Name: Avishek Mukherjee

E-Mail: amukherjee3@albany.edu

Homework1

1. Modified the code on [https://stanfordnlp.github.io/CoreNLP/download.html point 4](https://stanfordnlp.github.io/CoreNLP/download.html%20point%204).

for file in `find /Users/ovishake/Desktop/Docker/ -name "\*.jar"`; do export

CLASSPATH="$CLASSPATH:`readlink $file`"; done

File A Output:

Avisheks-MacBook-Pro:stanford-corenlp-full-2018-10-05 ovishake$ java -Xmx1g -cp "\*" edu.stanford.nlp.pipeline.StanfordCoreNLP -annotators tokenize, ssplit, pos, lemma, ner -outputFormat json -file Feb28\_GroupA.txt

[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator tokenize

Processing file /Users/ovishake/Desktop/Docker/stanford-corenlp-full-2018-10-05/Feb28\_GroupA.txt ... writing to /Users/ovishake/Desktop/Docker/stanford-corenlp-full-2018-10-05/Feb28\_GroupA.txt.json

Annotating file /Users/ovishake/Desktop/Docker/stanford-corenlp-full-2018-10-05/Feb28\_GroupA.txt ... done **[0.2 sec].**

Annotation pipeline timing information:

TokenizerAnnotator: 0.2 sec.

TOTAL: 0.2 sec. for 6960 tokens at 43500.0 tokens/sec.

Pipeline setup: 0.1 sec.

Total time for StanfordCoreNLP pipeline: **0.4 sec.**

Avisheks-MacBook-Pro:stanford-corenlp-full-2018-10-05 ovishake$

You should put the annotators to get the result quickly.

Avisheks-MacBook-Pro:stanford-corenlp-full-2018-10-05 ovishake$ java -Xmx1g -cp "\*" edu.stanford.nlp.pipeline.StanfordCoreNLP -annotators tokenize, ssplit, pos, lemma, ner -outputFormat json -file Feb28\_GroupB.txt

[main] INFO edu.stanford.nlp.pipeline.StanfordCoreNLP - Adding annotator tokenize

Processing file /Users/ovishake/Desktop/Docker/stanford-corenlp-full-2018-10-05/Feb28\_GroupB.txt ... writing to /Users/ovishake/Desktop/Docker/stanford-corenlp-full-2018-10-05/Feb28\_GroupB.txt.json

Annotating file /Users/ovishake/Desktop/Docker/stanford-corenlp-full-2018-10-05/Feb28\_GroupB.txt ... done [0.1 sec].

Annotation pipeline timing information:

TokenizerAnnotator: 0.1 sec.

TOTAL: 0.1 sec. for 6462 tokens at 48954.5 tokens/sec.

Pipeline setup: 0.0 sec.

Total time for StanfordCoreNLP pipeline: 0.3 sec.

Time taken to run the

Avisheks-MacBook-Pro:stanford-corenlp-full-2018-10-05 ovishake$

2.a.

Null Hypothesis:

Male chat participants and female chat participants use the same number of words in their sentences.

2.b.

Null Hypothesis:

Male chat participants use the same number of pronouns as the female chat participants.

3. a.

Raw Hash Maps.

﻿runfile('/Users/ovishake/Desktop/Docker/soco hw/groupa.py', wdir='/Users/ovishake/Desktop/Docker/soco hw')

{'jake': 0, 'josh': 0, 'ted': 0, 'tony': 0, 'george': 0, 'nick': 0, 'michael': 0, 'david': 0, 'matthew': 0, 'aaron': 49, 'rick': 0, 'luke': 18, 'john': 0}

{'rebecca': 0, 'lynn': 0, 'meg': 0, 'caroline': 24, 'mara': 0, 'cintihia': 0, 'ariel': 0, 'eva': 0, 'leah': 0, 'macy': 0, 'amy': 0, 'michelle': 0, 'july': 0, 'vicky': 0, 'judith': 18, 'tia': 0, 'rita': 81, 'cinthia': 0}

﻿runfile('/Users/ovishake/Desktop/Docker/soco hw/groupb.py', wdir='/Users/ovishake/Desktop/Docker/soco hw')

{'jake': 0, 'josh': 30, 'ted': 37, 'tony': 0, 'george': 0, 'nick': 0, 'michael': 0, 'david': 0, 'matthew': 70, 'aaron': 0, 'rick': 0, 'luke': 0, 'john': 0}

{'rebecca': 0, 'lynn': 88, 'meg': 0, 'caroline': 0, 'mara': 0, 'cintihia': 0, 'ariel': 0, 'eva': 0, 'leah': 0, 'macy': 0, 'amy': 0, 'michelle': 0, 'july': 0, 'vicky': 0, 'judith': 0, 'tia': 0, 'rita': 0, 'cinthia': 24}

Group A

|  |  |  |  |
| --- | --- | --- | --- |
| **Male** | **Pronoun** | **Female** | **Pronoun** |
| Jake | 0 | Rebecca | 0 |
| Josh | 0 | Lynn | 0 |
| Tony | 0 | Meg | 0 |
| George | 0 | Caroline | 24 |
| Nick | 0 | Mara | 0 |
| Michael | 0 | Cintihia | 0 |
| David | 0 | Ariel | 0 |
| Matthew | 0 | Eva | 0 |
| Aaron | 49 | Leah | 0 |
| Rick | 0 | Macy | 0 |
| Luke | 18 | Amy | 0 |
| John | 0 | Michelle | 0 |
|  |  | July | 0 |
|  |  | Vicky | 0 |
|  |  | Judith | 18 |
|  |  | Tia | 0 |
|  |  | Rita | 81 |
|  |  | Cinthia | 0 |

Group B

|  |  |  |  |
| --- | --- | --- | --- |
| **Male** | **Pronoun** | **Female** | **Pronoun** |
| Jake | 0 | Rebecca | 0 |
| Josh | 30 | Lynn | 88 |
| Tony | 37 | Meg | 0 |
| George | 0 | Caroline | 0 |
| Nick | 0 | Mara | 0 |
| Michael | 0 | Cintihia | 0 |
| David | 0 | Ariel | 0 |
| Matthew | 70 | Eva | 0 |
| Aaron | 0 | Leah | 0 |
| Rick | 0 | Macy | 0 |
| Luke | 0 | Amy | 0 |
| John | 0 | Michelle | 0 |
|  |  | July | 0 |
|  |  | Vicky | 0 |
|  |  | Judith | 0 |
|  |  | Tia | 0 |
|  |  | Rita | 0 |
|  |  | Cinthia | 24 |

3.B. Average.

Conversation A:

MALES: 5.58

FEMALES: 6.83

Conversation B:

MALES: 11.41

FEMALES: 3.66

3.c.

t-value = 0.97949

p-value = 0.430625

for p<0.05 (Significance Level)

Conclusion:

Nothing could be said for sure, as this establishes a Null Hypothesis.

Men and Women use almost the same number of pronouns.