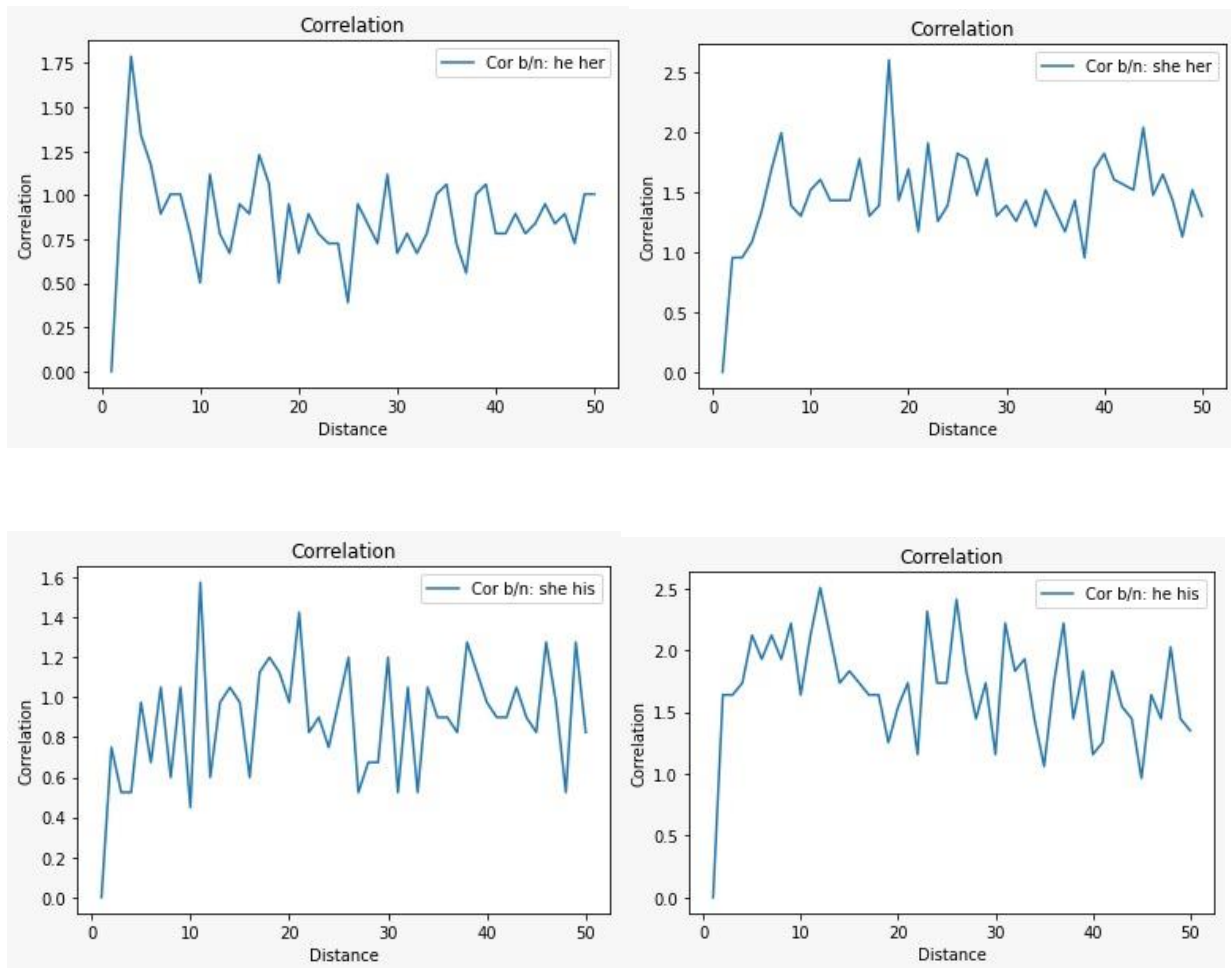


Question 1:



It can be observed (subtle) that the correlation between the mentioned words achieve the highest correlation before the distance of 20 and later the corr value drops or fluctuates extremely depending on the words occurring in this span between the target words W1 and W2.

Question 2:

3. Kneser-Ney Smoothing

a. Results:

1. Count of Longbourn: 142 & Pleasure: 142
2. Count of Bi-grams ending with Longbourn: 11 & Pleasure: 25
3. Negative $\log_2(P \text{ Lidstone})$ for Longbourn: 10.437752072323603 & Pleasure: 10.437752072323603
4. Negative $\log_2(P \text{ Kneser-Ney})$ for Longbourn: 12.014179013458373 & Pleasure: 10.829754442320946

b. Idea behind the Kneser Ney smoothing technique (Ref: Stanford NLP)

- KN smoothing is an extension of absolute discounting with a clever way of constructing the lower-order model.
- The lower-order model constructed above is significant only when count is very small in the higher-order model and thus it should be optimized.

Results for task 1 and 2 in question could be found in the python notebook, kindly refer the same!