In going across the six registers of JA\_BI, JA\_HI, PS\_BI, PS\_HI, TB\_BI, and TB\_HI, I searched for noun + noun sequences when two sequential words are common nouns. Using the NLTK module within Python, I tagged anything that was a “NN” or “NNS.” “NNS” accounts for common plural nouns such as: scotches, products, bodyguards, etc. In order to avoid personal pronouns, I excluded any that were tagged “PPS”

To ensure accuracy, I checked several of the actual corpus’ texts and ensured that reoccurring noun + noun compounds could be found within the final report. Doing this helped me recognize that my program was considering ‘%’, for example, as a noun. The program also had difficulty with identifying the meaning of single letters that were found within the text. This helped me change my program to better handle these exceptions. I was unsure on the exact definition of a compound noun and further screening of compound words could be done. (For example, three compounded nouns, etc.)

It was interesting to see the different types of noun + noun combinations across the six separate registries. For example, within JA\_BI I saw a high frequency of heart disease, heart failure, body weight, and mucus bacteria. Clearly very technical medical and biological terms. However, if we look at PS\_BI (which still deals with biology) we see that their studies involve compound nouns such as cocoa beans, fossil record, life patterns, and male snakes.

Similarly, the HI registers differed. PS\_HI contains a high frequency of the compounded words of status quo, state governments, and punch cards. In TB\_HI we find words such as food needs, merchant ships and trading state. Both of these registers clearly deal with business and state affairs but clearly emphasize different areas. PS\_HI had a high number of concentration camp. JA\_HI had a smaller but still significant count of army careers and TB\_HI had a high count of war origins and air force. Each dealt with the significance of national conflict, but each had its own area of discussion.

In conclusion, I would need to enhance my definition of common and compound nouns and ensure that I incorporate further testing into the program to better my script. However, even from a simple program, such as this one, we can clearly see a difference not only between different types of genres, but even within those genres.