This semester I have learned much concerning the English language as well as grammar that I did not know before. Not only have I been able to learn more about linguistics but I also had the opportunity to gain skills that help me analysis corpuses and texts. For example, learning regular expressions has helped me learn how to be more efficient when analyzing text strings within a program. I have already used my knowledge concerning regular expressions within the internship interviewing process.

However, I am not a linguistics major and have never thought of analyzing large data sets. With that being said, I am an avid reader and enjoy spending my time in many of the classical works. But perhaps even more interesting to me is the trends and topics taught by LDS prophets. Recently my brother started to manually go through each Conference edition of the Ensign and see what sources were the most often quoted. For example, did General Authorities site C.S. Lewis more than PMG? Are poets more often quoted than the Church Handbook? His research intrigued me and made me come to the decision to analyze Church text.

I specifically want to research which General Authorities reference which sources of text. I believe that many times Latter-Day Saints don’t fully utilize all four standard works, but rather they focus on one set of work. By conducting research on the way that General Authorities use different sources of inspired words within their talks or articles will shed more light on how we as members should approach scripture study.

I believe that this will help me develop programming skills in learning how to effectively use a web crawler, efficiently accommodate regular expressions to match older styles of English, and accurately analyze trends within talks and speeches.

My father has all of the editions of the Journal of Discourses and I have since tired of searching for topics, people, or ideas manually. I intend to at least use the Journal of Discourses and if I am able to, I plan on using all the of the General Conference talks up to this last conference. If I am unable to use the more recent General Conference talks, I will analyze older Church texts and newspapers like *History of the Church, Times and Seasons* and the *Morning Star*.

I plan on getting this data by creating a web crawler to go through websites that contain these texts. I will also contact linguistic professors to try and see if they can point me in the right direction to obtaining a larger corpus. After creating a web crawler and getting the text, I will ensure that we get rid of punctuation, uppercase, etc. to facilitate the process of using regular expressions. Ensuring accuracy is always a difficult task when dealing with large amounts of text. However, I’m not the only one who has ever undertaken the task to analyze these texts. I can compare my results with other databases’ results and see if we are relatively same ball park.

First and foremost, I will need to find a way to access the data. Once I have found a website that I can most easily use, I will need to create a web crawler that effectively gets the data that I want and ignore the data that I don’t want. After having gotten the text, I will need regular expressions and lists of words that I am searching for. Within Python I will need to use modules like re, dictionaries, etc.

Hopefully by the end of this project I will have learned more programming skills and also enhanced my understanding of ways General Authorities approach the scriptures.