**Paradata**

**Introduction**

At the beginning of the project process I had a lot of different ideas on what I wanted to do. At first I wanted to do this project on the topic of war. Specifically, I wanted to see how many times in all the editions the frequency that war was mentioned, and to see which decades were the most popular for referencing war in the newspaper. I then ran the idea past Dr. Graham and he recommended that I do something smaller and more particular. He said that I should maybe do something like the Korean War. I decided that may be a good idea so I then downloaded all the editions of the Shawville Equity from the years 1950-1953, which are the years that the Korean War took place in. I then decided to try and figure out a way to clean the data so that I only had the articles where the Korean War was mentioned. First I tried using the Stanford Ner to find this information, but I had no luck. I was unable to use the program to get the data that I wanted. I was unable to figure out the code to put in so that it would give me just the frequency of the term Korean War being used. When following the steps from the workbook I wasn’t even able to get it to show the location, names, etc that we learn to do in the module. I then decided to change my topic. I then wanted to use the TEI to clean up an issue that had the Korean War in it, but once again I struggled to get it to work. In the actual week that we did that exercise I had a lot of the trouble with the TEI and once again I wasn’t able to get my edition of the equity into the machine. The next step I took was to use RegExr. This is where I finally found some success.

I first put a random edition of the Equity into the RegExr and made an expression that would show how many times “Korean War” was said. I was unable to get any hits from it. So I tried another edition and once again was unable to get any hits. So I decided to look through the edition for something that I found interesting. That’s when I saw cities in Ontario being referenced. I thought that it would interesting to see how many times that Ontario is being referenced to compared to how often Quebec is referenced. I knew that the paper was out of Quebec, but I wanted to see how much influence Ontario had on the news that was being published in Quebec. I thought that there would be much more influence from Ontario then there would be from Quebec. This is because I felt as though Ontario has most of the mainstream news coming out of it and it would be shared more frequently, even in Quebec, than news that is Quebec specific. My key questions would be; Did Ontario have more of a media influence on the Shawville Equity then Quebec media would, and if so, by how much more was Ontario referenced than Quebec?

**Methods**

As mentioned early, I decided to use RegExr to find out how many times Quebec and Ontario were referenced in random articles from the Equity. I then had to create an expression for how to get the number of hits for this provinces. I was having trouble to find a way to create an expression but was able to come up with one. The expression for Quebec was (Quebe)\w+. This code looks incomplete as the ‘c’ from the end of Quebec is not in the code. This is because after messing around with the expression for quite a while, I was unable to make it search up the term Quebec unless the ‘c’ was missing at the end. I am sure there was another way to do this expression, but this way actually worked for me, and gave me the result that I needed as only Quebec came up as hits in the article. I then did the same for Ontario, using the code (Ontari)\w+. Once again the word is missing its last letter, but it was the only way to get the result I wanted. At first I only did the years 1950-1953 and recorded the hits for both Quebec and Ontario for each year. I decided to do the rest of the decade to get a higher number to look at. After I had gathered the numbers I had to find a way to put them into a visualization. I was looking through the old modules and came across RawGraphs. This tool gave me a way to show the data in a graph to see the results of the data in a much better way. I then had to put the data into Excel so that I could put the information into the RawGraphs website. I had to do it in two separate graphs, so I had to make two different excel sheets. I then made the two graphs and had them to compare.

**Results**

The results did not match up with what I believed was going to happen. The first year I did was 1953 and this was the only year that the edition actually matched up with what I believed. This edition had 31 for Ontario and 20 for Quebec. Other than this year, every other year was in favour of Quebec. In total there was 177 for Quebec and 108 for Montreal. This does not match up with my thesis that Ontario would have more of an influence on Quebec media than Quebec would have. The graphs can be referenced in my workbook from part 1. There could be a couple reasons for the results. The first thing that could have happened is that major cities such as Toronto or Ottawa could have been referenced, but without Ontario put behind them. I think this is unlikely because from what I read, anytime a city was referenced, the province with be referenced right there after it. Another thing that could of affected it is the editions that I picked. I find this highly unlikely however, because I looked at one edition from each year in a whole decade, and only one of the years Ontario was referenced more than Quebec. The key questions would be answered by the fact that Quebec was mentioned significantly more than Ontario was through the 10-year period. This means that Ontario did not have a bigger media influence on Quebec than Quebec influenced their own media.

**Conclusion/Growth over the Semester**

This class has been quite the challenge for me. Going into it I believed that I was pretty good with technology and understanding new techniques. However, I was very wrong. When I would do the modules it was like reading instructions in a completely different language to me. It became very frustrating and I was losing a lot of motivation. I ended up asking for help and the help that I received from Dr. Graham and fellow classmate Jeff helped me to work through most of the week’s exercises with a lot less trouble than before. At first my blog posts and fail logs were very weak compared to the standards they should have been at. My fail logs became much better as the semester went on, but my blog posts were still weak until my final one which I believe I did my best work on. My final project, while it probably will not be the best project in the whole class, is still a good effort for my ability to use these tools. I tried to work through a lot of the tools with no luck. While my project is very basic, I was still able to get the information I wanted to find, and be able to show it in some sort of graphic that helps to accompany the data I found. If I was able to do this class again then I would probably have asked for more help, then I already did when I needed help instead of just trying to work through it myself. This class has taught me a lot about different techniques that can be used for finding and cleaning data. While I may never have to use them again, it has been a very good learning experience and if I need to use them then I should be equipped to try and work through it. Even though I felt as though I struggled with a lot of the class, I feel as though I worked hard through the modules, and am able to use the tools and techniques to a basic degree. Overall this class has been a great work experience and I feel as though I was grown as a digital historian.