Q10) The following ciphertext is obtained using a substitution cipher. Give a clearly written description of steps you followed to decrypt the ciphertext. F → W

The first thing that I noticed from the ciphertext description is that this is a substitution cipher and as F → W, then it is a monoalphabetic substitution cipher.

i) The first thing I did was perform a frequency analysis on the ciphertext to identify frequency of N-grams. 1-gram is the single letter occurrences, so 26 monograms in the english alphabet. I also find a frequency of diagrams and triagrams and 4-grams. Frequency has been computed using a small script written on Python.

1-Gram frequencies =

{ C:37, G:24, S:20, K:18, Y:15, I:15, V:14, N:13, Z:13, E:12
O:10, F:9, D:8, L:7, X:7, J:7, P:6, M:5, W:5, H:5, A:5, Q:4}

From the 1-gram frequencies we can see that only 22 characters appear in the ciphertext and these are 4 characters which were never encrypted. The 4 characters to not occur in ciphertext are 'B', 'R', 'T', 'V'. These ciphertexts most probably represent very infrequently occurring letters in the english alphabet such as X, Q, J and Z.

ii) Similarly I perform bigram, trigram and quadgram and pentagram frequency analysis of the ciphertext which leads to the following result :—

- Bigram Frequencies

CG : 7
ZC : 7
NC : 5
YS : 5
GO : 5
CK : 5
AC : 5
CN : 5
SF : 4
GY : 4
YK 4
FZ : 4
MG : 3
GL : 3
US : 3
IC : 3
SI : 3
KU : 3
KS : 3
XE : 3
CJ : 3
SH : 3
XC : 3

OI : 3
KG : 3
CC : 3
CI : 3
ZE : 3
WY : 2
NS : 2
YK : 2
DP : 2
VM : 2
JC : 2
PK : 2
UG : 2
OL : 2
GI : 2
GA : 2
SA : 2
KZ : 2
KX : 2
DS : 2
LK : 2
IU : 2
IG : 2
ND : 2

DG : 2
YY : 2
EO : 2
JN : 2
EU : 2
JU : 2
UC : 2
UZ : 2
CF : 2
CS : 2
EJ : 2
EM : 1
LO : 1
OJ : 1
SU : 1
UP : 1
DC : 1
GD : 1
DN : 1
CU : 1
⋮

From the bigram frequencies we observe that {CG and ZC} occur many times. The most common bigrams in English are :-

th, he, in, en, nt, re, er, an, ti .....

but we have already assumed that C → e so G could be either {n, t, a} or some other consonant. We also see FZ occurring 4 times. We know F → W so, Z could be {i, h, o} but we also know that ZC occurs very often and C → e so Z can't be 'o' and could be 'i' or 'h'.

iii) Now, we compute 3-gram (Trigram Frequencies)

YSF : 3
GOI : 3
FZC : 3
ZCC : 3
CCN : 3
CYK : 2
JCK : 2
GOL : 2
ICG : 2
[illegible]
CGI : 2
NCG : 2

GAC : 2
CKS : 2
SAC : 2
CKX : 2
KSH : 2
ZCN : 2
KGO : 2
CDN : 2
NDG : 2
DGY : 2
GYY : 2
YYS : 2
JNC : 2

CJU : 2
UZC : 2
CFZ : 2
ZEJ : 2
EMG : 1
MGL : 1
GLO : 1
⋮

We observe from the bigram frequencies that we get {YSF, QOI, FZC} 3 times

FZC → wZe so Z should most likely be h

Hence Z → h

We also see that {QYY, YYS} occur 2 times and {YY} occurs 2 times in the bigram frequencies. The most common same letter bigrams in the English language are {'ee', 'll', 'nn', 'mm', 'rr'} so Y probably belongs to one of them.

(iv) Performing quadgram (4-grams) frequency analysis

FZCC : 3	EMQL : 1
ZCCN : 3	MQLO : 1
ICQE : 2	QLOS : 1
CCND : 2	⋮
CNDQ : 2	
NDQY : 2	
DQYY : 2	
QYYS : 2	
CFZC : 2	

We see, we have ZCCN occurring 3 times. We also know ZCCN → heeN so N could be {'l', 'r'} etc.

Now, let us perform the quadgram (5-grams) frequency analysis

v) 5-Gram (Pentagram Frequency Analysis)

FZCCN : 3	DGYYS : 2
ZCCND : 2	GYYSF : 2
CCNDG : 2	CFZCC : 2
CNDGY : 2	EMGLO : 1
NDGYY : 2	MGLOS : 1
	⋮

We have FZCCN occurring 3 times, so this could most probably be :-

FZCCN → WZeeN → WheeN so, N is most probably 'h' and we have FZCCN → wheel

Now, from monogram frequencies we can see

C : 37	K : 18	N : 13	H : 5
G : 24	Y : 15	Z : 13	A : 5
S : 20	V : 14	E : 12	Q : 1

So, we can see from lower frequencies that most probably

{Q, B, R, T, V} → {K, X, Q, J, Z} and

{G, S, K, Y, V} → {E, T, A, O, I, N}

We know C → e so {G, S, K, Y, V} → {T, A, O, I, N}

also

{N, Z, E, O, F, D, L} → {S, R, H, D, L, U, C}

We know Z → h and N → l so,

{E, O, F, D, L} → {S, r, d, u, c}

vii) Now, we see occurrences of F with other ciphertext characters.

FZE in trigram frequency

FZE → WhE , so E may be 'i'

The most frequent bigraph in ciphertext is 'CG'.

CG → eG so G may be 'a', 'l', 'n', h' etc.

So, as has frequency of 4, or G → a.

We have test out this mapping on a few examples.

CNDGY → elDaY

If we take C → a, we can apply it to solve GYY and GYS

YY → {'ll', 'ee', 'hh', 'mm'....}

GYY → aYY This implies Y → r

CNDGY → elDar so D → {'C', 'B', 'M', 'N'} etc.

Observing quadgram and pentagram frequencies for D, we see DGYY:2, DGYYS:2, DGYYSF:2

Darr DarrS DarrSw

arrSw → arrow would be a plausibility

so, we see that {S → o} and that leaves only one possibility {D → b}, hence

DGYYSF → barrow

vii) Going back to monogram frequencies, we had seen {G, S, K, Y, U} → {t, a, e, i, n} but now we know G → a and Y → 'r' so

{S, K, U} → {t, o, i, n}

The total encryptions we have deciphered are:—

W → F
h → Z
e → C
l → N
b → D
a → G
r → Y
o → S

Now, let us decipher the ciphertext using what we know

~~EMaLl~~

EMGLOSU D C GDN CU SWY S FHNSFCY KD PUMLWGYI.

EMaLOoU beable Uo WroW Hlower Kb PUMLWarI ....

Using this section of the deciphered plaintext, we can make out Hlower → is most probably flower and Uo is most probably to. So, we get f → H and t → U in encryption.

viii) Now, the decryption we know are

f → H
l → N
o → S
w → F
e → C
r → Y
h → Z
b → D
a → G
t → V

EMGLOSUDCGDNCUSWYSFHNSFCYKDPUMCWGYICOXY ICOX
YSIPTCKQPKUGKMGOLICGINCGACKSNISACYKZSCKX.

EMaLOotbe able to Wrow flowerK bPt MG WarIeS
XroIPJeK QPKtakMPJh deaI leaAek. ...

from this decrypted message we can further identify :-
{S → K, g → W, μ → P, d → I, v → A} in the encryption table

ix) We again decipher the encrypted message :-
EMaLOot be able to grow flowers but MG gardeS
XrodwJes rust as MaOL dead leaves old over
shoe sXEJes of roXe and bushells of dead
grass as anLbodLs and todaL E bought
a wheel barrow to helX in EnJlearEng Et it....

Using the above text we identify words and once again identify the substitutions: {m → M, y → L, P → X, c → J, j → Q

n → O, y → L, i → E, c → J

x) We again decipher the ciphertext

I i may not be able to grow flowers but my garden produces just as many dead leaves old over shoe species of rope and bushells of dead grass as anybodys and today i bought a wheel barrow to help in clearing it up i have always loved and respected the wheel barrow it is the one wheeled vehicle of which i am perfect master.

So, we end up with the following encipherment table:-

P: a b c d e f g h i j k l m n o p q r s t u v w x y z
C: G D J I C H W Z E Q - N M O S X _ Y K U P A F _ L _

c

Here we never see the letters k, q, r and z and hence they can represent any of the following: Q, B, R, T, V