

RegData

Quantitative Analysis of National Laws in India

March 2021



Copyright ©2021

First published in 2021
by Centre for Civil Society

Authors: Jayana Bedi and Prashant Narang
Centre for Civil Society¹

In collaboration with Mercatus Center

Data Visualisation: Ahmed Pathan

For more information and other requests, write to:

Centre for Civil Society
A-69, Hauz Khas, New Delhi – 110016
Phone: +91 11 26537456
Email: ccs@ccs.in Website: www.ccs.in

¹ We thank Arjun Krishnan for his support in preparing this summary note.

Contents

- Introduction** **1**

- Methodology** **1**
 - Word count 1
 - Binding words 1
 - How did we collate binding words for national laws? 1
 - Linguistic complexity 2

- Summary of Findings** **4**
 - Volume of national laws 4
 - Restrictiveness of national laws 4
 - Linguistic complexity of national laws 5
 - Shannon entropy 5
 - Conditional count 6
 - Flesch reading ease score 6

Introduction

RegData, an initiative of the [Mercatus Center](#), is an effort to quantify various aspects of regulation. The Mercatus Center created RegData with an aim to “introduce an objective, replicable, and transparent methodology for measuring regulation.” It uses custom-made text analysis and a machine-learning algorithm to measure the different features of law. These features include volume, restrictiveness and linguistic complexity. Together, these metrics indicate the regulatory burden a law, department or ministry imposes. Some variables, like the restrictive terms or ‘binding words’, demonstrate associations with economic growth and productivity (McLaughlin, Strosko, and Jones 2019).

In collaboration with the Mercatus Center, we obtained quantitative metrics for all 876 national laws of India. For this purpose, we used the list of laws made available on the official [portal](#) of the Government of India¹. This empirical analysis, along with our categorisation of laws by the Ministry and Department, will help open way for further research on the burden imposed by laws.

Methodology

Below we give an overview of the three metrics studied. A detailed methodology for each of these metrics and the algorithms used, can be found on the [QuantGov website](#).

Word count

Word count per law: This metric quantifies the number of words in a law.

Word count per year: This metric documents the average number of words per law, in a particular year (from 1857-2019). It helps document the change in the volume of laws over the years.

Word count per Ministry: This metric quantifies the average number of words per law in each Ministry in India. The total number of words in legislations introduced by a Ministry, helps indicate the extent to which the sector is regulated and the effort it would require on part of the regulatees, to acquaint themselves with the law.

Binding words

Number of binding words per law: RegData uses a text analysis program to count the number of binding words or “restrictions” in a law, that create an obligation to comply or limit choice sets for private actors. These include phrases such as: ‘shall’, ‘must’, ‘may not’, ‘required’, and ‘prohibited’.

Normalised binding words: Binding words are likely to be higher in laws that are lengthy. To get an estimate of the density of restrictiveness, we also calculate the ‘normalised binding words’ for each law. This metric refers to the average number of words after which a binding word appears. For instance, if the normalised binding words for a law is 300, it means that on average a binding term appears after every 300 words in the law. Lower normalised binding words would imply that a law is more restrictive.

How did we collate binding words for national laws?

To identify terms that bind actions of individual actors in Indian laws, we adopted the following 5 step approach:

Step-I: Database used. We used the official list of laws available on the Ministry of Law and Justice [website](#). The website lists all national level laws enacted up till March 2020, in chronological order.

Step-II: Sampling process. We used the chronological list of laws to select a sample of 40 laws by using a two-stage systematic random sampling process (without replacement). In the first round,

1. This list was last updated in October 2020

we selected every 43rd law in the list. In the second round, we selected every 41st law in the list. This produced a sample of 40 laws (4.5% of the total laws listed).^{2x}

Step-III: Collate the universe of binding words. All 40 laws selected in the sample were studied in detail to identify terms that were used, at least once, in the ‘restrictive’ sense, i.e. to restrict the actions of individuals, companies or any non-government entity. All such terms were listed down, irrespective of their frequency. For instance, use of the term ‘cancel’ in Section 22(4) of the Banking Regulation Act 1949 is restrictive. The provision mentions that if any company fails to comply with the conditions mentioned in the Act, then the Reserve Bank of India will ‘cancel’ their licence.

Step-IV: Prepare a restriction counter for the sample. Once the universe of binding words was collated, we prepared a restriction counter for all 40 laws. Against each law, we recorded the number of times a particular binding word appeared. This aggregate number was broken down into two categories: restrictive and nonrestrictive. Under the ‘restrictive’ category, we recorded the number of occurrences of term in a context where it bound the actions of private actors. All other appearances of the term were recorded as non-restrictive (including ones that restrict the actions of the public officials, limit their scope of power and instill accountability). For instance, the term ‘ought to’ appears to be a binding word. However, the majority of its appearances are in the non-restrictive context. Under the Air (Prevention and Control of Pollution) Act 1981, ‘ought to’ is used in Section 5 in a non-restrictive context. The section prescribes the constitution of a State Board and requires the State Government to ascertain which interests ‘ought to’ be represented in the board. The term does not restrict actions of individuals/ private players.

Step-V: Threshold to pick the final set of terms. After recording the number of binding words for each law, we prepared aggregates of their usage in the restrictive and nonrestrictive context. That is, the total number of times a term appeared as a restriction (across the 40 laws) and the total number of times it appeared in the non-restrictive context. Words that appeared in a restrictive context in more than 80 per cent instances were chosen as binding words. Words that appeared in restrictive context in less than 80 percent instances and words that appeared only once across all 40 laws were removed from the final list of words chosen.

We chose the following 12 binding words: “prohibited”, “prohibition”, “prohibit”, “shall be punishable”, “shall be punished”, “imprisonment”, “fine”, “restrictions”, “shall be liable”, “cancel”, “impose”, and “guilty of”.

Linguistic complexity

This metric measures the complexity of a given law. Complexity is understood by how a law fares on the following four sub-categories:

1. *Shannon entropy* refers to the “likelihood of encountering new words and concepts in a given body of text” (McLaughlin et al. 2020). Although, higher Shannon entropy implies higher complexity of a given text it is difficult to make inferences from an entropy score in absolute terms. The entropy scores cannot be meaningfully aggregated. These scores are best understood in relative terms. See [here](#) for more information.
2. *Sentence length* measures the average or mean length of sentences in a law. A higher mean length may result in greater difficulty in reading and understanding a legislation.
3. *Conditional count* calculates the frequency of branching words in a legislation. This includes terms like “if”, “but”, and “provided” that represent logical branches in a law. As the conditional count increases, the difficulty of comprehending a law may also increase.
4. *Flesch Reading Ease score* measures the readability of a given piece of text based on the average length of sentences and the average number of syllables per word. Each law is assigned a score

2. Our aim in the first round was to collate a sample of 20 laws. For this purpose, we divided the 876 laws into 20 parts (i.e. picked every 43rd law). To ensure that our sample is representative, we later expanded the sample size to 40 laws and picked another 20 laws in the second round (i.e. every 41st law).

on a scale of 0-100 with 0 being the most difficult read and 100 being the easiest to read. The higher the Flesch score of a law, the easier it is to read a law. A Flesch reading score can also be negative.

These four metrics help provide an understanding of how easy or difficult it is to comprehend a law. A law that is tough to comprehend may also increase the compliance costs for regulated entities (in terms of effort, time and money)(McLaughlin et al. 2020).

Summary of Findings

Our interactive dashboards provide details of all three quantitative metrics across all national laws. Below we present a summary of the key findings.

Volume of national laws

The average number of words in a national law is 9602.3 words. Table 1 lists the top 10 national laws with the highest number of words.

| Name of law | Number of words |
|---|-----------------|
| The Income Tax Act, 1961 | 5,30,288 |
| Companies Act, 2013 | 1,77,733 |
| Code of Civil Procedure, 1908 | 1,76,971 |
| The Code of Criminal Procedure Act, 1973 | 1,30,630 |
| The Merchant Shipping Act, 1958 | 1,26,670 |
| The Delhi Municipal Corporation Act, 1957 | 98,462 |
| The New Delhi Municipal Council Act, 1994 | 83,695 |
| The Indian Penal Code, 1860 | 79,684 |
| The Customs Act, 1962 | 73,976 |
| Cantonments Act, 2006 | 69,943 |
| On average these 10 laws have 1,05,004 words. | |

The shortest law is the Supreme Court (Number of Judges) Act, 1956, with 117 words.

Ministry analysis: The top 5 Ministries with the most voluminous laws are: Ministry of Corporate Affairs (29,686 words on average); Ministry of Finance (15,069 words on average); Ministry of Shipping (13,694 words on average); Ministry of Environment, Forest and Climate Change (12,137 words on average) and Ministry of Commerce and Industry (11,787 words on average).

The Ministry of Culture has the least voluminous laws, with an average of 4229 words per law.

Temporal analysis: The average words per law have increased in the post-independence era from 16,264 to 18,537 words. Post-independence, the year 1973 reached a peak with an average of 1,30,630 words per law. This is India's highest average across the pre and post-independence era. In the last decade (2010-2020), the year 2013 had the highest average number of words per law (55,527 words). The year 2011 saw the lowest average number of words per law (3,608 words). Between 1947 and 2010, the lowest average number of words was seen in the year 1993 (4,279 words).

Restrictiveness of national laws

The average number of binding words per law is 19.3. The most commonly used term is “fine” with 4.66 uses per law, followed by “imprisonment” at 4.39 uses per law (on average). Apart from the Code of Criminal Procedure Act 1973 and the Indian Penal Code 1860, the two civil laws that have the highest occurrences of the term ‘fine’ and ‘imprisonment’ are: the Merchant Shipping Act, 1958 and Companies Act 2013.

The words “prohibition” and “prohibited” are used 1.97 times per law. The term “cancel” is the least commonly used binding word, at 0.45 uses per law.

To identify the most restrictive laws, we also calculated the normalised binding words for all laws.³ Table 2 lists the top 10 laws with the highest density of binding words.

3. This is measured as the total number of words/number of binding words.

| Name of law | Normalised binding words |
|---|--------------------------|
| White Phosphorus Matches Prohibition Act, 1913 | 38 |
| The Indian Penal Code, 1860 | 47 |
| Explosive Substances Act, 1908 | 66 |
| Prevention of Damage to Public Property Act, 1984 | 71 |
| Prohibition of Child Marriage Act, 2007 | 71 |
| Weapons of Mass Destruction and Their Delivery Systems (Prohibition of Unlawful Activities) Act, 2005 | 77 |
| The Calcutta Metro Railway (Operation and Maintenance) Act, 1985 | 80 |
| Lotteries (Regulation) Act 1998 | 83 |
| Legal Metrology Act, 2010 | 84 |
| State Emblems of India (Prohibition of Improper Use) Act, 2005 | 88 |
| On average, binding words occur after every 70.5 words in these 10 laws | |

Interestingly, laws like the Legal Metrology Act 2010 have a greater density of restrictive terms as compared to the Code of Criminal Procedure Act 1973.

Temporal analysis: While the density of restrictive terms in laws have reduced post-independence, the change is marginal. In the pre-independence era, on average, a restrictive term appeared after 265 words. Post-independence, on average, a restrictive term appears after 383 words. In the post-independence era, the year 1973 was the most restrictive. In this year, a restrictive term appeared after every 109 words (on average). The least restrictive year post-independence was 1961. In this year, a restrictive term appeared after every 1891 terms, on average

Linguistic complexity of national laws

Shannon entropy

The average Shannon score of all national laws studied is 8.13. The Easements (Extending Act 5 of 1882) 1891 has the lowest score of 5.32. The Andhra Pradesh Reorganisation Act 2014, has the highest score of 10.14. The average Shannon Entropy score (at the part level) in the Code of Federal Regulations (CFR) for 2019 is 7.86. Compositions by Shakespeare tend to have an entropy score between 9.3 and 9.7 (McLaughlin et al. 2020).

Ministry analysis: The Ministry of Labour and Employment records the highest average Shannon entropy score of 9.162. The Ministry of Personnel, Public Grievances and Pensions records the lowest average entropy score of 8.005. This is still higher than the average in the CFR.

Conditional count

On average, a national law has 57.07 conditionals. Table 3 lists the top 10 laws with the highest number of conditionals.

| Name of law | Conditional counts |
|--|--------------------|
| The Income Tax Act, 1961 | 3344 |
| Code of Civil Procedure, 1908 | 1664 |
| Companies Act, 2013 | 1525 |
| The Code of Criminal Procedure Act, 1973 | 1275 |
| The Merchant Shipping Act, 1958 | 1069 |
| Indian Succession Act, 1925 | 620 |
| The Delhi Municipal Corporation Act, 1957 | 610 |
| The Indian Penal Code, 1860 | 608 |
| Central Goods and Services Act, 2017 | 573 |
| The New Delhi Municipal Council Act, 1994 | 556 |
| On average, these 10 laws have 1184.4 conditionals | |

Flesch reading ease score

In sum, 608 laws fall in the ‘fairly difficult’, ‘difficult’ or ‘very difficult’ category. Only 60 laws are ‘easy’ or “fairly easy” to read. 211 laws are rated as ‘standard’. The Supreme Court (Number of Judges) Act, 1956 has the highest score of 88, and is ranked as the easiest to read. Table 4 lists the 10 most difficult to read laws

| Name of Law | Flesch reading score |
|---|----------------------|
| Sick Industrial Companies (Special Provisions) Repeal Act, 2004 | -1 |
| Calcutta High Court (Jurisdictional Limits) Act, 1919 | 2 |
| The Parel Investments And Trading Private Limited And Domestic Gas Private Limited (Taking Over Of Management) Repeal Act, 2005 | 4 |
| National Capital Territory of Delhi Laws (Special Provisions) Second Act, 2009 | 5 |
| Right to Information Act, 2005 | 10 |
| National Capital Territory of Delhi Laws (Special Provision) Act, 2011 | 18 |
| Sports Broadcasting Signals (Mandatory Sharing with Prasar Bharti) Act, 2007 | 19 |
| National Medical Commission Act 2019 | 21 |
| Mussalman Wakf Act 1923 | 23 |
| Provisions of the Panchayats (Extension to the Scheduled Areas) Act, 1996 | 23 |
| On average, these 10 laws have a flesch reading score of 9.7 | |

References

- McLaughlin, Patrick A, Kofi Ampaabeng, Ethan Greist, Jonathan Nelson, Thurston Powers, and Stephen Strosko. 2020. “RegData 3.2 User’s Guide.” *Mercatus Center, George Mason University*, https://new-website-files.s3.amazonaws.com/regdata3_2_users_guide.pdf.
- McLaughlin, Patrick A, Stephen Strosko, and Laura Jones. 2019. “RegData Canada: A Snapshot of Regulatory Restrictions in Canada’s Provinces.” *Mercatus Center, George Mason University, Regulatory Snapshot*.