Data Analysis:

Language is also a powerful medium for depicting tone, characters, mood, and theme in fictional literature. For all the literary tools involved in informing the reader of a work's happenings, adjectives play a pivotal role in coloring a tale with mood, characterization, and sensory details. This essay examines how adjectives play out within three classic works of literature: The Phantom of the Opera by Gaston Leroux, Winnie-the-Pooh by A.A. Milne, and Mary Shelley's Frankenstein through the use of a Natural Language Processing (NLP) method for the extraction and analysis of the frequency of adjectives. Examining these descriptions closely, the analysis seeks to showcase how language varies across genres, target audiences, and thematic content.

Project Data:

The data sources used in this project are The Phantom of the Opera by Gaston Leroux, Winnie-the-Pooh by A.A. Milne, and Mary Shelley's Frankenstein, all of which were obtained through Project Gutenberg's repository. The text files housing the contents of these books were saved locally and analyzed using the Python code linked above through the usage of the Natural Language Toolkit (NLTK).

Assumptions:

Before analyzing data, the script had made some vital assumptions. The assumption to begin with was that text files were encoded using UTF-8 and followed the usual pattern for Project Gutenberg, having *** START and *** as headers and footers, but not within the body of the work. A quick string slice was attempted to exclude this metadata with *** START and

*** being there. Although this approach is useful, it is sensitive and can fail to remove non-narrative material cleanly when these tags differ between texts.

Second, the analysis was dependent on the accuracy of part-of-speech (POS) tagging with NLTK's PerceptronTagger. The tagger applies grammatical labels to each word in the text so that the program can identify adjectives. Precisely, the analysis was interested in tags 'JJ' (basic adjectives), 'JJR' (adjectives of comparison), and 'JJS' (adjectives of superlative degree). This choice supposes that the PerceptronTagger works well enough concerning these grammatical classes even when it is applied to earlier or stylistically unusual English from 19th- and early 20th-century writing.

Methodology:

The script operated in three stages, the first being tokenization. In this stage, the entirety of the text was split into word tokens separately using NLTK's word_tokenize function. This function resulted in the story being converted into a list of words and punctuation, which would enable the data to be tagged.

In the second stage of the script, the aforementioned tagging would take place. Using PerceptronTagger each word was assigned the corresponding part of speech tag (either 'JJ', 'JJR', or 'JJS').

In the third stage, the frequency analysis took place. First, the adjectives pulled were converted to lowercase to avoid duplication (such as 'Little' being marked differently than 'little') , then the Python counter class was used in order to count the frequency of each adjective in the texts.

Results:

The results revealed distinctive adjective patterns in each literary work, strongly reflective of genre, tone, and intended audience.

Phantom of the Opera:	Winnie-the-Pooh:	Frankenstein:
little: 215	little: 74	own: 104
first: 98	good: 45	first: 78
other: 98	long: 31	other: 76
last: 95	other: 29	such: 70
old: 62	last: 24	many: 68
great: 62	first: 23	miserable: 63
young: 59	more: 20	same: 62
few: 56	happy: 17	few: 61
good: 55	big: 17	dear: 60
more: 51	small: 16	little: 55
whole: 45	right: 16	human: 55
poor: 45	sure: 15	several: 51
same: 45	much: 15	old: 50
own: 44	old: 14	more: 49
terrible: 43	many: 13	happy: 46
dead: 43	great: 13	great: 45
such: 42	dear: 12	poor: 38
black: 39	loud: 12	new: 38
persian: 37	funny: 11	last: 33
red: 36	large: 10	good: 32
curious: 28	next: 10	gentle: 31
white: 30	nice: 10	strange: 31

As one can tell at a glance, Leroux's *Phantom of the Opera* houses adjectives reflective of its mysterious and symbolic nature, as well as more physical descriptors for its characters. Using words like "dead", "persian", "red" to imply the character focus of the play, whilst words like "curious" suggest heavy thematic focus. Differing from the work of Shelley's *Frankenstein* uses words like "miserable", "happy", and "human" more frequently, showcasing the philosophical and introspective nature of the story. The contrast between misery and happiness evokes extensive thought from readers. Both of which are contrasted by Milne's

Winnie-the-Pooh, in which not a single adjective exceeds two syllables in length, betraying the story's young and inexperienced audience.

Conclusion:

Through this analysis of adjectives included in these works, the value of computational linguistics in literary studies is evident. By tokenizing and observing adjectives housed across different sub-genres, the ability to develop an understanding of how authors curate mood, develop their characters, and communicate theme. Each piece examined exhibits an adjective use appropriate for the conventions of the genre, target audience, and narrative aims of the work.

Works Cited:

Leroux, Gaston, and Nancy Holder. *Phantom of the Opera*. Naperville, Illinois, Poisoned Pen Press, 2020.

Shelley, Mary. *Frankenstein*. Lackington, Hughes, Harding, Mavor & Jones, 1 Jan. 1818.

Milne, A A. *Winnie-The-Pooh*. New York, Ny, E.P. Dutton & Company, 1926.