mentions of postal service. Similarly, many regulations that include communication requirements will list increasingly outdated means of communications (e.g. fax) alongside more modern ones (e.g. Email). At the same time, this approach has its downsides when technology advances, but regulations are not amended. The *Tax Court of Canada Rules of Procedure respecting the Canada Pension Plan* from 1990, for instance, specifies that an "application to extend time may be served by filing it in the Registry in which the notice of appeal or notice of intervention was filed or to which it was *mailed, or by sending a letter, telegram, telex or fax* to that Registry." It omits Email.

## 6. Website

Accompanying this report is a github repository and an interactive website, which further details the results and findings from this project. The website is built using the React JS framework based on JavaScript programming language. We selected React JS because of the numerous opensource libraries, the available online documentation and component-oriented design. There are three main components in the website. First, the 'Normal Display', which displays the index value of the regulation based on enabling authority. Second, 'Normal Display – Metadata Table', which displays the regulations' metadata in the chart above. Lastly, the 'Mean Display' chart, which aggregates the regulations based on year and generates an average value based on the year.

**How data is transferred from R to the web application.**

The data displayed in the regulation app is passed from a CSV file created in R which contains all the regulation's metadata, such as prescriptivity, flexibility or age. Afterwards the file is manually converted to JSON, so it can be accessed in JS without further libraries. Within the file index.js, the function *processDataSet* processes the dataset before it is stored as a global object variable. The function *processDataSet* retrieves the data points which contain the specified enabling authority. The function *processMeanDataSet* aggregates the mean for all the data points for each year.

### Data Preprocessing and Data Display

The function *handleConfig* removes the regulations that are repealed and have less than 500 words. Also, the function accepts a highcharts configuration file used to generate the graphs. The other methods inside index.js are used to update the graphs whenever the user interacts with the settings.

### Deployment

Currently, the application is deployed using Heroku, an opensource lightweight deployment tool. The application can be access through this URL: https://regulations.herokuapp.com.

### Github Repository

All regulation data, the script for analyzing the data as well as our wordlists are available on: https://github.com/WAlschner/Semantic-Analysis-Regulations.

## 7. Future Work

This project is a first attempt to conduct an automated semantic analysis of federal Canadian regulations. More work remains to be done on several fronts.

First, the measures created as part of the project need to be scrutinized by the regulatory community. Do the measures align with a manual assessment of the same texts? Have we captured most the important signaling words? Should our measures be combined to simplify their interpretation? A thorough engagement with this report and the accompanying data and website will help answer these questions.

Second, we have identified questions we could not answer. Most importantly, our readability analysis reached an impasse as off-the-shelf measures did not perform reliably on statutory texts. Dedicated research is needed to devise readability measures uniquely suited for the legal domain.

Third, and most importantly, our measures should be further contextualized to inform policy analysis. As we highlighted in relation to prescriptivity, high prescriptivity is not necessarily an indicator of a poor regulation – but it might be. More work is needed to contextualize our measure in order to identify best practices and candidate regulations for regulatory reform.