Cryptography has been an integral part of human history, evolving alongside our need to protect sensitive information. This necessity has only intensified in the modern era, driven by technological advancements and the growing demand for secure communication and data protection. To meet these needs, various encryption algorithms have been developed. These range from early techniques, such as the monoalphabetic substitution of the Caesar Cipher and the linear, alphanumeric Affine Cipher, to more complex methods like the polyalphabetic Vigenère Cipher. Today, modern cryptographic algorithms, such as AES, form the backbone of secure data transmission. While the early encryption methods are no longer used in practice, they serve an important role in educating the public about the foundational principles of cryptography.

Cryptography - Cryptanalysis

B) 1)

- The text was selected based on the criterion of being a general news article, and was therefore chosen from the *World* section of *The Guardian* website: https://theguardian.com/world/ from the news. The title of the text :***Trump praises North Korea's 'progress' as expert warns against optimism***

- With the help of the Cryptool tool – by selecting from the menu Analysis-> Tools -> Histogram and N-gram we obtain the following results :

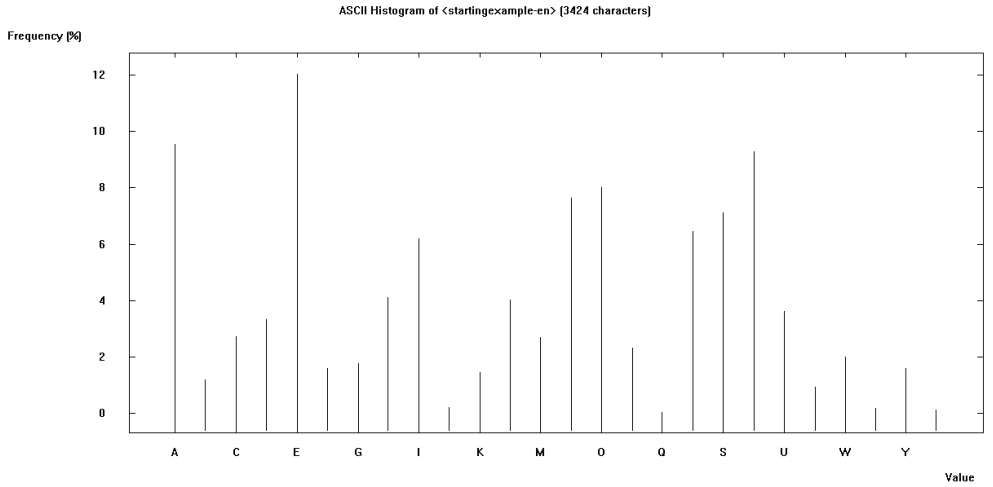


Figure 1 – Histogram

The histogram illustrates the frequency of letters in the text, which closely align with the typical English letter frequency distribution. Notably, the letter **E** ranks highest in frequency, followed by **A** and **T**, while **Z** and **J** appear the least often. Further analysis through the N-gram option provides a more detailed breakdown of letter frequencies. Specifically, the letter **E** occurs with a frequency of 12.0327, **A** appears with a frequency of 9.5502, and **T** has a frequency of 9.2874. As we continue down the frequency list, we see that the letter **Z** is the least frequent, with a frequency of just 0.0292.

- By selecting "Encrypt/Decrypt" and then applying the Caesar cipher, we observe that the text is transformed into an

encrypted version, with each letter shifted according to the specified encryption key.

We follow the same steps for analyzing histograms and letter frequency, observing how the letters appear with the highest frequency in the text.

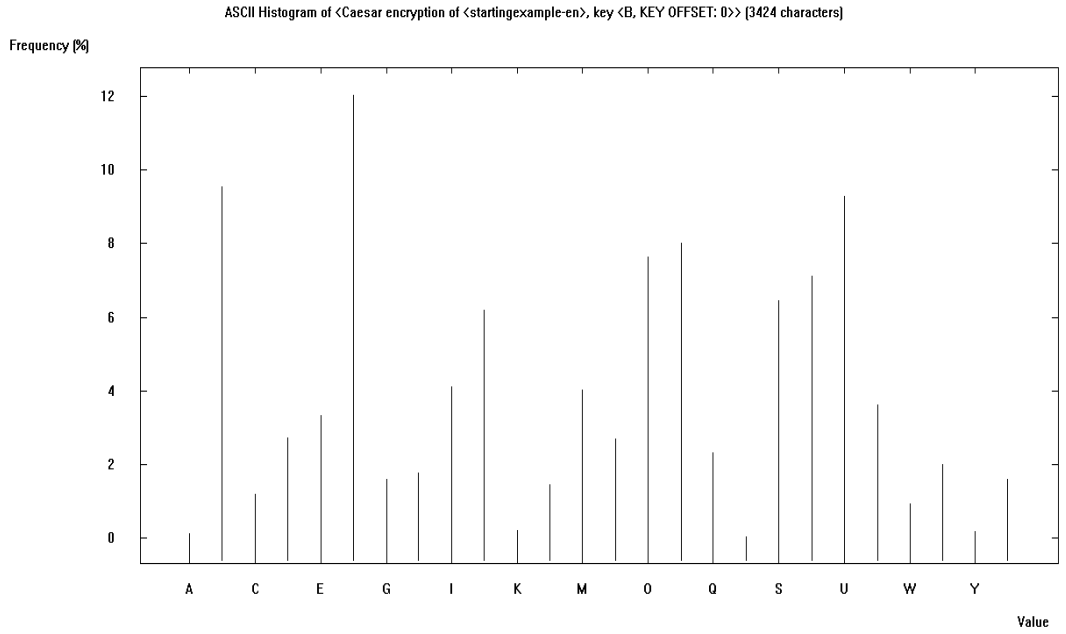


Figure 2

Below are the frequencies of the letters in the encrypted text. In this case, the letter **E** from the original text is now replaced by **F**, which occupies the second position with a frequency of 9.5502.

Similarly, the letter **B** has taken the place of the letter **A**, and at the end of the frequency list, the letter that replaced **Z** appears with a frequency of 0.0292.

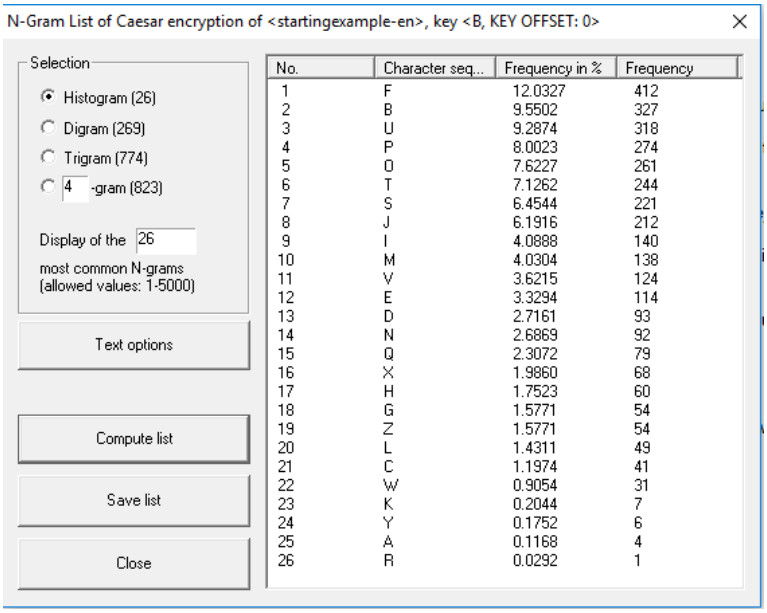


Figure 3 – N-Gram

B) 2)

The second text to be analyzed is a technical piece titled *A Computational Framework for Conceptual Blending*, sourced from the website:

https://www.sciencedirect.com/science/article/pii/S000437021730142X

In this text, the letter **E**, the most frequent letter in the English alphabet, occurs with a frequency of 10.6111%, while **T** follows closely at 9.2680%. At the opposite end, the letter **Z** has the lowest frequency, with an occurrence of just 0.0672%.

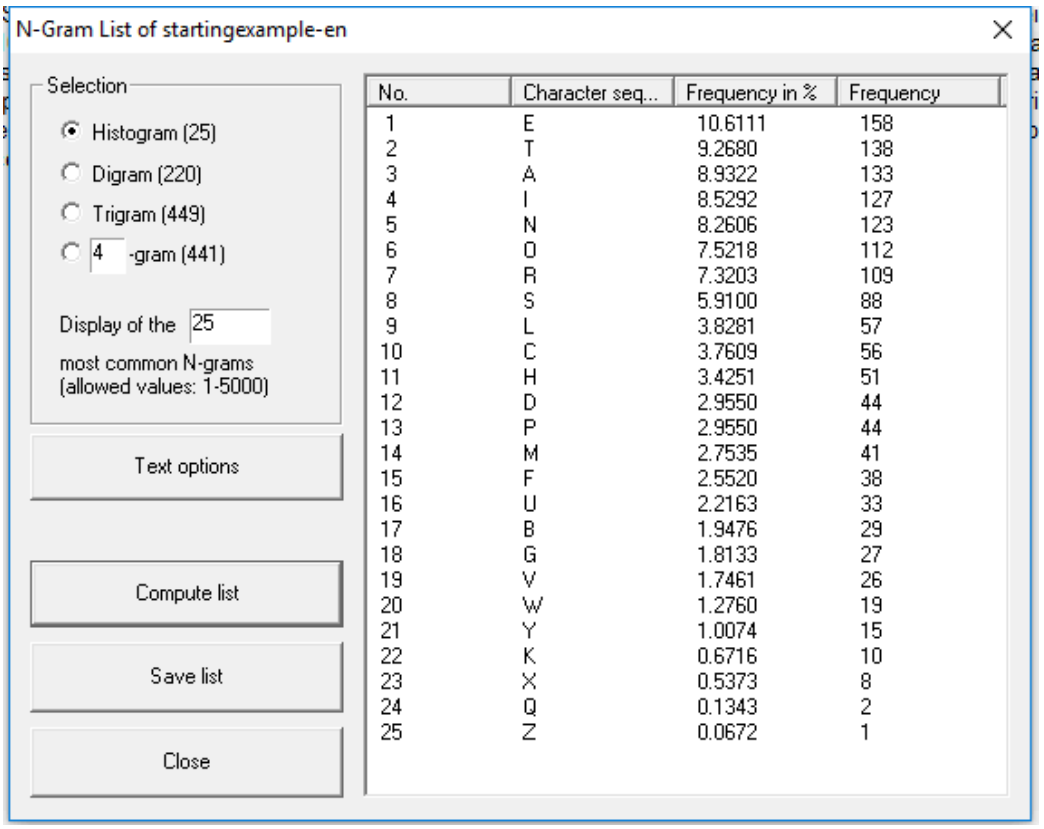


Figure 4

After encrypting the text, the frequency of letters changes according to the encrypted text and they appear as follows:

Frequency in letter

F – 10.6111
U – 9.2680
…..……..
R – 0.1343
A – 0.0672

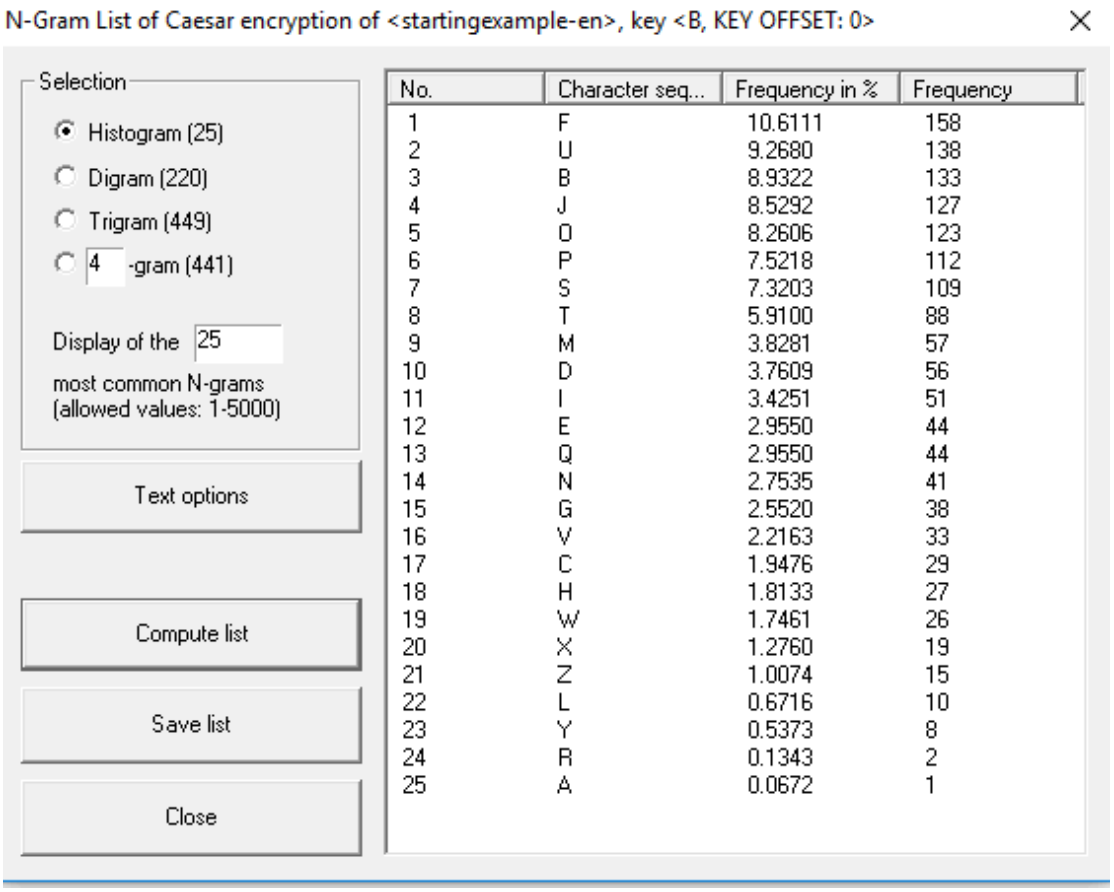


Figure 5

The histogram below shows the letters with the highest and lowest frequency in analytical order after mapping.

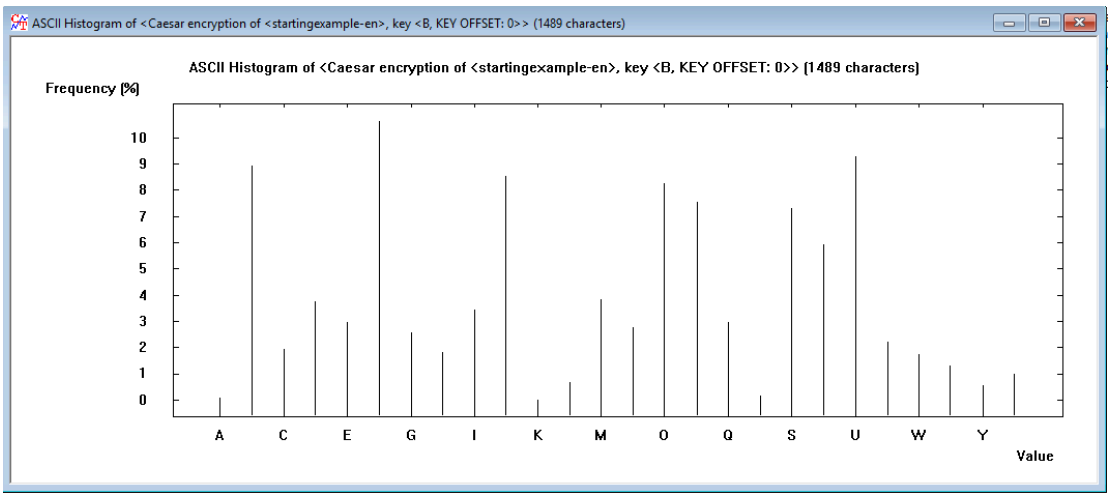


Figure 6

In the first text, the letter frequencies closely match those of the standard English alphabet frequency distribution. However, in the second text, there is a noticeable deviation. For example, the letter **E** appears with a frequency of 10.6911%, and similar discrepancies are observed across several other letters, leading to significant differences in the overall frequency distribution.

C)

- The first text was encrypted using the Caesar cipher (Figure 7), with the letter **G** from the English alphabet serving as the encryption key. When analyzed using the N-Gram option in CrypTool, the frequency distribution mirrors that of the original text, though the letters themselves are shifted according to the cipher. For example, a frequency of 12.0629 corresponds to the letter **X** instead of **E**. From this, we can easily conclude that the cipher is a monoalphabetic substitution algorithm. Additionally, by examining the digrams, we would further confirm that it is a sliding cipher, specifically the Caesar cipher.

- The second text exhibits "manipulated" frequencies, with the letter **D** appearing at a frequency of 14.4860, among other deviations. This suggests that **D** has likely replaced the letter **E** from the original text. Despite these alterations, we can still conclude that the cipher is a Caesar cipher, as the pattern reflects a shift in the alphabet.

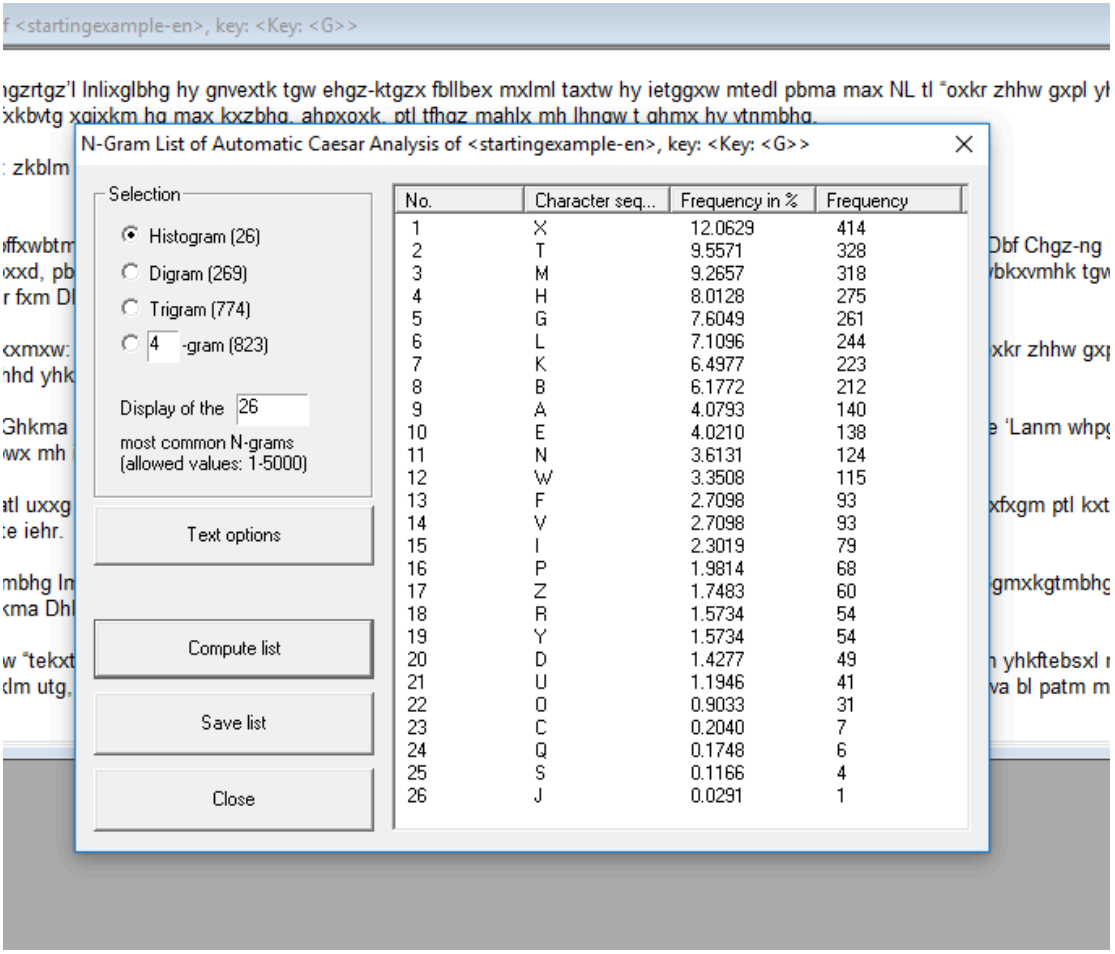
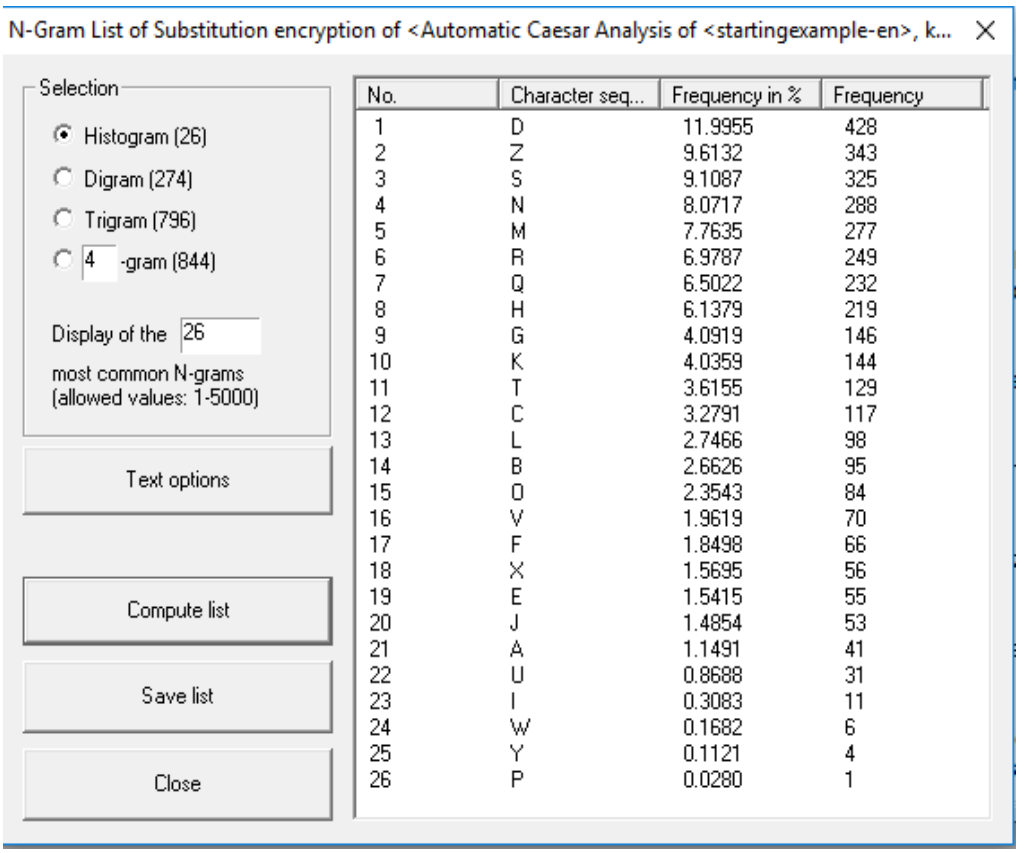


Figure 7

For the simple substitution encryption (Figure 8), the key **Z** was used in the first text. The resulting frequencies, as analyzed using the N-Gram tool (Figure 8), show that the letter **D** now appears with a frequency of 11.9955, and the frequency distribution closely resembles that of the English alphabet. This suggests that a simple substitution algorithm has been applied. Notably, the letter **Z** appears with a frequency of 9.6132, which is significantly higher than its usual occurrence in English text. Given that it appears frequently, we can infer that **Z** is likely the key. Therefore, we can confidently conclude that the encryption is based on a simple substitution cipher, with **Z** as the key.

In the second text, the letter frequencies are similar to those of the original, pre-encrypted text. However, here, **C** appears first with a frequency of 14.4860, followed by the letter **R** at 9.6184, and so on. Due to the specialized nature of the text, the frequencies deviate slightly from the standard English alphabet distribution. Despite these variations, we can confidently conclude that the Poseidon algorithm is a simple substitution cipher. It is likely that **C** has replaced **E**, **R** has replaced **A**, and other letters have been substituted similarly.



Εικόνα 8

- For the permutation algorithm applied to the first text (Figure 9), a key of **k=3{3,2,1}** was used. The resulting letter frequencies closely resemble those of the standard English alphabet. For instance, **E** appears with a frequency of 11.9955, and **A** with a frequency of 9.6132, among others. Based on these frequency patterns, we can confidently conclude that the algorithm belongs to the category of monoalphabetic

substitution with permutations. To precisely identify the specific algorithm, however, further analysis would be required.

- In the second text, after encryption, the letter frequencies closely match those of the English alphabet. Specifically, **E** has a frequency of 13.6808, **T** appears with a frequency of 9.3961, and so on.

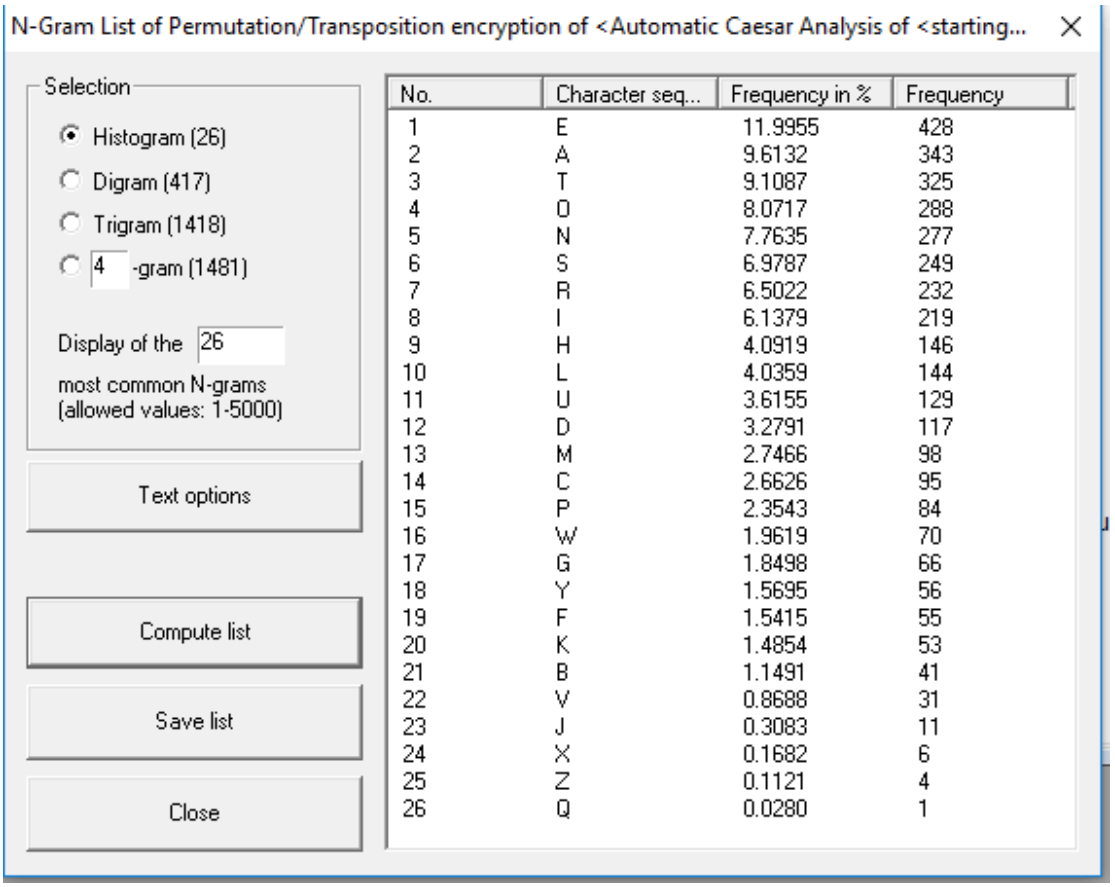


Figure 9

- For the Vigenère cipher applied to the first text (Figure 10), the key **LUCKY** was used. The N-Gram analysis reveals that the letter frequencies appear "broken," suggesting that each letter has been encoded with a different, seemingly random letter based on the key. This pattern strongly indicates that the Vigenère encryption algorithm has been used.

- In the second text, the frequencies also appear "broken," suggesting that the Vigenère cipher is likely being used.

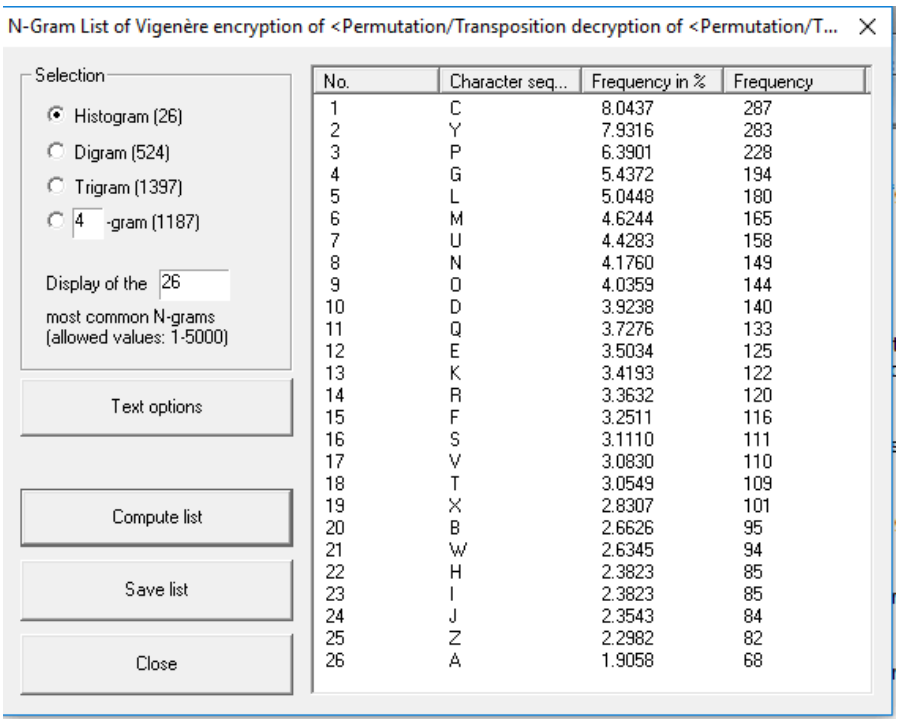


Figure 10

D) For the Affine cipher, a C language implementation is provided.

E)

a) In the text provided in the file **encr_substitution.txt**, we attempt cryptanalysis using the CrypTool application, focusing on the letter frequency distribution. This method allows us to identify some of the words, such as "THE," "SIDE," "KEY," and "CODE," but the entire message remains undeciphered.

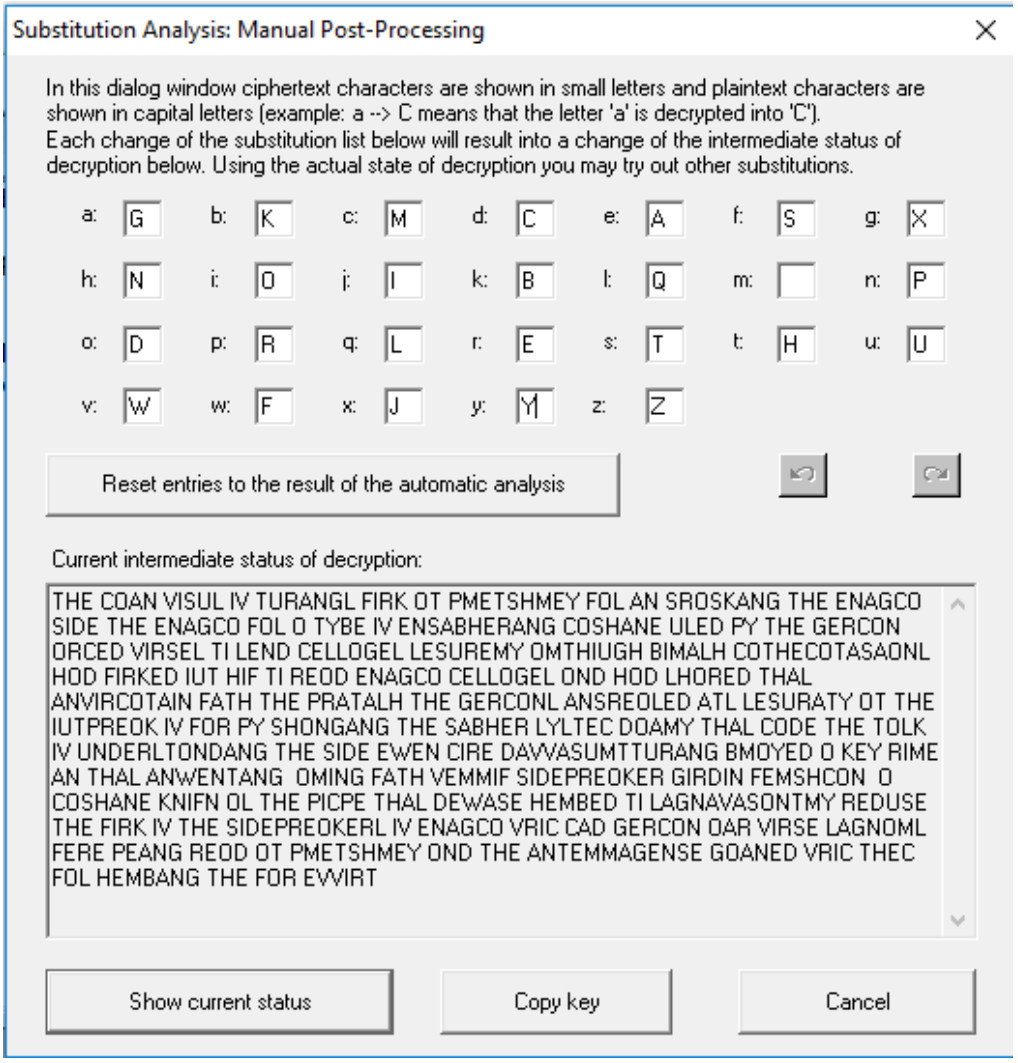


Figure 11

b)

In the attempt to recover the message again, knowing everything about the ENIGMA encryption machine, where we see from the text that a word similar to the word ENIGMA – ENAGCO, from which we conclude that the letter I has been replaced by the letter A, the letter M has been replaced by the letter C, and the letter A by the letter O. We make the replacements and move on to the next letters.

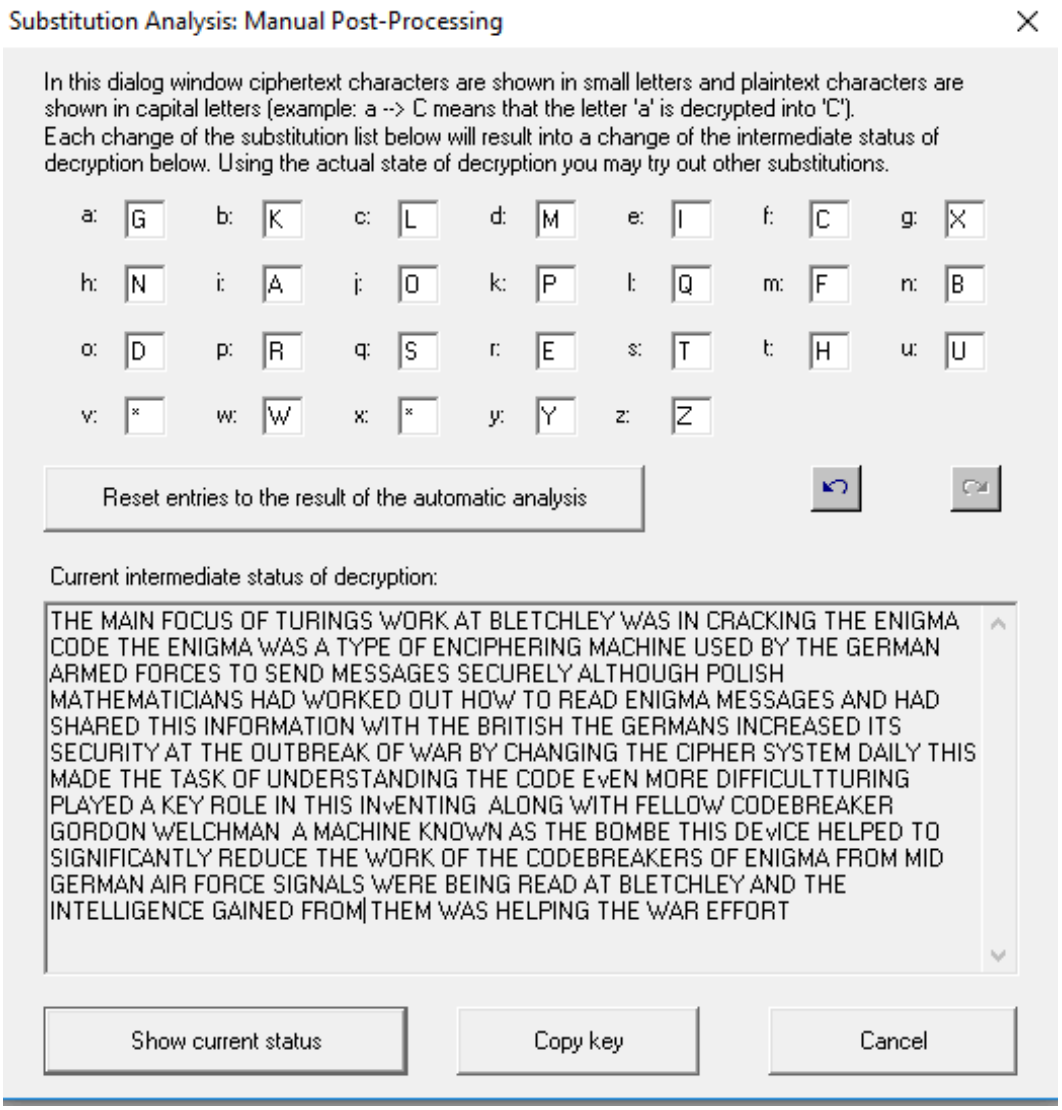


Figure 12

The decrypted message is presented below, making substitutions for words that look familiar, we end up with this message:

THE MAIN FOCUS OF TURINGS WORK AT BLETCHLEY WAS IN CRACKING THE ENIGMA CODE THE ENIGMA WAS A TYPE OF ENCIPHERING MACHINE USED BY THE GERMAN ARMED FORCES TO SEND MESSAGES SECURELY ALTHOUGH POLISH MATHEMATICIANS HAD WORKED OUT HOW TO READ ENIGMA MESSAGES AND HAD SHARED THIS INFORMATION WITH THE BRITISH THE GERMANS INCREASED ITS SECURITY AT THE OUTBREAK OF WAR BY CHANGING THE CIPHER SYSTEM DAILY THIS MADE THE TASK OF UNDERSTANDING THE CODE EvEN

MORE DIFFICULT TURING PLAYED A KEY ROLE IN THIS INvENTING ALONG WITH FELLOW CODEBREAKER GORDON WELCHMAN A MACHINE KNOWN AS THE BOMBE THIS DEvICE HELPED TO SIGNIFICANTLY REDUCE THE WORK OF THE CODEBREAKERS OF ENIGMA FROM MID GERMAN AIR FORCE SIGNALS WERE BEING READ AT BLETCHLEY AND THE INTELLIGENCE GAINED FROM THEM WAS HELPING THE WAR EFFORT

Z) Using the encrypted text encoded with the Vigenère cipher, we perform cryptanalysis with a focus on determining the key length.

The ciphertext: **hitciwtwvzzxciwdeopchfvlppopw**

By observing the characters and their recurrence we see that the trigram : ciw appears twice in the text. The distance between the first occurrence of the trigram until the second occurrence is likely to give us the key or its divisors. In this way we conclude that the possible key size is 9 and its divisor is 3.

Therefore we "break" the message into columns observing the possible versions.

We start the tests with a possible key of 3.

H	I	T
C	I	W
T	W	V
Z	Z	X
C	I	W
D	E	O
P	C	H
F	V	L
P	P	O
P	W	

x	x	x

We apply the Caesar cipher to each column, analyzing the frequencies as follows:

- In the first column, **p** is shifted to **e** (shift of 11).
- In the second column, **i** is shifted to **e** (shift of 4).

Third column:

W	E	T
R	E	W
I	S	V
O	V	X
R	E	W
S	A	O
E	Y	H
U	R	L
E	L	O
E	S	

WE(T)**RE**(W)**IS**(V)**OV**(X)**RE**(W)**SA**(O)**EY**(H)**UR**(L)**EL**(O)**ES**

The decrypted message:

WEAREDISCOVEREDSAVEYOURSELVES

Where the key is: LET

And it means: **WE ARE DISCOVERED SAVE YOURSELVES**

(We are discovered, save yourselves.)

Sources used:

https://www.dcode.fr/caesar-cipher
http://www.multiwingspan.co.uk/cipher.php?page=caesar

For the image:

http://giuliodagostino.com/2018/03/cryptography-cryptanalysis-tools-part-1/

Ambel Basha
Athens May 2018