Comp 6751 **Project 4** **Fall 2023**

Project 4: Extending a Feature Grammar

Due date December 11, 2023

Goal: extend your feature grammar for sentiment analysis to cover more complex cases. Compare resulting sentiment analysis to a baseline system. Extend your system to identify *stance* of the movie review towards a named entity.

Data: you should be guided by examples in the NLTK data > `sentence_polarity` > `rt-polarity.neg` You may simplify data samples to focus on the topic of the assignment, as for Project 3, but for Project 4, you should attempt more complex, longer sentences.

Description: You should extend your feature grammar for NLTK's feature based Earley's Chart parser (parser class "FeatureEarleyChartParser").

You are trying to develop the best grammar you can: wider coverage, less acceptance of ungrammatical information, and trying to connect sentiment and discourse relation information.

Suggested coverage

1. declarative sentences with sentiment-bearing words
2. include relative clauses
3. include conjunction (*and, or, but*) of sentiment bearing adjectives or nouns (*I saw a dull and scary movie, It was a mess and a hazard*)
4. conjunction (*and, or, but*) of sentiment bearing sentences (*It was too long but it was entertaining*)

Data: you may continue to use the polarity dataset of Project 3 as inspiration, or you may consult the movie_reviews dataset in NLTK.

SSAP baseline: run the aFinn Simplest Sentiment Analysis in Python

- download the aFinn sentiment lexicon (`https://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=6010`)
- get the SSAP to work (`https://finnaarupnielsen.wordpress.com/2011/06/20/simplest-sentiment-analysis-in-python-with-af/`)
- use your sentences for Project 3 and 4 and record the performance of SSAP
- record the performance of SSAP as a baseline and compare the performance of your Project 4 in both tabular form and with error analysis on 5 examples of different complexity

Deliverables:

Create a file *Good* with your training sentences that your grammar parses and labels correctly.

Create a file *False* with your training sentences that your grammar does not parse or label correctly.

- 1 file: your well-annotated grammar as a .pdf document (2pts, Grad Attr. 4,5,6)
- 1 file: *Good* with annotations as a ASCII document (.txt) (1pt, Grad Attr. 4,5,6)
- 1 file: *False* with annotations as a ASCII document (.txt) (1pt, Grad Attr. 4,5,6)
- 1 file: A report that includes:
  - one page explaining and critiquing your grammar design for complex declarative sentences (2pts, Grad Attr. 1, 6)
  - one page explaining and critiquing your grammar design for conjunction (*and, or, but*) of sentiment bearing adjectives or nouns (*I saw a dull and scary movie, It was a mess and a hazard*) with and without negation (1.5pts, Grad Attr. 1, 6)
  - one page explaining and critiquing your grammar design for conjunction (*and, or, but*) of sentiment bearing sentences (*It was too long but entertaining*) with and without negation (1.5pts, Grad Attr. 1, 6)
  - one page comparing SSAP and your grammar-based sentiment analysis of small paragraphs of text that contain sentiment words (1.5pt, Grad Attr. 1, 6)
  - one page explaining and critiquing your approach for stance assignment (1.5pts, Grad Attr. 1, 6)
  - a Demo file in .pdf format that illustrates your program running on two interesting examples (1pt, Grad Attr 1, 6)

Note on grading: if your Project 4 is better than your Project 3, it will count for both projects. Note also, that the report has to be in your words, reflecting on the indicated aspects.