

Using Computational Text Mining to Understand Public Priorities for Disability Policy Towards Children in Canadian National Consultations

Derrick L. Cogburn
American University
dcogburn@american.edu

Denise Keiko Shikako Thomas
McGill University
keiko.thomas@mcgill.ca

Jonathan Lai
McGill University
jonathan.lai3@mail.mcgill.ca

Abstract

Identifying policy preferences from public consultations presents a challenge to national and local governments. Computational text mining approaches provide a useful strategy for analyzing the large-scale textual data emerging from these policy processes. In this study, we developed an inductive and deductive text mining approach to understand disability-related policy priorities. This approach is applied to data from the nationwide disability policy consultation conducted in 2016 by the Government of Canada. This process included 18 town hall meetings, 9 thematic roundtables, and online submissions from 92 stakeholders. Transcripts of these consultations were made available to researchers. Three broad research questions were asked of this data, focused on key themes; differences by city size and type of consultation; and impact of two global policy frameworks. The study identified a number of key themes and saw differences by city size. The study identified content related to both the CRPD and CRC.

1. Introduction

National and local governments around the world have turned increasingly to using public consultations to inform complex policy-making processes. Many of these consultations are being conducted with electronic collaboration tools or with closed captioning-based transcripts creating a text-based record of the contributions and deliberations. However, while these practices are increasing, accurately discerning public sentiment and public policy preferences from the participants remains difficult. Fortunately, this is a data analytic problem well suited for text mining approaches.

In this study, we developed an inductive and deductive text mining approach designed to exploit this rich source of textual data. We then tested the approach by analyzing the complex disability-related policy priorities within a specific public consultation.

The disability policy domain is extremely complex, especially as it relates to children with disabilities. In this context, local and national policy is conditioned by and engages with global treaty obligations. Once ratified, these international treaty commitments require the “State Party” to enact implementing legislation, so that the national legal environment is aligned fully with the country’s international obligations. In the case of disability policy, there are two inter-related international treaties, the United Nations Convention on the Rights of Persons with Disabilities (CRPD) and the UN Convention on the Rights of the Child (CRC).

For example, since 2010 the Government of Canada has been a state party to the United Nations Convention on the Rights of Persons with Disabilities (CRPD) and in 2017 it acceded to the Optional Protocol. After years of public and civil society consultations, in June 2018, the Canadian government tabled its proposed federal accessibility legislation in the House of Commons. Though *an Act to ensure a barrier-free Canada*, also called the *Accessible Canada Act* does not make any explicit mention of the CRPD, its objective is “to enhance the full and equal participation of all persons, especially persons with disabilities, in society”.

As part of its process to develop a national law for accessibility, the Minister of Public Services and Procurement and Accessibility, the Honourable Carla Qualtrough, launched a nationwide public process to assess the perspectives held by community members regarding policies to support persons with disabilities. This process included 18 town hall meetings across the country and 9 thematic roundtables with experts in areas including transportation, built environment, and employment. Transcripts of these consultations were made available to the research team along with online submissions from 92 stakeholder organizations.

The purpose of this study is to better understand the ways in which computational text mining can assist governments in exploiting the rich data textual emerging from public policy consultations. A secondary purpose is to determine the specific disability-related public policy priorities of participants in the national

consultation process in Canada, particularly as they relate to children with disabilities. Finally, we want to understand the impact of the CRPD and CRC on these deliberations. Given the complexity of this policy domain, it is highly likely this approach is adaptable to other complex policy domains.

2. Analytical Approach

Our analytical approach is driven by three main dynamics, covered extensively in the literature on deliberative dialogues and e-government. These dynamics include: 1) public engagement in public policy, 2) public engagement in human rights frameworks, especially the UN Convention on the Rights of Persons with Disabilities and UN Convention on the Rights of Children, and 3) An Evidence-informed framework for policymaking [8]. From the public engagement in public policy we adopt the understanding of community participation as an important vehicle for enhancing public policy deliberation [1][2][3][4][5][6][7][8]. When these dialogues are captured via transcripts, they become an invaluable source of data for analyzing these processes using text mining approaches. From the human rights frameworks, we understand that international treaties are one way of aggregating global consensus around key policy areas. As two of the leading global policy frameworks in this emerging policy environment, it is possible the CRPD and the CRC have influenced the perspectives of these participants and capturing participants' discourses may also inform local and international development and implementation of these treaties. From the evidence-informed policymaking framework, we adopt the understanding that research information moves in a cyclical way from questions and problems identified by the public to the adaptation to local contexts and application into larger scales of public policy. Figure 1. suggests a cycle moving from public engagement to public policy changes and the steps involving data mining and knowledge translation contribute to more inclusive policies.

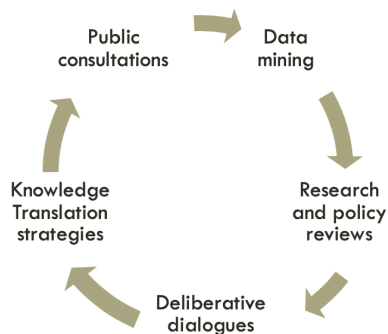


Figure 1. Overview of Analytical Approach

3. Research questions

Based on this analytical approach we have identified three overarching or “grand tour” research questions that drive this study, each with several subsidiary and operationalized questions.

1. What are the key elements (i.e., terms, phrases, topics, and named entities) that occurred within the public consultations?
 - a. RQ1.1 What are the key terms across the entire dataset?
 - b. RQ1.2 What are the key phrases across the entire dataset?
 - c. RQ1.3 What topics are identified across the entire dataset?
 - d. RQ1.4 What named entities are identified across the entire dataset?
2. To what degree to these elements differ by city size (i.e., small, medium, large) and type of consultation (i.e. public consultation, stakeholder forum)
 - a. RQ2.1 How do key terms in the dataset differ by city size?
 - b. RQ2.2 How do key terms in the dataset differ by consultation type?
 - c. RQ2.3 How do key phrases in the dataset differ by city size?
 - d. RQ2.4 How do key phrases in the dataset differ by consultation type?
 - e. RQ2.5 How do topics in the dataset differ by city size?
 - f. RQ2.6 How do topics in the dataset differ by consultation type?
3. What elements of the international treaties related to children with disabilities (i.e., UN Convention on the Rights of Persons with Disabilities (CRPD) and UN Convention on the Rights of the Child (CRC)) are reflected in the public consultation?
 - a. RQ3.1 Which international treaty, CRPD or CRC, is reflected more in the dataset?
 - b. RQ3.2 What elements in the CRPD are included in the dataset?
 - c. RQ3.3 What elements of the CRC are included in the dataset?
 - d. RQ3.4 How does the representation of the CRC and CRPD differ by city size?
 - e. RQ3.5 How does the representation of the CRC and CRPD differ by consultation type?

4. Methodology

This project uses a statistical “bag of words” text mining methodology. It also uses the Cross-Industry Standard Protocol for Data Mining (CRISP-DM)

approach to organize our work. CRISP-DM structures text mining projects into six phases.

In the initial phase, the focus of the project is on clearly articulating the text mining problem you are trying to address. This stage is addressed above as we discussed the purpose of this study.

This stage is followed by identifying and understanding the available data. The data for this study are transcripts of the public consultation process, and will be discussed in the next section.

In the next stage, we prepare the data for text mining, using a variety of preprocessing techniques, including applying a standard English exclusion list. Some of the other data cleaning techniques include customizing the exclusion list by adding terms that may be artifacts of the data collection process and may not send any signal as to the policy content of the consultations. As we analyzed the data, we also combined words that were referencing the same concept but were written differently such as “People with Disabilities” and “Persons with Disabilities”.

After the data are cleaned and prepared, the next stage of the CRISP-DM is to develop and apply a number of analytical approaches and models. These approaches include a range of inductive approaches (e.g. keyword frequency analysis, phrase frequency analysis, named entity extraction, and topic modeling). For several of these techniques, we explore the data further by applying a measure of significance called term frequency by inverted document frequency (TF*IDF). This is a fairly standard text mining technique that helps to measure “importance” of terms and phrases within a text corpus, by highlighting words that occur frequently in the dataset, but minimizing or penalize those terms that occur in too many documents. When using TF*IDF, the higher the weighting, the more unique and “important” the term. One other technique we will use is to explore the data using word “frequency”, but in order to attempt to control for the size of each document, we will normalize the data by using “rate per 10,000 words”. This approach makes the frequency analysis less sensitive to the size of each individual document. The penultimate stage of the CRISP-DM is to evaluate our models. We do that in our section on findings, where we attempt to answer all our research questions based on this analysis. Finally, stage six focuses on deployment, which we do through a closing section on recommendations.

5. Data collection

The data for the project were provided to the research team by the Ministry of Public Services and Procurement and Accessibility. These data were saved

in 83 text files, and were separated into six categories: (1) government summaries (n=4); (2) forum transcripts (n=17); (3) forum summaries (n=18); (4) roundtable transcripts (n=9); (5) roundtable summaries (n=8); and (6) online submissions (n=27). summarized in Table 1.

Table 1. Overview of Data for the Study

	VALUE	FREQUENCY	TOTAL PERCENT	VALID PERCENT
Government Summary	4		4.8%	4.8%
Forum Summary	18		21.7%	21.7%
Forum Transcript	17		20.5%	20.5%
Roundtable Summary	8		9.6%	9.6%
Roundtable Transcript	9		10.8%	10.8%
Online Submission	27		32.5%	32.5%

These files were important into WordStat 8, a commercial text mining tool to create the corpus for the study (<http://provalisresearch.com>). Once the corpus was created, we added three key variables to each file, which were: (1) document type, (2) meeting location, and a variable called (3) city-size (which was based on public data on the population of that locality. These variables were added to aid in our crosstabulation and comparative analysis.

Before beginning data analysis, we first preprocessed and cleaned the data. These preprocessing steps included applying a standard English exclusion list (stopwords) and customized it by adding stopwords unique to this dataset (this includes words and phrases that are artifacts of the data collection and transcription process). For this analysis, the researchers also decided to filter the data to remove the “summaries” so that only raw data was included. We also made a decision to include cases/documents that included missing data on any of our variables. This filtering resulted in 53 documents being included in this analysis. No stemming, lemmatization, or other pre-processing steps were taken at this time.

6. Findings

While the research team continues to refine the analytical strategy for this data, we have produced some initial analyses that may be of interest to a variety of audiences, including: academic, community, government, and policy analysts. These findings are organized along the lines of our initial grand tour and operational research questions, which begin with descriptive inductive questions, and then move to specific deductive questions.

6.1 Inductive Analysis

We begin our inductive analysis by a descriptive keyword frequency analysis across the entire dataset. This will give us an initial understanding of our data, similar to a descriptive analysis of variable frequencies and measures of central tendency in statistical analysis.

6.1.1. Keyword frequency analysis. Keyword frequency analysis is one of the simplest text mining techniques, but it can still be extremely powerful, and shed substantial light on the nature of our data. This is particularly true when we combine it with a cross tabular analysis of the additional variables we have added to the documents in our dataset. Our first broad research question asks: What are the key elements (i.e., terms, phrases, topics, and named entities) that occurred within the public consultations?

RQ1.1 What are the key terms in the entire dataset? When we look across the entire dataset, we find certain keywords immediately emerge to highlight the diverse nature of this dataset and its focus on persons with disabilities. In order to demonstrate the impact of using TF*IDF, Figure 2. below is a wordcloud and Figure 3. is a barchart highlighting the top 20 terms.



Figure 2. Wordcloud of Top 20 Terms

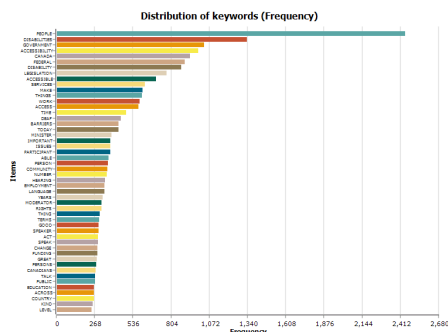


Figure 3. Barchart of Top 20 Terms Using TF*IDF

These keywords are what we would expect, and if we did not see something like these keywords, we would know we probably had a different dataset than what we

were expecting. Even so, you can start to see some priorities in the dataset, including accessible, services, barriers, employment, funding, education, legislation, and several others.

RQ1.2 What are the key phrases in the entire dataset? **6.1.2. Phrase frequency analysis.** Phrases frequency analysis can start to tell us more about the dataset. We have set our analysis to include 2-5-word phrases (NB: in order to capture if people are referring to the CRPD and CRC, we should probably change that to 2-7 word phrases). The addition of phrases to our frequency analysis, what are called “n-grams” in text mining and data science, helps add more specificity to where we begin with the keyword frequency analysis. We can begin to see key themes in the dataset more clearly as shown in Table 1.

Table 2. Top 20 Phrases in the Dataset

Without Using TF*IDF	Using TF*IDF
People with Disabilities	Situation de Handicap
Persons with Disabilities	Personnes en Situation de Handicap
Federal Government	Personnes Handicapees
Human Rights	People with Disabilities
Sign Language	Women with Disabilities
Accessibility Legislation	Sexual Health
Deaf People	Sexual and Reproductive Health
Minister Qualtrough	Deaf People
Situation de Handicap	Persons with Disabilities
Canadians with Disabilities	Sign Language
Government of Canada	Deaf Blind
Across the Country	Culture Change
Hard of Hearing	Minister Qualtrough
Mental Health	Social Procurement
Personnes Handicapees	Federal Government
Disabilities Act	Developmental Disabilities
Across Canada	Hard of Hearing
Deaf Blind	Mental Health
Built Environment	Etats Unis
Federal Accessibility Legislation	People with Developmental Disabilities

RQ1.3 What topics are identified in the entire dataset? Another useful inductive or exploratory text mining technique is called topic modeling. Topic modeling may be implemented in a number of different ways. One of the most popular is called Latent Dirichlet Analysis (LDA). In Wordstat 8, the Provalis Research Corporation has implemented topic modeling using a variant of statistical factor analysis. With this approach,

words and phrases that have an underlying statistical relationship cluster together on a “factor” or in this case a “topic”. The researcher can then examine the grouping of words to try to interpret what the factor/topic represents. In Wordstat, the software even suggests a word or phrase that might represent this factor/topic. This approach has been shown to be equal to or better than LDA in recent research papers.

Once we cleaned the topic model we had nine coherent topics in our dataset. These topics in order of “coherence” are as follows: (1) Situation of Persons with Disabilities; (2) Sign Language; (3) Mental Health; (4) Federal Government; (5) Built Environment/ Building Code; (6) Minister Qualtrough; (7) Tax Credit/ Income; (8) Barriers for Persons with Disabilities; (9) People with Disabilities. When we cross-tabulate these topics with key variables such as “location” or “document type” we are able to determine if these topics are clustered across our dataset, or if they were more prominent in on location or type of document. For example Figure 4 below summaries our nine topics cross-tabulated with the location of the meeting that produced that document. For several of our topics, such as Federal Government, built environment, and people with disabilities, we see that topic being fairly evenly distributed across the locations. However, for some key topics, such as “sign language”, which is illustrated in the figure, we can see it predominantly linked to the online submissions. Similarly, “Mental Health” is clearly influenced by one online submission. Other examples not shown, were “Built Environment” and location online submissions. Alternatively, the topic “Minister Qualtrough” was not mentioned as frequently in the online submissions. Figure 4. presents our topic model by location.

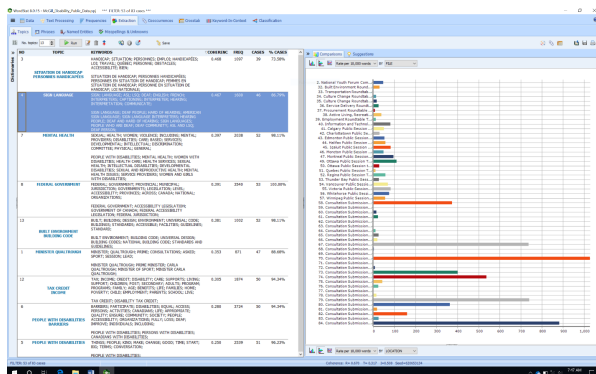


Figure 4. Topic Model by Location

6.1.3. Named Entity Recognition. Another technique for exploring the dataset is called “Named Entity Recognition” NER. With this technique, we can identify specific named entities within the dataset (e.g. persons, places, organizations, locations), and their frequency.

Some of these might be expected, such as Canada and Government of Canada. Others, might be surprising, such as: Minister Qualtrough (n=64); British Columbia (n=36), United States (n=25), and CRPD (n=24). Table 3 illustrates top named entities.

Table 3. Top Named Entities in the Dataset

ENTITY	TOTAL	UNIQUE
Canada	1225	880
Canadians	405	382
Canadian	309	183
Ontario	157	126
Qualtrough	115	30
Québec	113	110
Government of Canada	89	79
Federal Government	86	82
LL	83	56
Alberta	74	42
English	72	72
ASL	67	67
Disabilities Act	66	58
CRTC	65	61
Minister Qualtrough	64	64
BC	60	49
French	59	59
LSQ	57	56
DOESN	54	54
DIDN	52	51
MANITOBA	51	45
Ottawa	46	38
American	46	14
Canadiens	44	44
British	44	4
Edmonton	43	40
Columbia	43	4
REEI	41	41
Deaf-Blind	41	41
Toronto	41	36
Rights of Persons	40	38
Saskatchewan	37	34
Lauzon	37	28
British Columbia	36	33
États-Unis	35	35
Prime Minister	35	30
CCD	34	32
DTC	34	29
WASN	33	33
Calgary	33	28
Isn	32	28
Winnipeg	31	29
Americans	30	30
AODA	30	29
Quebec	29	25
Loi	26	26
United States	25	22
VRS	24	24
CRPD	24	23
American Sign Language	23	23

6.1.4. Hierarchical Cluster Analysis. Another way of looking for themes within the data is to do a hierarchical cluster analysis. This approach looks for words that are substantially overlapping each other, and these clusters helps to identify some pretty interesting insights into key concepts in the data. Figure x below is a small

portion of the dendrogram we produced, and although these clusters have to be interpreted by the researchers, it already indicates some very interesting thematic areas in the data. The basic way to interpret this figure is essentially, the closer together the words are the stronger their “thematic connection”. As you move farther to the right, you see broader thematic clusters. It is a very powerful technique.

For example: these clusters from top to bottom seem to be addressing (1) the role of technological abilities in the world; (2) the consultation was producing a positive conversation; (3) access services to enhance persons with disabilities access to, information and transportation, education and employment, and the role of accessibility legislation; (4) something about across the country and the provinces; (5) the necessity to provide funding and support programs for persons with disabilities; (6) jurisdiction leadership role?; (7) engagement/importance of various levels of government, including both municipal and provincial; (8) the need to improve national organizations; and (9) the CRPD as a human rights treaty. Figure 5. Presents a portion of our hierarchical cluster analysis.

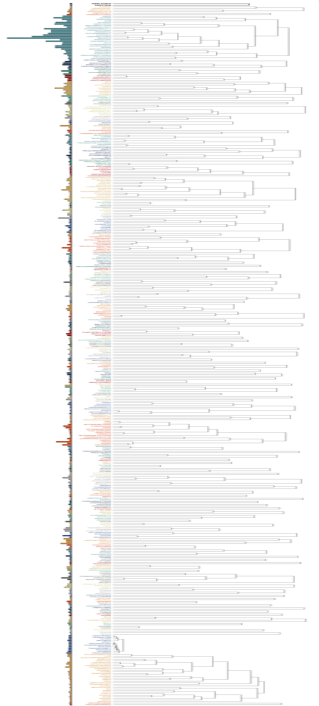


Figure 5. Sample of hierarchical Cluster Analysis

To what degree to these elements differ by city size (i.e., small, medium, large) and type of consultation (i.e. public consultation, stakeholder forum)

RQ2.1a How do key terms in the dataset differ by city-size? Figure 6. below illustrates how key terms such as autism, deaf, education, and employment are distributed across the dataset by city size.

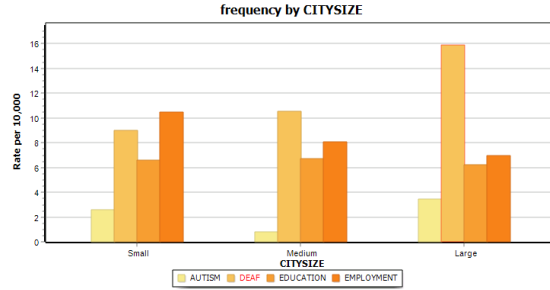


Figure 6. Key Terms by City Size

Several key phrases have distinct distributions by size of city for the consultations. For example, the phrase “Deaf Blind” is almost exclusively used in the large cities as shown below in Figure 7.

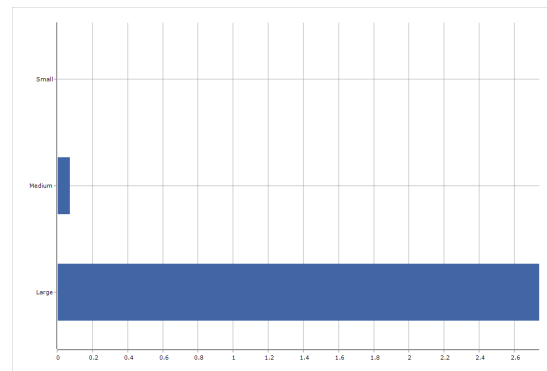


Figure 7. Phrase Deaf-Blind by City Size

In contrast, Figure 8. shows how the phrase intellectual disabilities is mentioned almost exclusively in small towns, and almost never in large cities.

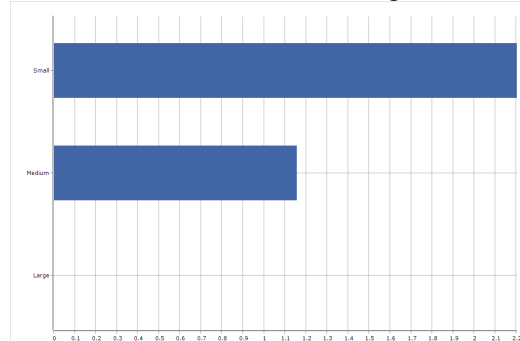


Figure 8. Intellectual Disabilities by City Size

RQ2.1 How do key terms in the dataset differ by consultation location? When we look at the keyword frequency analysis cross-tabulated by the consultation location variable, there are some interesting findings. For example, the words “Health” and “Education” occur far more frequently in one online submission 78 than they do anywhere else in the dataset. Figure 9. below illustrates this crosstabulation.

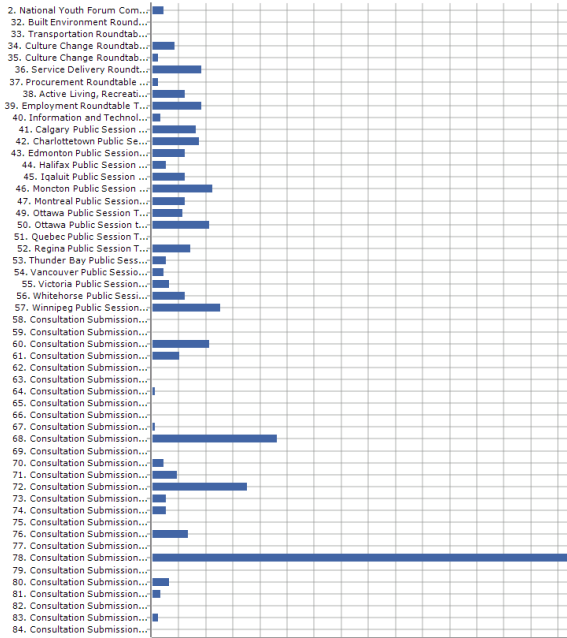


Figure 9. Keyword Health by Type

RQ2.2 How do key terms in the dataset differ by consultation type? For example, Figure 10. illustrates how the keyword “deaf” was mentioned far more frequently in the online submissions than in the other two types.

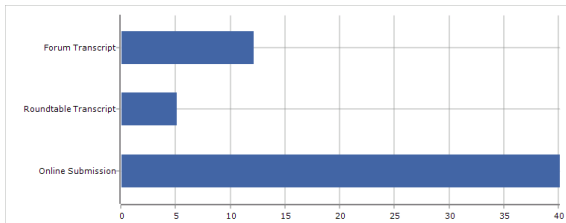


Figure 10. Keyword Deaf by type

Similarly, Figure 11. shows that employment is also mentioned more frequently in the online submissions.

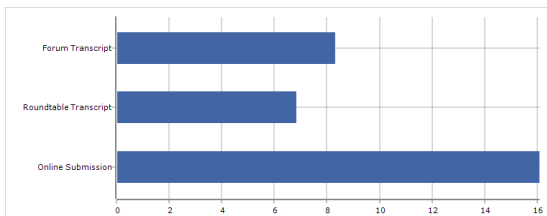


Figure 11. Keyword Employment by Type

RQ2.4 How do key phrases in the dataset differ by consultation type? When analyzing how key phrases in

the dataset differ by consultation type, we find several interesting examples. For example, Figures 12 illustrates how the phrases “Deaf People” is mentioned far more frequently in the online submissions, than in the forums or roundtables.

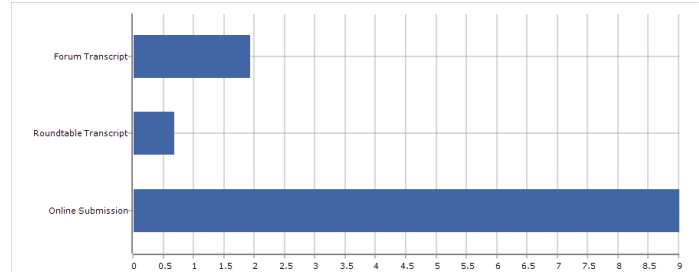


Figure 12. Phrase Deaf People by Type

Another interesting example is the phrase “women with disabilities” which Figure 13. Shows is mentioned disproportionality in the online submissions, and almost never in the forums or roundtables.

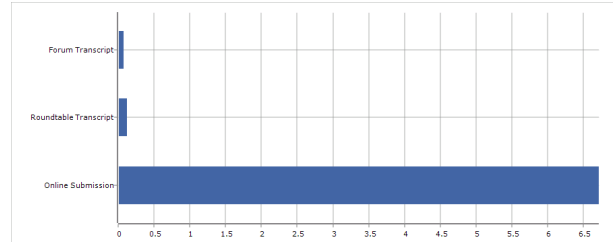


Figure 13. Phrase Women with Disabilities by Type

In contrast, Figure 14. shows how the phrase “Built Environment” is mentioned more frequently in the roundtables than in other document types.

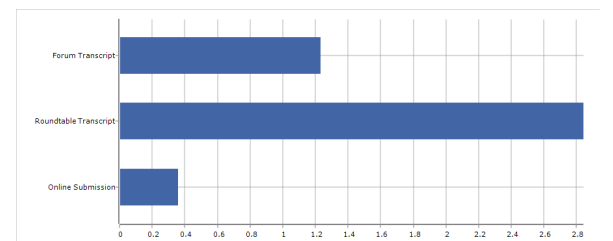


Figure 14. Phrase Built Environment by Type

Sign language was only mentioned in large cities. RQ2.5 How do topics in the dataset differ by city size? When analyzing topics by city size, we see in Figure 15. that “sign language” is relatively evenly distributed, with a slight increase in large cities.

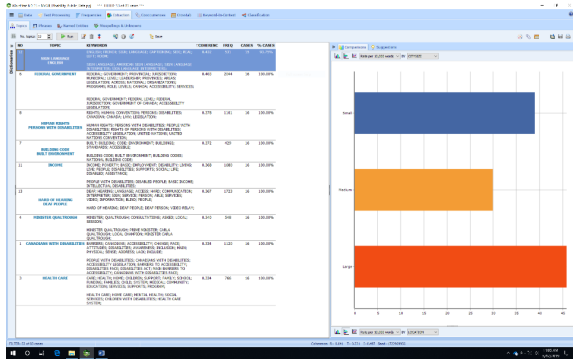


Figure 15. Sign Language Topic by City Size

However, in Figure 16. we see the “local champions” topic seems to be heavily focused on the small towns.

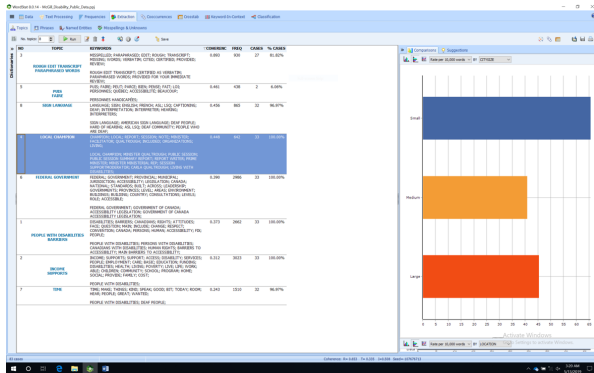


Figure 16. Local Champions Topic by City Size

However, when we look at topics crosstabulated by city size, all nine topics are evenly distributed. For example, in Figure 17 we see built environment by city size all showing a relatively even distribution.

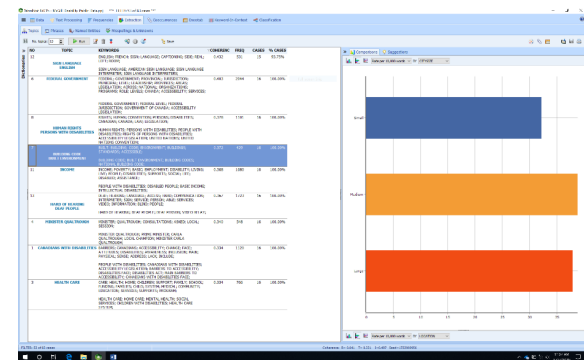


Figure 17. Built Environment by City Size

RQ2.6 How do topics in the dataset differ by consultation type? When we look at topics in the dataset by consultation type, Figure 18 education, human rights and sign language.

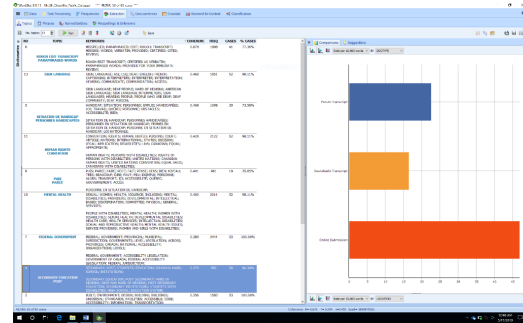


Figure 18. Education by Consultation Type

6.2. Deductive Analysis

In our deductive analysis, we developed two concepts, a priori, and assessed the extent to which they are present in the dataset. These concepts are related to two international treaties related to children with disabilities, specifically the CRPD and the CRC. To what degree are they reflected in the consultation?

6.2.1. Overview of CRPD and CRC Dictionaries.

In order to assess the degree to which elements of the CRPD and CRC are reflected in the consultation, two detailed, multi-level categorization models (or dictionaries) were developed for each treaty. In each instance, the research team started with the specific articles we were interested in as the highest functional categories (the highest category was CRPD and CRC to allow us to compare the two treaties in aggregate form). Figure 19. illustrates a portion of our model.

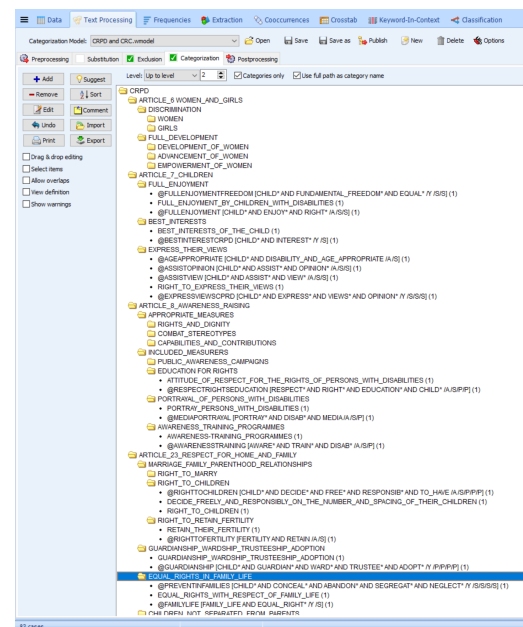


Figure 19. Sign Language Topic by City Size

Then, below the level of the specific articles, we looked at the focus of each specific paragraphs and sub-paragraphs and made that a sub-category (and sub-sub categories). Once all the categories and sub-categories were exhausted, we included specific words and phrases that would indicate a discussion about those themes. Then, we built “proximity rules”, including liberal use of wildcard searches that were intended to capture multiple variations of the word (e.g., disab* to capture, word variations such as: disabled, disabilities, etc.). These searches could contain up to four levels of association between a target word, phrase, or category and these search terms, and be limited or expanded to the proximity of the sentence, paragraph, or document. In Figure 20, elements of the CRPD are found more frequently in the dataset than the CRC.

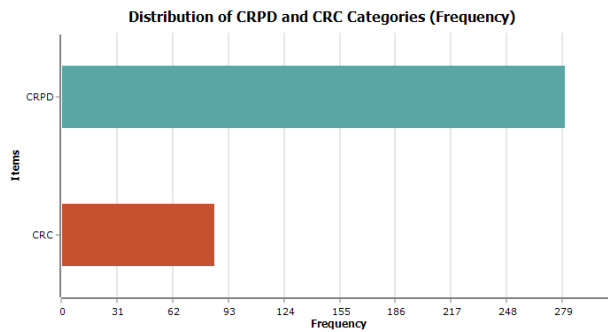


Figure 20. Comparing the CRPD to the CRC

Also, Figure 21, shows that the CRC is discussed more frequently in large cities; while the CRPD is discussed far more frequently in medium-sized cities.

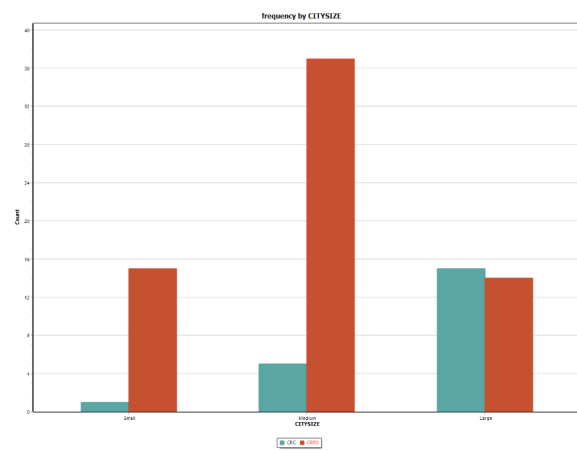


Figure 21. Comparing CRPD and CRC by City Size

And Figure 22, Illustrates the distribution by Document Type, showing both the CRPD and CRC occurring more frequently in online submissions.

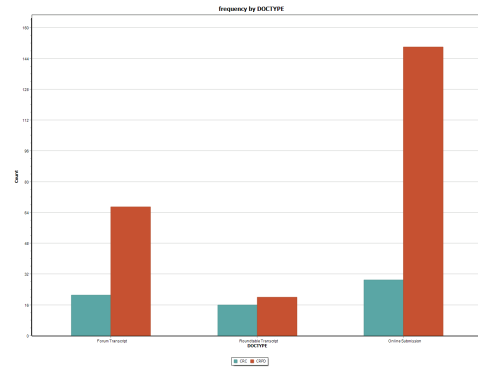


Figure 22. Comparing CRPD and CRC by Type

Figure 23 illustrates how the Montreal public session talked more about the CRC than the CRPD, and in fact contributed more discussion of the CRC than any other session (followed by the Active Living Recreation and Sport Roundtable). Also, the Moncton public session contributed more discussion of the CRPD than any other consultation location. Several sessions had discussions of one treaty and not the other. For example, the Victoria and Iqulquit public sessions had a fairly substantial discussion of the CRPD, but no discussion of the CRC. In contrast, the Active Living Recreation and Sport Roundtable had discussed the CRC but had no discussion of the CRPD.

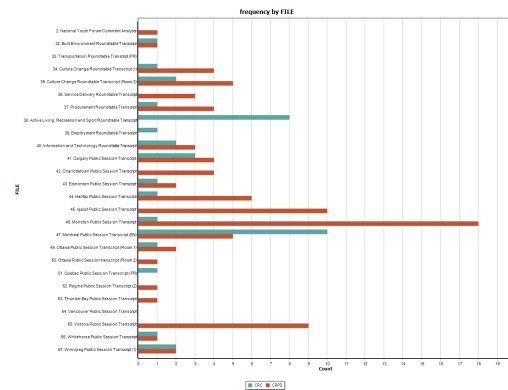


Figure 23. Crosstab of Location by CRPD and CRC

Figure 27 illustrates CRC Article 31 on Leisure, Play, and Culture receiving the most discussion.

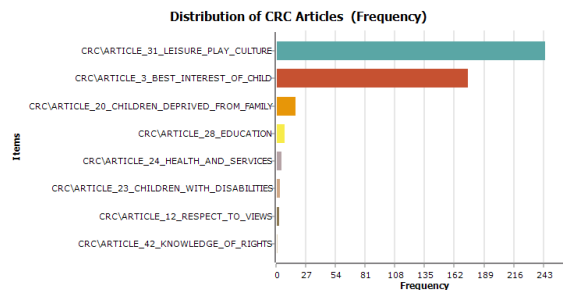


Figure 24. Distribution of CRC Articles

Similarly, Figure 25 shows CRPD Article 30 receiving the most discussion. In contrast, whereas the CRC Article 28 on Education was the least referenced, CRPD Article 24 on Education was second most.

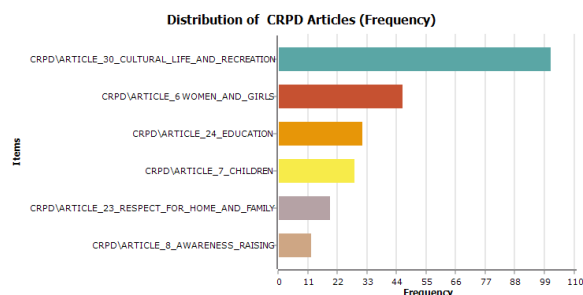


Figure 25. Distribution of CRPD Articles

7. Discussion

The findings of this study strengthen our understanding of the overall disability and child-related public policy priorities in Canada. We have learned how these priorities differ by city size, but also by consultation type.

However, the analytical structure proposed here has the potential to inform several directions in understanding public discourse and the presence of human-rights language in public discourse. The analytical structure we propose can also inform analysis of an extensive body of text such as public policies in many countries and establish a multitude of analysis and comparisons locally and across countries. For instance: congruences and dissonances between human rights treaties (e.g. the CRC and the CRPD) in countries that are signatories of both.

This analytical model can also inform the analysis of public discourse and its application into actual legislation, policy guidelines, acts and practical policy changes in one country. The use of international frameworks as the bases of the analytical structure allows also for comparison between countries of similar socio-economic and political status and provide insight into comparisons among countries ranging in different socio-economic development status. Results from this project will inform the sentiment and knowledge of the Canadian public in relation to the CRC and the CRPD, informing possible awareness-raising campaigns, monitoring, reporting, and complaint mechanisms for these treaties.

Further, these models will be able to use globally and longitudinally to track developments in the implementation of these treaties. Moreover, using these methods to understanding text-heavy data sources, such as the public discourse in any field, and

comparing those data to various standards will be useful in various disciplines. For instance, in healthcare, understanding adherence to clinical guidelines and reasons for clinically unwarranted variations in care can be explored by mining medical records. Overall, these tools provide a way to capture patterns compared to a “gold standard” (the model), allowing for ways to synthesize any large repertoire once built.

8. Limitations

While extremely valuable, there are a number of limitations inherent to a study like this. Some limitations of this study are: (1) it was conducted in one policy domain related to the process of one piece of legislation in Canada. Therefore, findings may not be applicable to other policy areas and in other countries. Other government may not have a similar process of public consultation and allowed for access to the data. (2) the models were developed by three content experts and are limited as such. However, subject-matter experts are seen as legitimate sources in this domain.

9. References

- [1] Degeling, C., Carter, S. M. & Rychetnik, L., «Which public and why deliberate? – A scoping review of public deliberation in public health and health policy research», *Social Science & Medicine* 131 (2015) pp. 1-10.
- [2] Jones, C.D., A.B. Smith, and E.F. Roberts,
- [3] J. Street, K. Duszynski, S. Krawczyk, A. Braunack-Mayer, “The use of citizens’ juries in health policy decision-making: A systematic review”, *Social Science & Medicine*, Elsevier, 109 (2014), 1-9.
- [4] C. de Freitas, “Public and patient participation in health policy, care and research”, *Porto Biomedical Journal*, 2017; 2(2):31-32.
- [5] C. Degeing, L. Eychetnik, J. Street, R. Thomas, SM., Carter “Influencing health policy through public deliberation: Lessons learned from two decades of Citizen’s/community juries. *Social Science & Medicine*, 179 (2017) 166-171.
- [6] S.M. Meffert, C.E. McCulloch, T.C., Neylan, M.Ghandi, C. Lund, “Increase of perceived frequency of neighborhood domestic violence is associated with increase of women’s depression symptoms in a nationally representative longitudinal study in South Africa, *Social Science and Medicine* 131 (2015) 89-97.
- [7] K. K., LI, J. Abelson, M. Giacomini, D. Contandriopoulos, “Conceptualization the use of public involvement in health policy-decision-making”, *Social Science & Medicine* 138 (2015) 14-21.
- [8] Bowen S, Zwi AB (2005) Pathways to “Evidence-Informed” Policy and Practice: A Framework for Action. *PLoS Med* 2(7): e166. <https://doi.org/10.1371/journal.pmed.00201>