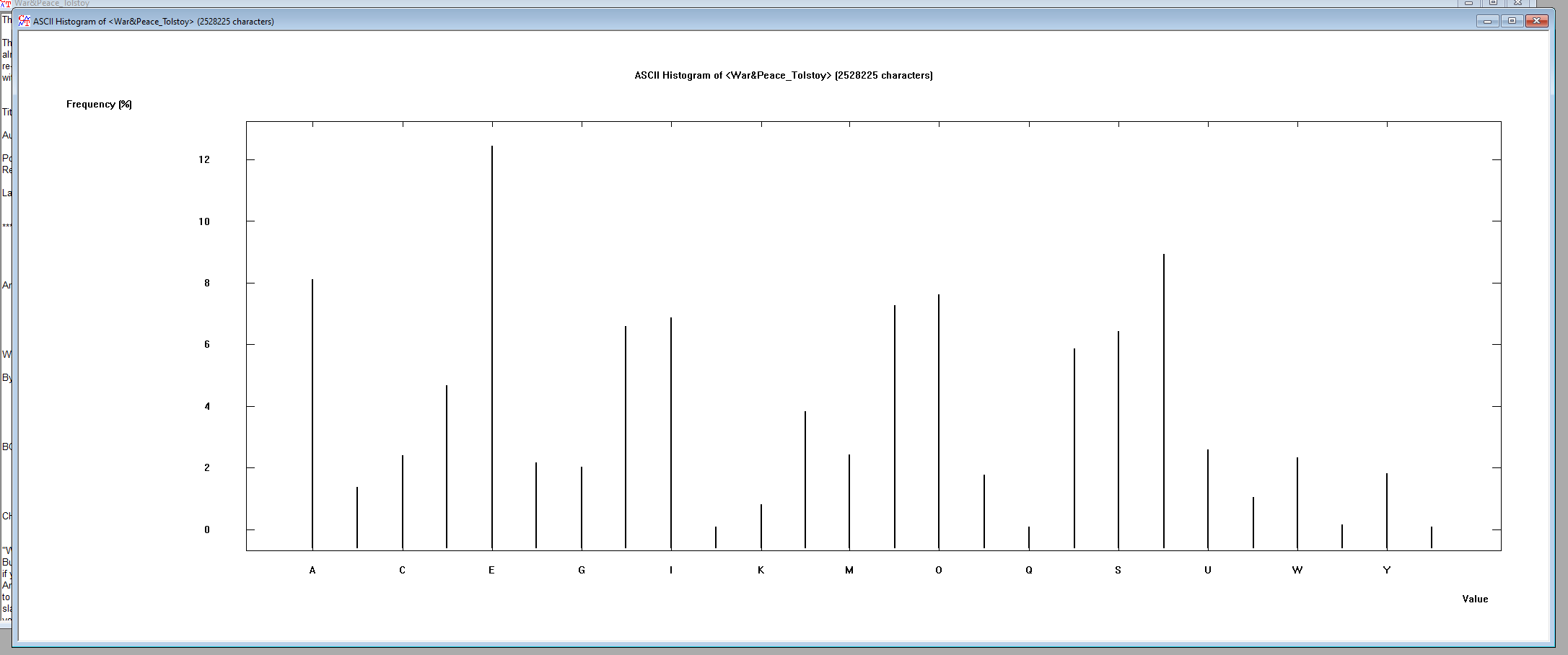
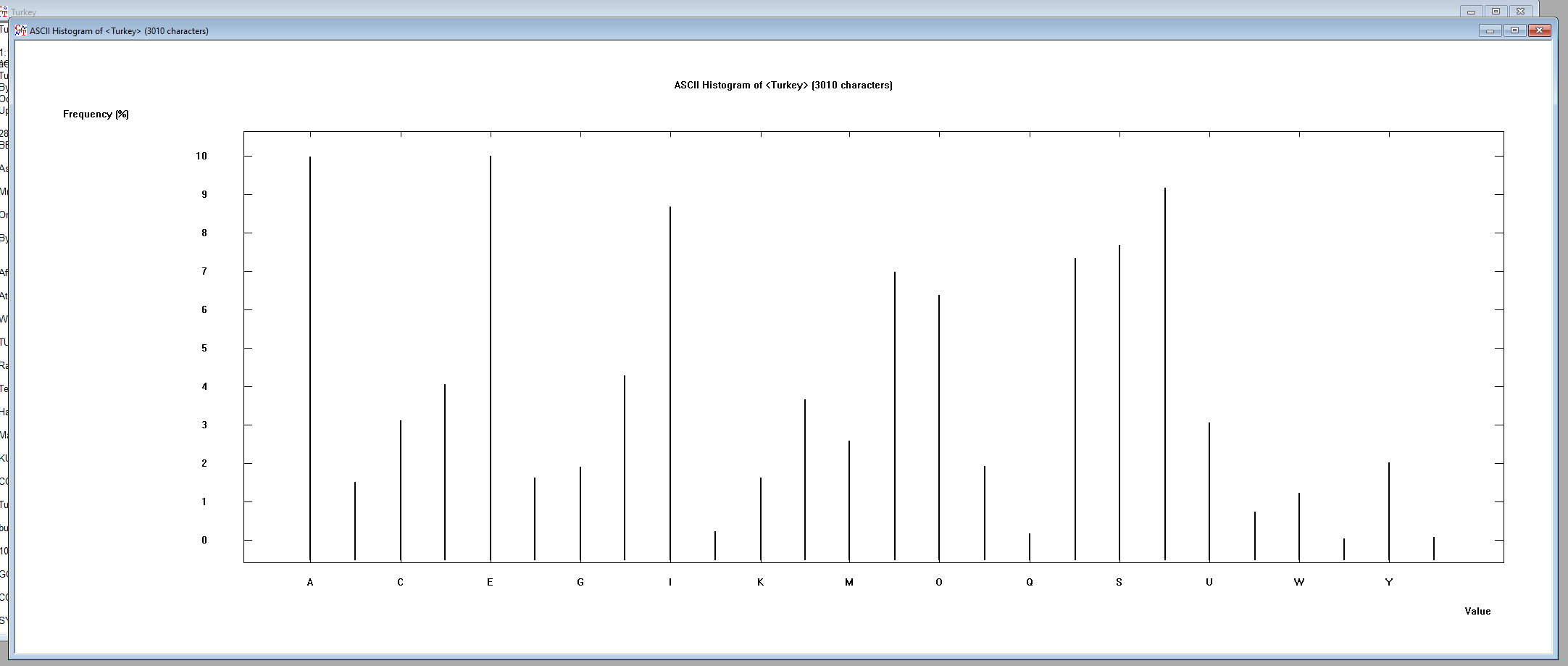
# Crypto Assignment

## Task 1 (4 points)

* (1.5 points) Create a histogram for each text that display the relative frequency of letters in a graphical form. For this, go to Analysis > Tools for Analysis. Provide the two histograms in screenshots.





* (1.5 points) Do the two histograms depend significantly on the texts you have provided?

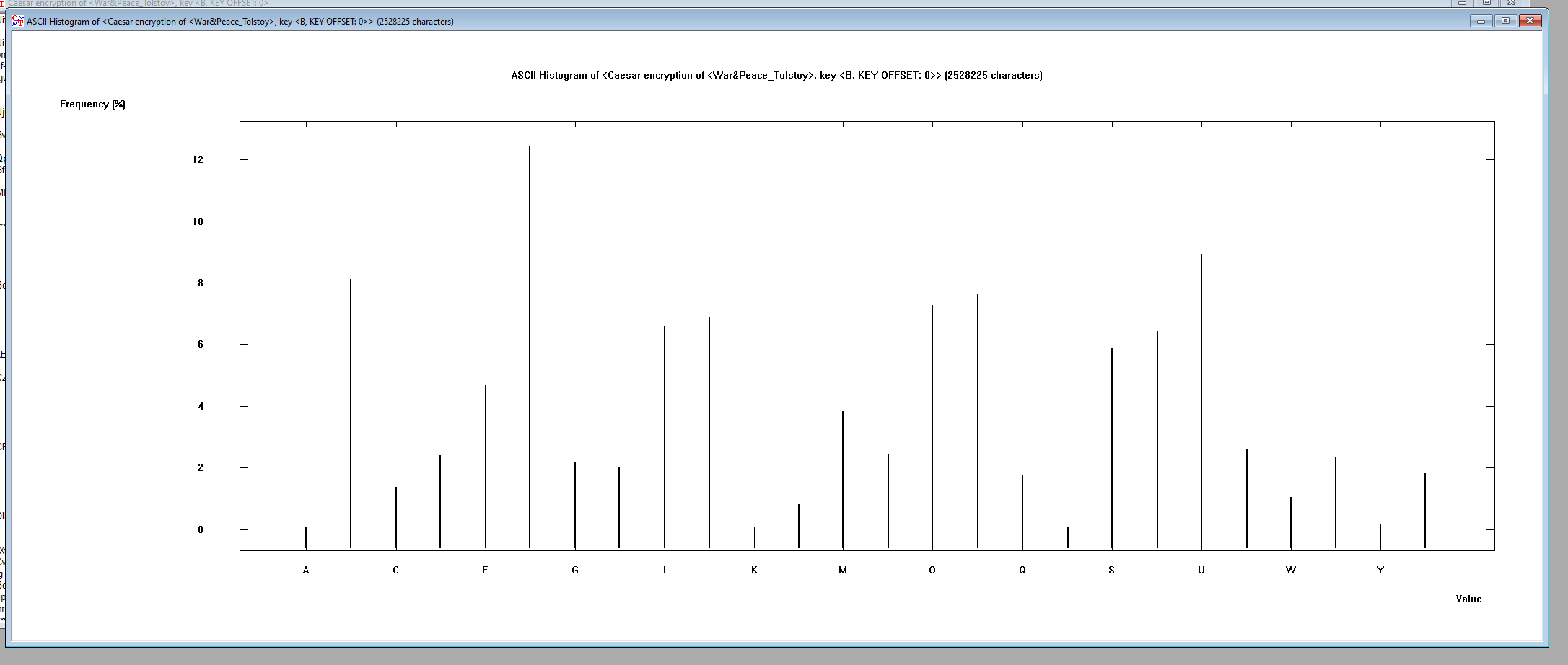
No, they do not since the histogram shows which letters appear the most. This is not dependent on the text since many of the letters that are common on one of the texts are common on the other as well. This is a constant of the language used since letters like E appear almost equally as high despite being completely different texts

* (1 point) Calculate the entropy of each text. For this, Analysis > Tools for Analysis > Entropy. What would you conclude from the comparison of the entropies? The entropy shows that both are 4.16 by using all 26 letters of the English alphabet. This proves that both texts use the same letters and are extremely similar in the language that they use.

## Task 2 (4 points)

* (2 points) What are the characteristic features of the obtained distribution compared with the original text? Provide a screenshot to support your answer.

The main difference between the new and the old histogram is that the new one is shifted to the right due to the letters being shifted to the right with the letter E becoming the letter F. Thus, the letter to the right in the alphabet became the same in usage as the original letter following the encryption.



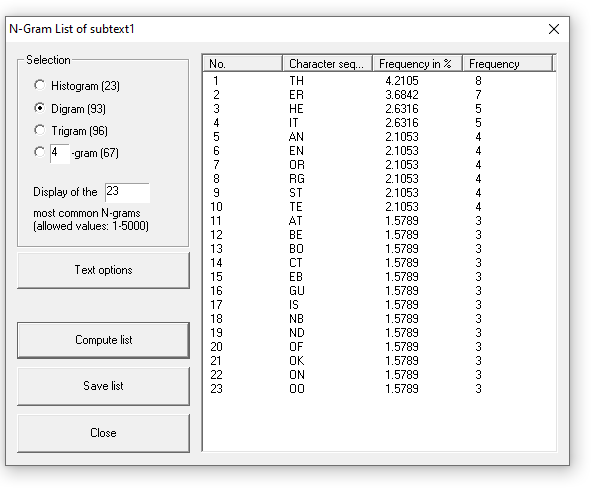
* (2 points) How would you apply the features you have discovered in cracking the key?

If English is the language of the document, I could know that the letter E is the most often used and the fact that it is not the most popular letter used and the most popular letter is to the right of it could indicate a possible Caesar encryption. I would then run a decryption of the text with B set as the character to indicate a shift to the right. This would allow me to determine if my presumption is correct or not, which in this case it would be.

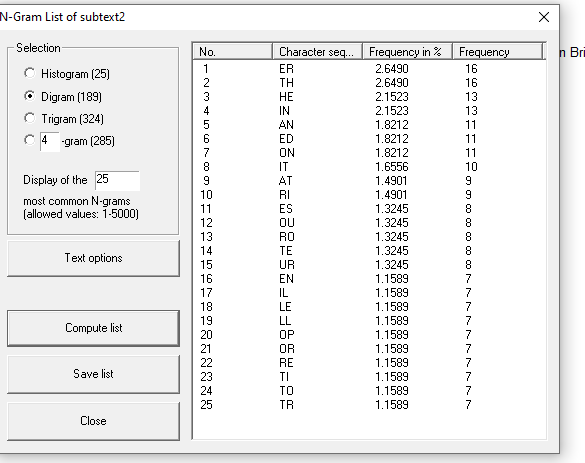
## Task 3 (4 points)

* (1.5 points for War&Peace\_Tolstoy.txt) Summarize your observation of the frequency distribution. Also, provide a screenshot of the frequency distribution.

The frequency indicated the most popular character sequence to be TH which occurred 4.2105% of the time. This is useful since cracking the most frequent character sequence is good step towards cracking the encryption.



* (1.5 points for a NYT article) Summarize your observation of the frequency distribution. Also, provide a screenshot of the frequency distribution.

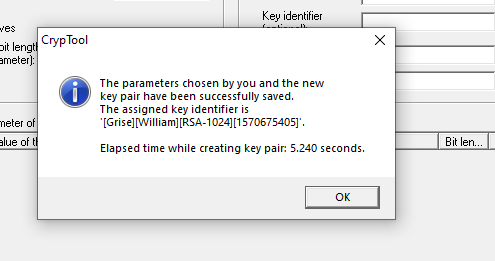


The frequency indicated the two most popular character sequence to be ER and TH which both occurred 2.6490% of the time.

* (1 point) Compare and contrast the two frequency distributions. The major difference between the two is that the TH sequence occurred 2% more in the War and Peace article than in the NYT article. This is not a big issue since the top three-character sequences are the same for both articles. Those three-character sequences are ER, TH, and HE. This indicates that these 3-character sequences are common in the English language and if they are able to be decipher when encrypted are big towards decrypting the rest of the document.

## Task 4 (4 points)

* (2 points) Attach a screen shot that shows the successful creation of the key pair.



* (2 points) Attach a screenshot that displays the signature verification.



## Task 5 (4 points)

* (1 points) List the cipher and the key for Ciphertext 1.

Vigenere with a key of the letter R

* (1 points) List the cipher and the key for Ciphertext 2.
* (2 points) List the cipher and the key for Ciphertext 3.

ADFGVX was used with the standard matrix and the transposition password being password