CPSC 4310/5310/7310 – Spring 2020
Natural Language Processing (NLP)

Assignment 4 [80 points]

Due on April 3rd, 2020

The problems are adopted from the textbook.

1. **[20 points]**

   Give the CFG and draw the parse trees for the following sentences:

   (a) I need a return flight between Calgary and Vancouver.

   (b) I would like to fly on Air Canada.

   (c) Is there an Air Canada flight from Lethbridge to Calgary?

   (d) Does Air Canada have a flight between 5:00 am and 6:00 am from Lethbridge to Calgary?

2. **[20 points]**

   Using the PCFG seen in class (i.e., worked example), run by hand the CKY parser on the following sentence:

   People fish tanks with rods

3. **[40 points]**

   Using the Brown corpus, write a program that classify texts according to one of the categories/genres, and evaluate the results of your program. You should use Word2Vec to represent the words, and the two following classifiers:

   (a) Logistic regression or single-layer perceptron, and

   (b) Multi-layer neural network (about 3 or 5 layers with 2 or 3 units each).

   You should split the data into training and testing (e.g., 90% and 10%), mix up the testing data and see if your classifiers will succeed to split them into categories/genres again.