Ministry of Education of Republic of Moldova

Technical University of Moldova

Faculty of Computers, Informatics and Microelectronics

Software Engineering Department

Cryptography And Security

Laboratory work #2

Cryptanalysis of monoalphabetic ciphers.

Author:
Alexandru Rudoi
std. gr. FAF-231

Verified:
Maia Zaica
asist. univ.

Chișinău 2025

Contents

Objective

The objective of this laboratory work is to understand and apply frequency–analysis techniques for breaking monoalphabetic substitution ciphers. Concretely, we will: (i) compute symbol frequencies for a given ciphertext; (ii) compare them to the empirical distribution of letters in English; (iii) iteratively propose and refine a plaintext↔ciphertext mapping using linguistic cues (common digraphs/trigraphs, doubled letters, and short high-frequency words); and (iv) reconstruct the original message and the substitution key. The report documents each step, justifies the substitutions, and presents the recovered plaintext and key alphabet.

Theoretical Background

Monoalphabetic substitution ciphers map each plaintext letter to a unique ciphertext letter via a fixed permutation of the alphabet. While such ciphers obscure individual symbols, they preserve *statistical structure*. Over sufficiently long texts, the frequency of letters in the ciphertext approaches that of the underlying language. By aligning the ciphertext's frequency profile with the known English distribution and validating hypotheses with linguistic patterns (e.g., `THE`, `AND`, doubled letters like `LL`, and one-letter words "A", "I"), a cryptanalyst can progressively recover the substitution.

Letter frequencies in English

Table 1 lists typical letter frequencies for English (percent of occurrence). Figure 2 shows a bar chart that will be used as a visual reference when comparing against the intercepted ciphertext.

A	B	C	D	E	F	G	H	I	J	K	L	M
8,17	*1,49*	*2,78*	*4,25*	*12,7*	*2,23*	*2,01*	*6,09*	*6.97*	*0,15*	*0,77*	*4,03*	*2,41*
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
6,75	*7,51*	*1,93*	*0,09*	*5,99*	*6,33*	*9,06*	*2,76*	*0,98*	*2,36*	*0,15*	*1,97*	*0,07*

Figure 1: Letter frequencies in English.

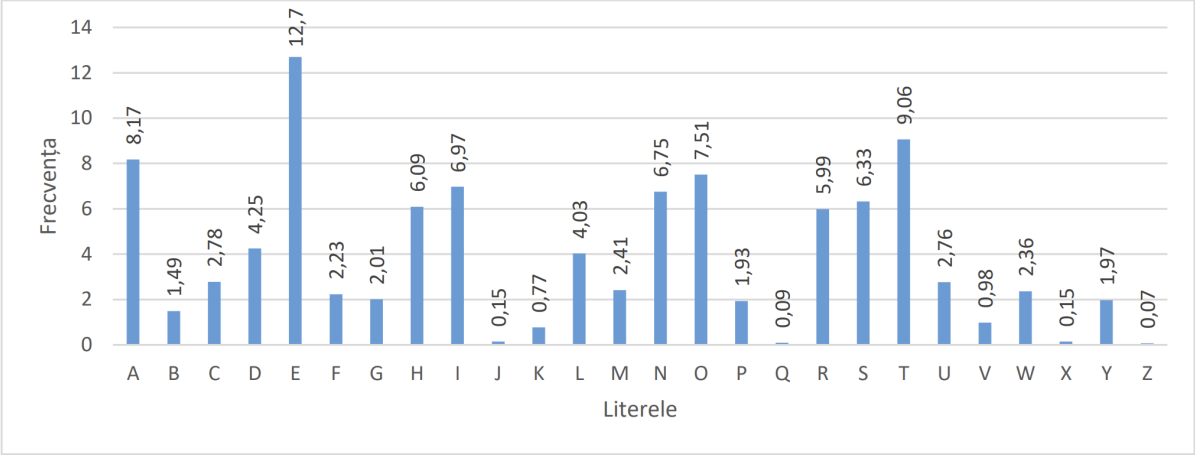

Figure 2: English letter frequency distribution (visual reference).

Frequency Analysis Attack Methodology

We can use information about the frequency of letter occurrences in a language to attempt to break a monoalphabetic substitution cipher. This is possible because, for example, in a message written in English, the letter "E," which has the highest frequency, might be encrypted as "X." In that case, every "X" in the encrypted text would correspond to an "E" in the plaintext. Consequently, the most frequent letter in the encrypted text should be "X."

Thus, if we intercept an encrypted message and the most frequent letter in it is "P," we can assume that "P" was used to encrypt "E," and we can replace all "P"s with "E"s. Of course, not every text has exactly the same frequency, and as noted above, "T" and "A" also have high frequencies, so "P" could represent one of these. However, it is unlikely to be "Z," which is rarely encountered in English. By repeating this process with the next most frequent letter, we can make progress in breaking the message.

If we were to put all the letters in order and replace them according to the frequency table, it is most likely that we would not obtain the expected result. The cryptanalyst must use other "personality traits" of the letters to break the cryptogram. This may include examining pairs of letters (digraphs), the most common of which are TH, HE, AN, IN, ER, ON, RE, ED, ND, HA, AT, and EN. Triplets of letters (trigraphs) can also be very useful, with the most frequent in English being THE, AND, THA, ENT, ION, TIO, FOR, NDE, HAS, NCE, TIS, OFT, and MEN. Additionally, in English, only a few letters appear as doubles (SS, EE, TT, OO, and FF being the most frequent). There are only two meaningful single-letter words in English: "A" and "I."

Other frequent words also begin to emerge as we make some substitutions. For example, "T*E" may appear frequently after performing substitutions for "T" and "E." In this case, "T*E" is very likely to be "THE," a very common word in English.

The process of frequency analysis utilizes various subtle properties of the language,

and for this reason, it is almost impossible for a computer to do all the work. Inevitably, a human element is necessary in this process to make informed decisions about which letters should be replaced.

The Task

An encrypted message has been intercepted, known to have been produced using a monoalphabetic substitution cipher. By applying the frequency–analysis attack, the objective is to determine the original plaintext message, assuming that it was written in English. It is important to note that only the letters were encrypted, while all other characters such as spaces and punctuation marks remain unchanged.

Note: To support the decryption process, the following online tool can be used: Frequency Analysis – Breaking the Code.

My Variant (25 → 2)

```
WQV TOOXWXNG NC PVHIVHF WN WQV WITGPCNIZTWXNGP UINODHVOHIFUWNJITUQF. WIDV,
    XW RTP ZNIV NC T JTZV WQTG TGFWQXGJ VSPVXW PNDJQWWN OVSTF HNZUIVQVGPXNG
    CNI NGSF WQV PQNIWVPW UNPPXASV WXZV, GNW WQVSNGJVPWTGO WQV HIFUWTGTSFPXP
    RTP, SXLVRXPV, EDPW T UDMMSV. VJFUW'P RTPWQDP T BDTPX HIFUWNSNJF XG
    HNGWITPW WN WQV OVTOSF PVIXNDP PHXVGHV NC WNOTF.FVW JIVTW WQXGJP QTKV
    PZTSS AVJXGGXGJP, TGO WQVPV QXVINJSFUQP OXOXGHSDOV, WQNDJQ XG TG
    XZUVICVHW CTPQXNG, WQV WRN VSVZVGWP NC PVHIVHF TGOWITGPCNIZTWXNG WQTW
    HNZUIXPV WQV VPPVGWXTS TWWIXADWVP NC WQV PHXVGHV. TGOPN HIFUWNSNJF RTP
    ANIG. XG XWP CXIPW 3,000 FVTIP, XW OXO GNW JINR PWVTOXSF. HIFUWNSNJF
    TINPVXGOVUVGOVGWSF XG ZTGF USTHVP, TGO XG ZNPW NC WQVZ XW OXVO WQV OVTWQP
     NCXWP HXKXSXMTWXNGP. XG NWQVI USTHVP, XW PDIKXKVO, VZAVOOVO XG T
    SXWVITWDIV,TGO CINZ WQXP WQV GVYW JVGVITWXNG HNDSO HSXZA WN QXJQVI SVKVSP
    .ADW UINJIVPP RTP PSNR TGO EVILF. ZNIV RTP SNPW WQTG IVWTXGVO. ZDHQ NC
    WQVQXPWNIF NC HIFUWNSNJF NC WQXP WXZV XP T UTWHQRNIL, T HITMF BDXSW
    NCDGIVSTWVO XWVZP, PUINDWXGJ, CSNDIXPQXGJ, RXWQVIXGJ. NGSF WNRTIO
    WQVRVPWVIG IVGTXPPTGHV ONVP WQV THHIVWXGJ LGNRSVOJV AVJXG WN ADXSO DU
    TZNZVGWDZ. WQV PWNIF NC HIFUWNSNJF ODIXGJ WQVPV FVTIP XP, XG NWQVI RNIOP,
    VYTHWSF WQV PWNIF NC ZTGLXGO. HQXGT, WQV NGSF QXJQ HXKXSXMTWXNG NC
    TGWXBDXWF WN DPV XOVNJITUQXHRIXWXGJ, PVVZP GVKVI WN QTKV OVKVSNUVO ZDHQ
    IVTS HIFUWNJITUQF UVIQTUP CNI WQTW IVTPNG. XG NGV HTPV LGNRG CNI ZXSXWTIF
     UDIUNPVP, WQV11WQ-HVGWDIF HNZUXSTWXNG, RD-HQXGJ WPDGJ-FTN ("VPPVGWXTSP
    CINZ ZXSXWTIFHSTPPXHP"), IVHNZZVGOVO T WIDV XC PZTSS HNOV. WN T SXPW NC
    40 USTXGWVYWXWVZP, ITGJXGJ CINZ IVBDVPWP CNI ANRP TGO TIINRP WN WQV
```

```
IVUNIW NC TKXHWNIF, WQV HNIIVPUNGOVGWP RNDSO TPPXJG WQV CXIPW 40
XOVNJITZP NC TUNVZ. WQVG, RQVG T SXVDWVGTGW RXPQVO, CNI VYTZUSV, WN
IVBDVPW ZNIVTIINRP, QV RTP WN RIXWV WQV HNIIVPUNGOXGJ XOVNJITZ TW T
PUVHXCXVO USTHVNG TG NIOXGTIF OXPUTWHQ TGO PWTZU QXP PVTS NG XW.XG HQXGT'
P JIVTW GVXJQANI WN WQV RVPW, XGOXT, RQNPV HXKXSXMTWXNGSXLVRXPV OVKVSNUVO
 VTISF TGO WN QXJQ VPWTWV, PVKVITS CNIZP NC PVHIVWHNZZDGXHTWXNGP RVIV
LGNRG TGO, T UUTIVGWSF, UITHWXHVO. WQV TIWQT-PTPWIT, T HSTPPXH RNIL NG
PWTWVHITCW TWWIXADWVO WN LTDWXSFT, XG OVPHIXAXGJWQV VPUXNGTJV PVIKXHV NC
XGOXT TP UITHWXHTSSF IXOOSXGJ WQV HNDGWIF RXWQP UXVP, IVHNZZVGOVO WQTW
WQV NCCXHVIP NC WQV XGPWXWDWVP NC  PUXNGTJV JXKVWQVXI PUXVP WQVXI
TPPXJGZVGWP AF PVHIVW RIXWXGJ.UVIQTUP ZNPW XGWVIVPWXGJ WN HIFUWNSNJXPWP,
TZTWVDI NIUINCVPPXNGTS, XP WQTW KTWPFTFTGT'P CTZNDP WVYWANNL NC VINWXHP,
WQV LTZTPDWIT,SXPWP PVHIVW RIXWXGJ TP NGV NC WQV 64 TIWP, NI FNJTP, WQTW
RNZVGPQNDSO LGNR TGO UITHWXHV. WQV CNDIWQ JIVTW HXKXSXMTWXNG NC TGWXBDXWF
, WQVZVPNUN-WTZXTG, ITWQVI UTITSSVSVO VJFUW VTISF XG XWP
HIFUWNJITUQXHVKNSDWXNG, ADW WQVG PDIUTPPVO XW. WQDP, XG WQV STPW UVIXNO
NC HDGVXCNIZRIXWXGJ, XG HNSNUQNGP RIXWWVG TW DIDL (XG UIVPVGW-OTF XITB)
DGOVI WQVPVSVDHXO LXGJP XG WQV STPW CVR PHNIV FVTIP AVCNIV WQV HQIXPWXTG
VIT,NHHTPXNGTS PHIXAVP HNGKVIWVO WQVXI GTZVP XGWN GDZAVIP.
WQVVGHXUQVIZVGWXC PDHQ XW AVZTF QTKV AVVG NGSF CNI TZDPVZVGW NI WNPQNR
NCC.
```

Listing 1: Ciphertext for Variant 25→2

Technical Implementation

This section documents the practical steps used to break the monoalphabetic substitution in the assigned variant. We begin with letter–frequency analysis of the ciphertext, compare it to the empirical English distribution, then apply iterative substitutions guided by common linguistic patterns (high-frequency words, digraphs/trigraphs, doubled letters, and one-letter words). Each round shows the current mapping and the progressively decoded text.

Step 1 — Frequency analysis of the ciphertext

Using the provided tool, I compute the frequency of each symbol in the intercepted text (Variant: 25 → 2).

V	W	X	N	P	T	I	G	Q	H	S	O	U	F	Z	D	J	C	R	A	K	L	B	M	Y	E
273	250	201	195	188	185	169	165	111	89	84	79	66	64	61	59	59	56	42	21	20	15	7	7	5	2
11.0	10.1	8.1	7.9	7.6	7.5	6.8	6.7	4.5	3.6	3.4	3.2	2.7	2.6	2.5	2.4	2.4	2.3	1.7	0.8	0.8	0.6	0.3	0.3	0.2	0.1

Figure 3: Letter frequencies in the intercepted ciphertext (Variant 25→2).

Step 2 — Reference distribution (English)

We compare the ciphertext profile against the typical English letter distribution.

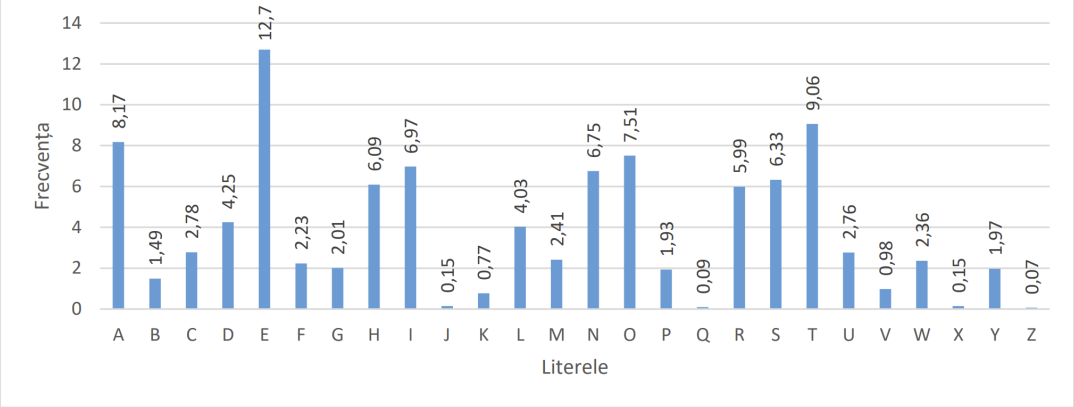

Figure 4: English letter frequency distribution (reference).

Step 3 — Initial anchors from top frequencies

The most frequent ciphertext letters suggest candidates for E and T. Following the observed counts, we tentatively assign:

$$\texttt{V} \rightarrow \texttt{e}, \qquad \texttt{W} \rightarrow \texttt{t}.$$

Known mappings so far: `V→e, W→t`

```
tQe TOOXtXNG NC PeHIeHF tN tQe tITGPCNIZTtXNGP UINODHeOHIFUtNJITUQF. tIDe,
    Xt RTP ZNIe NC T JTZe tQTG TGFtQXGJ eSPeXt PNDJQttN OeSTF HNZUIeQeGPXNG
    CNI NGSF tQe PQNItePt UNPPXASe tXZe, GNt tQeSNGJePtTGO tQe HIFUtTGTSFPXP
    RTP, SXLeRXPe, EDPt T UDMMSe. eJFUt'P RTPtQDP T BDTPX HIFUtNSNJF XG
    HNGtITPt tN tQe OeTOSF PeIXNDP PHXeGHe NC tNOTF.Fet JIeTt tQXGJP QTKe
    PZTSS AeJXGGXGJP, TGO tQePe QXeINJSFUQP OXOXGHSDOe, tQNDJQ XG TG
    XZUeICeHt CTPQXNG, tQe tRN eSeZeGtP NC PeHIeHF TGOtITGPCNIZTtXNG tQTt
    HNZUIXPe tQe ePPeGtXTS TttIXADteP NC tQe PHXeGHe. TGOPN HIFUtNSNJF RTP
    ANIG. XG XtP CXIPt 3,000 FeTIP, Xt OXO GNt JINR PteTOXSF. HIFUtNSNJF
    TINPeXGOeUeGOeGtSF XG ZTGF USTHeP, TGO XG ZNPt NC tQeZ Xt OXeO tQe OeTtQP
     NCXtP HXKXSXMTtXNGP. XG NtQeI USTHeP, Xt PDIKXKeO, eZAeOOeO XG T
```

```
SXteITtDIe,TGO CINZ tQXP tQe GeYt JeGeITtXNG HNDSO HSXZA tN QXJQeI SeKeSP
.ADt UINJIePP RTP PSNR TGO EeILF. ZNIe RTP SNPt tQTG IetTXGeO. ZDHQ NC
tQeQXPtNIF NC HIFUtNSNJF NC tQXP tXZe XP T UTtHQRNIL, T HITMF BDXSt
NCDGIeSTteO XteZP, PUINDtXGJ, CSNDIXPQXGJ, RXtQeIXGJ. NGSF tNRTIO
tQeRePteIG IeGTXPPTGHe ONeP tQe THHIetXGJ LGNRSeOJe AeJXG tN ADXSO DU
TZNZeGtDZ. tQe PtNIF NC HIFUtNSNJF ODIXGJ tQePe FeTIP XP, XG NtQeI RNIOP,
eYTHtSF tQe PtNIF NC ZTGLXGO. HQXGT, tQe NGSF QXJQ HXKXSXMTtXNG NC
TGtXBDXtF tN DPe XOeNJITUQXHRIXtXGJ, PeeZP GeKeI tN QTKe OeKeSNUeO ZDHQ
IeTS HIFUtNJITUQF UeIQTUP CNI tQTt IeTPNG. XG NGe HTPe LGNRG CNI ZXSXtTIF
 UDIUNPeP, tQe11tQ-HeGtDIF HNZUXSTtXNG, RD-HQXGJ tPDGJ-FTN ("ePPeGtXTSP
CINZ ZXSXtTIFHSTPPXHP"), IeHNZZeGOeO T tIDe XC PZTSS HNOe. tN T SXPt NC
40 USTXGteYtXteZP, ITGJXGJ CINZ IeBDePtP CNI ANRP TGO TIINRP tN tQe
IeUNIt NC TKXHtNIF, tQe HNIIePUNGOeGtP RNDSO TPPXJG tQe CXIPt 40
XOeNJITZP NC TUNeZ. tQeG, RQeG T SXeDteGTGt RXPQeO, CNI eYTZUSe, tN
IeBDePt ZNIeTIINRP, Qe RTP tN RIXte tQe HNIIePUNGOXGJ XOeNJITZ Tt T
PUeHXCXeO USTHeNG TG NIOXGTIF OXPUTtHQ TGO PtTZU QXP PeTS NG Xt.XG HQXGT'
P JIeTt GeXJQANI tN tQe RePt, XGOXT, RQNPe HXKXSXMTtXNGSXLeRXPe OeKeSNUeO
 eTISF TGO tN QXJQ ePtTte, PeKeITS CNIZP NC PeHIetHNZZDGXHTtXNGP ReIe
LGNRG TGO, T UUTIeGtSF, UITHtXHeO. tQe TItQT-PTPtIT, T HSTPPXH RNIL NG
PtTteHITCt TttIXADteO tN LTDtXSFT, XG OePHIXAXGJtQe ePUXNGTJe PeIKXHe NC
XGOXT TP UITHtXHTSSF IXOOSXGJ tQe HNDGtIF RXtQP UXeP, IeHNZZeGOeO tQTt
tQe NCCXHeIP NC tQe XGPtXtDteP NC  PUXNGTJe JXKetQeXI PUXeP tQeXI
TPPXJGZeGtP AF PeHIet RIXtXGJ.UeIQTUP ZNPt XGteIePtXGJ tN HIFUtNSNJXPtP,
TZTteDI NIUINCePPXNGTS, XP tQTt KTtPFTFTGT'P CTZNDP teYtANNL NC eINtXHP,
tQe LTZTPDtIT,SXPtP PeHIet RIXtXGJ TP NGe NC tQe 64 TItP, NI FNJTP, tQTt
RNZeGPQNDSO LGNR TGO UITHtXHe. tQe CNDItQ JIeTt HXKXSXMTtXNG NC TGtXBDXtF
, tQeZePNUN-tTZXTG, ITtQeI UTITSSeSeO eJFUt eTISF XG XtP
HIFUtNJITUQXHeKNSDtXNG, ADt tQeG PDIUTPPeO Xt. tQDP, XG tQe STPt UeIXNO
NC HDGeXCNIZRIXtXGJ, XG HNSNUQNGP RIXtteG Tt DIDL (XG UIePeGt-OTF XITB)
DGOeI tQePeSeDHXO LXGJP XG tQe STPt CeR PHNIe FeTIP AeCNIe tQe HQIXPtXTG
eIT,NHHTPXNGTS PHIXAeP HNGKeIteO tQeXI GTZeP XGtN GDZAeIP.
tQeeGHXUQeIZeGtXC PDHQ Xt AeZTF QTKe AeeG NGSF CNI TZDPeZeGt NI tNPQNR
NCC.
```

Listing 2: Round 1 — after V→e, W→t

Step 4 — Locking THE from the 3-letter pattern

The 3-letter pattern `tQe` appears frequently and matches the English word `the`, which fixes:

$$\texttt{Q} \rightarrow \texttt{h}.$$

Known mappings so far: `V→e, W→t, Q→h`

```
the TOOXtXNG NC PeHIeHF tN the tITGPCNIZTtXNGP UINODHeOHIFUtNJITUhF. tIDe,
    Xt RTP ZNIe NC T JTZe thTG TGFthXGJ eSPeXt PNDJhttN OeSTF HNZUIeheGPXNG
    CNI NGSF the PhNItePt UNPPXASe tXZe, GNt theSNGJePtTGO the HIFUtTGTSFPXP
    RTP, SXLeRXPe, EDPt T UDMMSe. eJFUt'P RTPthDP T BDTPX HIFUtNSNJF XG
    HNGtITPt tN the OeTOSF PeIXNDP PHXeGHe NC tNOTF.Fet JIeTt thXGJP hTKe
    PZTSS AeJXGGXGJP, TGO thePe hXeINJSFUhP OXOXGHSDOe, thNDJh XG TG
    XZUeICeHt CTPhXNG, the tRN eSeZeGtP NC PeHIeHF TGOtITGPCNIZTtXNG thTt
    HNZUIXPe the ePPeGtXTS TttIXADteP NC the PHXeGHe. TGOPN HIFUtNSNJF RTP
    ANIG. XG XtP CXIPt 3,000 FeTIP, Xt OXO GNt JINR PteTOXSF. HIFUtNSNJF
    TINPeXGOeUeGOeGtSF XG ZTGF USTHeP, TGO XG ZNPt NC theZ Xt OXeO the OeTthP
     NCXtP HXKXSXMTtXNGP. XG NtheI USTHeP, Xt PDIKXKeO, eZAeOOeO XG T
    SXteITtDIe,TGO CINZ thXP the GeYt JeGeITtXNG HNDSO HSXZA tN hXJheI SeKeSP
    .ADt UINJIePP RTP PSNR TGO EeILF. ZNIe RTP SNPt thTG IetTXGeO. ZDHh NC
    thehXPtNIF NC HIFUtNSNJF NC thXP tXZe XP T UTtHhRNIL, T HITMF BDXSt
    NCDGIeSTteO XteZP, PUINDtXGJ, CSNDIXPhXGJ, RXtheIXGJ. NGSF tNRTIO
    theRePteIG IeGTXPPTGHe ONeP the THHIetXGJ LGNRSeOJe AeJXG tN ADXSO DU
    TZNZeGtDZ. the PtNIF NC HIFUtNSNJF ODIXGJ thePe FeTIP XP, XG NtheI RNIOP,
    eYTHtSF the PtNIF NC ZTGLXGO. HhXGT, the NGSF hXJh HXKXSXMTtXNG NC
    TGtXBDXtF tN DPe XOeNJITUhXHRIXtXGJ, PeeZP GeKeI tN hTKe OeKeSNUeO ZDHh
    IeTS HIFUtNJITUhF UeIhTUP CNI thTt IeTPNG. XG NGe HTPe LGNRG CNI ZXSXtTIF
     UDIUNPeP, the11th-HeGtDIF HNZUXSTtXNG, RD-HhXGJ tPDGJ-FTN ("ePPeGtXTSP
    CINZ ZXSXtTIFHSTPPXHP"), IeHNZZeGOeO T tIDe XC PZTSS HNOe. tN T SXPt NC
    40 USTXGteYtXteZP, ITGJXGJ CINZ IeBDePtP CNI ANRP TGO TIINRP tN the
    IeUNIt NC TKXHtNIF, the HNIIePUNGOeGtP RNDSO TPPXJG the CXIPt 40
    XOeNJITZP NC TUNeZ. theG, RheG T SXeDteGTGt RXPheO, CNI eYTZUSe, tN
    IeBDePt ZNIeTIINRP, he RTP tN RIXte the HNIIePUNGOXGJ XOeNJITZ Tt T
    PUeHXCXeO USTHeNG TG NIOXGTIF OXPUTtHh TGO PtTZU hXP PeTS NG Xt.XG HhXGT'
    P JIeTt GeXJhANI tN the RePt, XGOXT, RhNPe HXKXSXMTtXNGSXLeRXPe OeKeSNUeO
     eTISF TGO tN hXJh ePtTte, PeKeITS CNIZP NC PeHIetHNZZDGXHTtXNGP ReIe
    LGNRG TGO, T UUTIeGtSF, UITHtXHeO. the TIthT-PTPtIT, T HSTPPXH RNIL NG
    PtTteHITCt TttIXADteO tN LTDtXSFT, XG OePHIXAXGJthe ePUXNGTJe PeIKXHe NC
    XGOXT TP UITHtXHTSSF IXOOSXGJ the HNDGtIF RXthP UXeP, IeHNZZeGOeO thTt
    the NCCXHeIP NC the XGPtXtDteP NC  PUXNGTJe JXKetheXI PUXeP theXI
    TPPXJGZeGtP AF PeHIet RIXtXGJ.UeIhTUP ZNPt XGteIePtXGJ tN HIFUtNSNJXPtP,
    TZTteDI NIUINCePPXNGTS, XP thTt KTtPFTFTGT'P CTZNDP teYtANNL NC eINtXHP,
    the LTZTPDtIT,SXPtP PeHIet RIXtXGJ TP NGe NC the 64 TItP, NI FNJTP, thTt
    RNZeGPhNDSO LGNR TGO UITHtXHe. the CNDIth JIeTt HXKXSXMTtXNG NC TGtXBDXtF
    , theZePNUN-tTZXTG, ITtheI UTITSSeSeO eJFUt eTISF XG XtP
    HIFUtNJITUhXHeKNSDtXNG, ADt theG PDIUTPPeO Xt. thDP, XG the STPt UeIXNO
    NC HDGeXCNIZRIXtXGJ, XG HNSNUhNGP RIXtteG Tt DIDL (XG UIePeGt-OTF XITB)
```

```
DGOeI thePeSeDHXO LXGJP XG the STPt CeR PHNIe FeTIP AeCNIe the HhIXPtXTG
eIT,NHHTPXNGTS PHIXAeP HNGKeIteO theXI GTZeP XGtN GDZAeIP.
theeGHXUheIZeGtXC PDHh Xt AeZTF hTKe AeeG NGSF CNI TZDPeZeGt NI tNPhNR
NCC.
```

Listing 3: Round 2 — after Q→h

Step 5 — Short common words (to / it) and single-letter words (I)

From two-letter tokens (`tN`, `Xt`), we test `to` and `it`. This yields:

$$\texttt{N} \to \texttt{o}, \qquad \texttt{X} \to \texttt{i}.$$

Why `it` and not `at`? Because I already tried it, and after some steps I stopped and some confusing new words.

Known mappings so far: `V→e, W→t, Q→h, N→o, X→i`

```
the TOOitioG oC PeHIeHF to the tITGPCoIZTtioGP UIoODHeOHIFUtoJITUhF. tIDe,
    it RTP ZoIe oC T JTZe thTG TGFthiGJ eSPeit PoDJhtto OeSTF HoZUIeheGPioG
    CoI oGSF the PhoItePt UoPPiASe tiZe, Got theSoGJePtTGO the HIFUtTGTSFPiP
    RTP, SiLeRiPe, EDPt T UDMMSe. eJFUt'P RTPthDP T BDTPi HIFUtoSoJF iG
    HoGtITPt to the OeTOSF PeIioDP PHieGHe oC toOTF.Fet JIeTt thiGJP hTKe
    PZTSS AeJiGGiGJP, TGO thePe hieIoJSFUhP OiOiGHSDOe, thoDJh iG TG
    iZUeICeHt CTPhioG, the tRo eSeZeGtP oC PeHIeHF TGOtITGPCoIZTtioG thTt
    HoZUIiPe the ePPeGtiTS TttIiADteP oC the PHieGHe. TGOPo HIFUtoSoJF RTP
    AoIG. iG itP CiIPt 3,000 FeTIP, it OiO Got JIoR PteTOiSF. HIFUtoSoJF
    TIoPeiGOeUeGOeGtSF iG ZTGF USTHeP, TGO iG ZoPt oC theZ it OieO the OeTthP
     oCitP HiKiSiMTtioGP. iG otheI USTHeP, it PDIKiKeO, eZAeOOeO iG T
    SiteITtDIe,TGO CIoZ thiP the GeYt JeGeITtioG HoDSO HSiZA to hiJheI SeKeSP
    .ADt UIoJIePP RTP PSoR TGO EeILF. ZoIe RTP SoPt thTG IetTiGeO. ZDHh oC
    thehiPtoIF oC HIFUtoSoJF oC thiP tiZe iP T UTtHhRoIL, T HITMF BDiSt
    oCDGIeSTteO iteZP, PUIoDtiGJ, CSoDIiPhiGJ, RitheIiGJ. oGSF toRTIO
    theRePteIG IeGTiPPTGHe OoeP the THHIetiGJ LGoRSeOJe AeJiG to ADiSO DU
    TZoZeGtDZ. the PtoIF oC HIFUtoSoJF ODIiGJ thePe FeTIP iP, iG otheI RoIOP,
    eYTHtSF the PtoIF oC ZTGLiGO. HhiGT, the oGSF hiJh HiKiSiMTtioG oC
    TGtiBDitF to DPe iOeoJITUhiHRIitiGJ, PeeZP GeKeI to hTKe OeKeSoUeO ZDHh
    IeTS HIFUtoJITUhF UeIhTUP CoI thTt IeTPoG. iG oGe HTPe LGoRG CoI ZiSitTIF
     UDIUoPeP, the11th-HeGtDIF HoZUiSTtioG, RD-HhiGJ tPDGJ-FTo ("ePPeGtiTSP
    CIoZ ZiSitTIFHSTPPiHP"), IeHoZZeGOeO T tIDe iC PZTSS HoOe. to T SiPt oC
    40 USTiGteYtiteZP, ITGJiGJ CIoZ IeBDePtP CoI AoRP TGO TIIoRP to the
    IeUoIt oC TKiHtoIF, the HoIIePUoGOeGtP RoDSO TPPiJG the CiIPt 40
```

```
    iOeoJITZP oC TUoeZ. theG, RheG T SieDteGTGt RiPheO, CoI eYTZUSe, to
    IeBDePt ZoIeTIIoRP, he RTP to RIite the HoIIePUoGOiGJ iOeoJITZ Tt T
    PUeHiCieO USTHeoG TG oIOiGTIF OiPUTtHh TGO PtTZU hiP PeTS oG it.iG HhiGT'
    P JIeTt GeiJhAoI to the RePt, iGOiT, RhoPe HiKiSiMTtioGSiLeRiPe OeKeSoUeO
     eTISF TGO to hiJh ePtTte, PeKeITS CoIZP oC PeHIetHoZZDGiHTtioGP ReIe
    LGoRG TGO, T UUTIeGtSF, UITHtiHeO. the TIthT-PTPtIT, T HSTPPiH RoIL oG
    PtTteHITCt TttIiADteO to LTDtiSFT, iG OePHIiAiGJthe ePUioGTJe PeIKiHe oC
    iGOiT TP UITHtiHTSSF IiOOSiGJ the HoDGtIF RithP UieP, IeHoZZeGOeO thTt
    the oCCiHeIP oC the iGPtitDteP oC  PUioGTJe JiKetheiI PUieP theiI
    TPPiJGZeGtP AF PeHIet RIitiGJ.UeIhTUP ZoPt iGteIePtiGJ to HIFUtoSoJiPtP,
    TZTteDI oIUIoCePPioGTS, iP thTt KTtPFTFTGT'P CTZoDP teYtAooL oC eIotiHP,
    the LTZTPDtIT,SiPtP PeHIet RIitiGJ TP oGe oC the 64 TItP, oI FoJTP, thTt
    RoZeGPhoDSO LGoR TGO UITHtiHe. the CoDIth JIeTt HiKiSiMTtioG oC TGtiBDitF
    , theZePoUo-tTZiTG, ITtheI UTITSSeSeO eJFUt eTISF iG itP
    HIFUtoJITUhiHeKoSDtioG, ADt theG PDIUTPPeO it. thDP, iG the STPt UeIioO
    oC HDGeiCoIZRIitiGJ, iG HoSoUhoGP RIitteG Tt DIDL (iG UIePeGt-OTF iITB)
    DGOeI thePeSeDHiO LiGJP iG the STPt CeR PHoIe FeTIP AeCoIe the HhIiPtiTG
    eIT,oHHTPioGTS PHIiAeP HoGKeIteO theiI GTZeP iGto GDZAeIP.
    theeGHiUheIZeGtiC PDHh it AeZTF hTKe AeeG oGSF CoI TZDPeZeGt oI toPhoR
    oCC.
```

Listing 4: Round 3 — after N→o, X→i

Step 6 — Single-letter word "a"

In English, the two one-letter words are `a` and `I`. Because we already fixed `X` → `i`, the standalone ciphertext symbol `T` must map to `a`:

$$\texttt{T} \rightarrow \texttt{a}.$$

Known mappings so far: `V→e`, `W→t`, `Q→h`, `N→o`, `X→i`, `T→a`

```
 the aOOitioG oC PeHIeHF to the tIaGPCoIZatioGP UIoODHeOHIFUtoJIaUhF. tIDe,
    it RaP ZoIe oC a JaZe thaG aGFthiGJ eSPeit PoDJhtto OeSaF HoZUIeheGPioG
    CoI oGSF the PhoItePt UoPPiASe tiZe, Got theSoGJePtaGO the HIFUtaGaSFPiP
    RaP, SiLeRiPe, EDPt a UDMMSe. eJFUt'P RaPthDP a BDaPi HIFUtoSoJF iG
    HoGtIaPt to the OeaOSF PeIioDP PHieGHe oC toOaF.Fet JIeat thiGJP haKe
    PZaSS AeJiGGiGJP, aGO thePe hieIoJSFUhP OiOiGHSDOe, thoDJh iG aG
    iZUeICeHt CaPhioG, the tRo eSeZeGtP oC PeHIeHF aGOtIaGPCoIZatioG that
    HoZUIiPe the ePPeGtiaS attIiADteP oC the PHieGHe. aGOPo HIFUtoSoJF RaP
    AoIG. iG itP CiIPt 3,000 FeaIP, it OiO Got JIoR PteaOiSF. HIFUtoSoJF
    aIoPeiGOeUeGOeGtSF iG ZaGF USaHeP, aGO iG ZoPt oC theZ it OieO the OeathP
     oCitP HiKiSiMatioGP. iG otheI USaHeP, it PDIKiKeO, eZAeOOeO iG a
```

```
SiteIatDIe,aGO CIoZ thiP the GeYt JeGeIatioG HoDSO HSiZA to hiJheI SeKeSP
.ADt UIoJIePP RaP PSoR aGO EeILF. ZoIe RaP SoPt thaG IetaiGeO. ZDHh oC
thehiPtoIF oC HIFUtoSoJF oC thiP tiZe iP a UatHhRoIL, a HIaMF BDiSt
oCDGIeSateO iteZP, PUIoDtiGJ, CSoDIiPhiGJ, RitheIiGJ. oGSF toRaIO
theRePteIG IeGaiPPaGHe OoeP the aHHIetiGJ LGoRSeOJe AeJiG to ADiSO DU
aZoZeGtDZ. the PtoIF oC HIFUtoSoJF ODIiGJ thePe FeaIP iP, iG otheI RoIOP,
eYaHtSF the PtoIF oC ZaGLiGO. HhiGa, the oGSF hiJh HiKiSiMatioG oC
aGtiBDitF to DPe iOeoJIaUhiHRIitiGJ, PeeZP GeKeI to haKe OeKeSoUeO ZDHh
IeaS HIFUtoJIaUhF UeIhaUP CoI that IeaPoG. iG oGe HaPe LGoRG CoI ZiSitaIF
 UDIUoPeP, the11th-HeGtDIF HoZUiSatioG, RD-HhiGJ tPDGJ-Fao ("ePPeGtiaSP
CIoZ ZiSitaIFHSaPPiHP"), IeHoZZeGOeO a tIDe iC PZaSS HoOe. to a SiPt oC
40 USaiGteYtiteZP, IaGJiGJ CIoZ IeBDePtP CoI AoRP aGO aIIoRP to the
IeUoIt oC aKiHtoIF, the HoIIePUoGOeGtP RoDSO aPPiJG the CiIPt 40
iOeoJIaZP oC aUoeZ. theG, RheG a SieDteGaGt RiPheO, CoI eYaZUSe, to
IeBDePt ZoIeaIIoRP, he RaP to RIite the HoIIePUoGOiGJ iOeoJIaZ at a
PUeHiCieO USaHeoG aG oIOiGaIF OiPUatHh aGO PtaZU hiP PeaS oG it.iG HhiGa'
P JIeat GeiJhAoI to the RePt, iGOia, RhoPe HiKiSiMatioGSiLeRiPe OeKeSoUeO
 eaISF aGO to hiJh ePtate, PeKeIaS CoIZP oC PeHIetHoZZDGiHatioGP ReIe
LGoRG aGO, a UUaIeGtSF, UIaHtiHeO. the aItha-PaPtIa, a HSaPPiH RoIL oG
PtateHIaCt attIiADteO to LaDtiSFa, iG OePHIiAiGJthe ePUioGaJe PeIKiHe oC
iGOia aP UIaHtiHaSSF IiOOSiGJ the HoDGtIF RithP UieP, IeHoZZeGOeO that
the oCCiHeIP oC the iGPtitDteP oC  PUioGaJe JiKetheiI PUieP theiI
aPPiJGZeGtP AF PeHIet RIitiGJ.UeIhaUP ZoPt iGteIePtiGJ to HIFUtoSoJiPtP,
aZateDI oIUIoCePPioGaS, iP that KatPFaFaGa'P CaZoDP teYtAooL oC eIotiHP,
the LaZaPDtIa,SiPtP PeHIet RIitiGJ aP oGe oC the 64 aItP, oI FoJaP, that
RoZeGPhoDSO LGoR aGO UIaHtiHe. the CoDIth JIeat HiKiSiMatioG oC aGtiBDitF
, theZePoUo-taZiaG, IatheI UaIaSSeSeO eJFUt eaISF iG itP
HIFUtoJIaUhiHeKoSDtioG, ADt theG PDIUaPPeO it. thDP, iG the SaPt UeIioO
oC HDGeiCoIZRIitiGJ, iG HoSoUhoGP RIitteG at DIDL (iG UIePeGt-OaF iIaB)
DGOeI thePeSeDHiO LiGJP iG the SaPt CeR PHoIe FeaIP AeCoIe the HhIiPtiaG
eIa,oHHaPioGaS PHIiAeP HoGKeIteO theiI GaZeP iGto GDZAeIP.
theeGHiUheIZeGtiC PDHh it AeZaF haKe AeeG oGSF CoI aZDPeZeGt oI toPhoR
oCC.
```

Listing 5: Round 4 — after T→a

Step 7 — High-frequency small words (or, off, is)

Examine candidates like `oI` (`on` if `I`→`n`), `oCC` (`off` if `C`→`f`), and the verb `is` (`P`→`s`). But after I choice `I` → `n` led to contradictions in subsequent words and stalled

progress, so I correct it by re-examining short tokens showed that `oI` fits `or` (not `on`):

$$\texttt{I} \to \texttt{r}, \quad \texttt{C} \to \texttt{f}\ (\texttt{oCC} \Rightarrow \texttt{off}), \quad \texttt{P} \to \texttt{s}\ (\texttt{is}).$$

Known mappings so far: `V→e, W→t, Q→h, N→o, X→i, T→a, I→r, C→f, P→s`

```
the aOOitioG of seHreHF to the traGsforZatioGs UroODHeOHrFUtoJraUhF. trDe,
    it Ras Zore of a JaZe thaG aGFthiGJ eSseit soDJhtto OeSaF HoZUreheGsioG
    for oGSF the shortest UossiASe tiZe, Got theSoGJestaGO the HrFUtaGaSFsis
    Ras, SiLeRise, EDst a UDMMSe. eJFUt's RasthDs a BDasi HrFUtoSoJF iG
    HoGtrast to the OeaOSF serioDs sHieGHe of toOaF.Fet Jreat thiGJs haKe
    sZaSS AeJiGGiGJs, aGO these hieroJSFUhs OiOiGHSDOe, thoDJh iG aG
    iZUerfeHt fashioG, the tRo eSeZeGts of seHreHF aGOtraGsforZatioG that
    HoZUrise the esseGtiaS attriADtes of the sHieGHe. aGOso HrFUtoSoJF Ras
    AorG. iG its first 3,000 Fears, it OiO Got JroR steaOiSF. HrFUtoSoJF
    aroseiGOeUeGOeGtSF iG ZaGF USaHes, aGO iG Zost of theZ it OieO the Oeaths
     ofits HiKiSiMatioGs. iG other USaHes, it sDrKiKeO, eZAeOOeO iG a
    SiteratDre,aGO froZ this the GeYt JeGeratioG HoDSO HSiZA to hiJher SeKeSs
    .ADt UroJress Ras sSoR aGO EerLF. Zore Ras Sost thaG retaiGeO. ZDHh of
    thehistorF of HrFUtoSoJF of this tiZe is a UatHhRorL, a HraMF BDiSt
    ofDGreSateO iteZs, sUroDtiGJ, fSoDrishiGJ, RitheriGJ. oGSF toRarO
    theResterG reGaissaGHe Ooes the aHHretiGJ LGoRSeOJe AeJiG to ADiSO DU
    aZoZeGtDZ. the storF of HrFUtoSoJF ODriGJ these Fears is, iG other RorOs,
    eYaHtSF the storF of ZaGLiGO. HhiGa, the oGSF hiJh HiKiSiMatioG of
    aGtiBDitF to Dse iOeoJraUhiHRritiGJ, seeZs GeKer to haKe OeKeSoUeO ZDHh
    reaS HrFUtoJraUhF UerhaUs for that reasoG. iG oGe Hase LGoRG for ZiSitarF
     UDrUoses, the11th-HeGtDrF HoZUiSatioG, RD-HhiGJ tsDGJ-Fao ("esseGtiaSs
    froZ ZiSitarFHSassiHs"), reHoZZeGOeO a trDe if sZaSS HoOe. to a Sist of
    40 USaiGteYtiteZs, raGJiGJ froZ reBDests for AoRs aGO arroRs to the
    reUort of aKiHtorF, the HorresUoGOeGts RoDSO assiJG the first 40
    iOeoJraZs of aUoeZ. theG, RheG a SieDteGaGt RisheO, for eYaZUSe, to
    reBDest ZorearroRs, he Ras to Rrite the HorresUoGOiGJ iOeoJraZ at a
    sUeHifieO USaHeoG aG orOiGarF OisUatHh aGO staZU his seaS oG it.iG HhiGa'
    s Jreat GeiJhAor to the Rest, iGOia, Rhose HiKiSiMatioGSiLeRise OeKeSoUeO
     earSF aGO to hiJh estate, seKeraS forZs of seHretHoZZDGiHatioGs Rere
    LGoRG aGO, a UUareGtSF, UraHtiHeO. the artha-sastra, a HSassiH RorL oG
    stateHraft attriADteO to LaDtiSFa, iG OesHriAiGJthe esUioGaJe serKiHe of
    iGOia as UraHtiHaSSF riOOSiGJ the HoDGtrF Riths Uies, reHoZZeGOeO that
    the offiHers of the iGstitDtes of  sUioGaJe JiKetheir sUies their
    assiJGZeGts AF seHret RritiGJ.UerhaUs Zost iGterestiGJ to HrFUtoSoJists,
    aZateDr orUrofessioGaS, is that KatsFaFaGa's faZoDs teYtAooL of erotiHs,
    the LaZasDtra,Sists seHret RritiGJ as oGe of the 64 arts, or FoJas, that
    RoZeGshoDSO LGoR aGO UraHtiHe. the foDrth Jreat HiKiSiMatioG of aGtiBDitF
```

```
, theZesoUo-taZiaG, rather UaraSSeSeO eJFUt earSF iG its
HrFUtoJraUhiHeKoSDtioG, ADt theG sDrUasseO it. thDs, iG the Sast UerioO
of HDGeiforZRritiGJ, iG HoSoUhoGs RritteG at DrDL (iG UreseGt-OaF iraB)
DGOer theseSeDHiO LiGJs iG the Sast feR sHore Fears Aefore the HhristiaG
era,oHHasioGaS sHriAes HoGKerteO their GaZes iGto GDZAers.
theeGHiUherZeGtif sDHh it AeZaF haKe AeeG oGSF for aZDseZeGt or toshoR
off.
```

Listing 6: Round 5 — after I→r, C→f, P→s

Step 8 — Medium-frequency anchors from context ("true", "addition")

Two clear word shapes resolve the next letters. The token `trDe` matches `true`, fixing

$$\texttt{D} \rightarrow \texttt{u}.$$

Likewise, `aOOitioG` match `addition`, which yields the doubled-`d` and final `n`:

$$\texttt{O} \rightarrow \texttt{d}, \qquad \texttt{G} \rightarrow \texttt{n}.$$

These updates remove earlier inconsistencies and make surrounding phrases read naturally.

Known mappings so far: `V→e`, `W→t`, `Q→h`, `N→o`, `X→i`, `T→a`, `I→r`, `C→f`, `P→s`, `D→u`, `O→d`, `G→n`

```
the addition of seHreHF to the transforZations UroduHedHrFUtoJraUhF. true,
    it Ras Zore of a JaZe than anFthinJ eSseit souJhtto deSaF HoZUrehension
    for onSF the shortest UossiASe tiZe, not theSonJestand the HrFUtanaSFsis
    Ras, SiLeRise, Eust a UuMMSe. eJFUt's Rasthus a Buasi HrFUtoSoJF in
    Hontrast to the deadSF serious sHienHe of todaF.Fet Jreat thinJs haKe
    sZaSS AeJinninJs, and these hieroJSFUhs didinHSude, thouJh in an
    iZUerfeHt fashion, the tRo eSeZents of seHreHF andtransforZation that
    HoZUrise the essentiaS attriAutes of the sHienHe. andso HrFUtoSoJF Ras
    Aorn. in its first 3,000 Fears, it did not JroR steadiSF. HrFUtoSoJF
    aroseindeUendentSF in ZanF USaHes, and in Zost of theZ it died the deaths
     ofits HiKiSiMations. in other USaHes, it surKiKed, eZAedded in a
    Siterature,and froZ this the neYt Jeneration HouSd HSiZA to hiJher SeKeSs
    .Aut UroJress Ras sSoR and EerLF. Zore Ras Sost than retained. ZuHh of
    thehistorF of HrFUtoSoJF of this tiZe is a UatHhRorL, a HraMF BuiSt
    ofunreSated iteZs, sUroutinJ, fSourishinJ, RitherinJ. onSF toRard
    theRestern renaissanHe does the aHHretinJ LnoRSedJe AeJin to AuiSd uU
    aZoZentuZ. the storF of HrFUtoSoJF durinJ these Fears is, in other Rords,
```

```
eYaHtSF the storF of ZanLind. Hhina, the onSF hiJh HiKiSiMation of
antiBuitF to use ideoJraUhiHRritinJ, seeZs neKer to haKe deKeSoUed ZuHh
reaS HrFUtoJraUhF UerhaUs for that reason. in one Hase LnoRn for ZiSitarF
 UurUoses, the11th-HenturF HoZUiSation, Ru-HhinJ tsunJ-Fao ("essentiaSs
froZ ZiSitarFHSassiHs"), reHoZZended a true if sZaSS Hode. to a Sist of
40 USainteYtiteZs, ranJinJ froZ reBuests for AoRs and arroRs to the
reUort of aKiHtorF, the HorresUondents RouSd assiJn the first 40
ideoJraZs of aUoeZ. then, Rhen a Sieutenant Rished, for eYaZUSe, to
reBuest ZorearroRs, he Ras to Rrite the HorresUondinJ ideoJraZ at a
sUeHified USaHeon an ordinarF disUatHh and staZU his seaS on it.in Hhina'
s Jreat neiJhAor to the Rest, india, Rhose HiKiSiMationSiLeRise deKeSoUed
 earSF and to hiJh estate, seKeraS forZs of seHretHoZZuniHations Rere
LnoRn and, a UUarentSF, UraHtiHed. the artha-sastra, a HSassiH RorL on
stateHraft attriAuted to LautiSFa, in desHriAinJthe esUionaJe serKiHe of
india as UraHtiHaSSF riddSinJ the HountrF Riths Uies, reHoZZended that
the offiHers of the institutes of  sUionaJe JiKetheir sUies their
assiJnZents AF seHret RritinJ.UerhaUs Zost interestinJ to HrFUtoSoJists,
aZateur orUrofessionaS, is that KatsFaFana's faZous teYtAooL of erotiHs,
the LaZasutra,Sists seHret RritinJ as one of the 64 arts, or FoJas, that
RoZenshouSd LnoR and UraHtiHe. the fourth Jreat HiKiSiMation of antiBuitF
, theZesoUo-taZian, rather UaraSSeSed eJFUt earSF in its
HrFUtoJraUhiHeKoSution, Aut then surUassed it. thus, in the Sast Ueriod
of HuneiforZRritinJ, in HoSoUhons Rritten at uruL (in Uresent-daF iraB)
under theseSeuHid LinJs in the Sast feR sHore Fears Aefore the Hhristian
era,oHHasionaS sHriAes HonKerted their naZes into nuZAers.
theenHiUherZentif suHh it AeZaF haKe Aeen onSF for aZuseZent or toshoR
off.
```

Listing 7: Round 6 — after D→u, O→d, G→n

Step 9 — Thematic vocabulary ("transformations", "secrecy", "great", "Egypt's")

With more plaintext structure visible, several domain words lock in additional letters:

- `transforZations` ⇒ `transformations` fixes `Z` → `m`.
- `seHreHF` ⇒ `secrecy` fixes `H` → `c` and `F` → `y`.
- `Jreat` ⇒ `great` fixes `J` → `g`.
- `egyUt's` ⇒ `Egypt's` fixes `U` → `p`.

Known mappings so far: V→e, W→t, Q→h, N→o, X→i, T→a, I→r, C→f, P→s, D→u, O→d, G→n, Z→m, H→c, F→y, U→p, J→g

```
the addition of secrecy to the transformations producedcryptography. true,
    it Ras more of a game than anything eSseit soughtto deSay comprehension
    for onSy the shortest possiASe time, not theSongestand the cryptanaSysis
    Ras, SiLeRise, Eust a puMMSe. egypt's Rasthus a Buasi cryptoSogy in
    contrast to the deadSy serious science of today.yet great things haKe
    smaSS Aeginnings, and these hierogSyphs didincSude, though in an
    imperfect fashion, the tRo eSements of secrecy andtransformation that
    comprise the essentiaS attriAutes of the science. andso cryptoSogy Ras
    Aorn. in its first 3,000 years, it did not groR steadiSy. cryptoSogy
    aroseindependentSy in many pSaces, and in most of them it died the deaths
     ofits ciKiSiMations. in other pSaces, it surKiKed, emAedded in a
    Siterature,and from this the neYt generation couSd cSimA to higher SeKeSs
    .Aut progress Ras sSoR and EerLy. more Ras Sost than retained. much of
    thehistory of cryptoSogy of this time is a patchRorL, a craMy BuiSt
    ofunreSated items, sprouting, fSourishing, Rithering. onSy toRard
    theRestern renaissance does the accreting LnoRSedge Aegin to AuiSd up
    amomentum. the story of cryptoSogy during these years is, in other Rords,
    eYactSy the story of manLind. china, the onSy high ciKiSiMation of
    antiBuity to use ideographicRriting, seems neKer to haKe deKeSoped much
    reaS cryptography perhaps for that reason. in one case LnoRn for miSitary
     purposes, the11th-century compiSation, Ru-ching tsung-yao ("essentiaSs
    from miSitarycSassics"), recommended a true if smaSS code. to a Sist of
    40 pSainteYtitems, ranging from reBuests for AoRs and arroRs to the
    report of aKictory, the correspondents RouSd assign the first 40
    ideograms of apoem. then, Rhen a Sieutenant Rished, for eYampSe, to
    reBuest morearroRs, he Ras to Rrite the corresponding ideogram at a
    specified pSaceon an ordinary dispatch and stamp his seaS on it.in china'
    s great neighAor to the Rest, india, Rhose ciKiSiMationSiLeRise deKeSoped
     earSy and to high estate, seKeraS forms of secretcommunications Rere
    LnoRn and, a pparentSy, practiced. the artha-sastra, a cSassic RorL on
    statecraft attriAuted to LautiSya, in descriAingthe espionage serKice of
    india as practicaSSy riddSing the country Riths pies, recommended that
    the officers of the institutes of  spionage giKetheir spies their
    assignments Ay secret Rriting.perhaps most interesting to cryptoSogists,
    amateur orprofessionaS, is that Katsyayana's famous teYtAooL of erotics,
    the Lamasutra,Sists secret Rriting as one of the 64 arts, or yogas, that
    RomenshouSd LnoR and practice. the fourth great ciKiSiMation of antiBuity
    , themesopo-tamian, rather paraSSeSed egypt earSy in its
    cryptographiceKoSution, Aut then surpassed it. thus, in the Sast period
```

```
of cuneiformRriting, in coSophons Rritten at uruL (in present-day iraB)
under theseSeucid Lings in the Sast feR score years Aefore the christian
era,occasionaS scriAes conKerted their names into numAers.
theenciphermentif such it Aemay haKe Aeen onSy for amusement or toshoR
off.
```

Listing 8: Round 7 — after Z→m, H→c, F→y, U→p, J→g

Step 10 — Easy anchors from common words

Several short, familiar words now appear clearly and let us fix the next substitutions:

- `Ras` ⇒ `was` ⇒ `R` → `w`.
- `smaSS` ⇒ `small` (double-`S` ⇒ double-`l`) ⇒ `S` → `l`.
- `Aeginnings` ⇒ `beginnings` ⇒ `A` → `b`.
- `antiBuity` ⇒ `antiquity` ⇒ `B` → `q`.
- `puMMle` ⇒ `puzzle` (double-`M` ⇒ double-`z`) ⇒ `M` → `z`.
- `ciKilivations` ⇒ `civilizations` ⇒ `K` → `v`.

Known mappings so far (extended):

`V→e, W→t, Q→h, N→o, X→i, T→a, I→r, C→f, P→s, D→u, O→d, G→n, Z→m, H→c, F→y, U→p, J→g, R→w, S→l, A→b, B→q, M→z, K→v`

```
the addition of secrecy to the transformations producedcryptography. true,
    it was more of a game than anything elseit soughtto delay comprehension
    for only the shortest possible time, not thelongestand the cryptanalysis
    was, liLewise, Eust a puzzle. egypt's wasthus a quasi cryptology in
    contrast to the deadly serious science of today.yet great things have
    small beginnings, and these hieroglyphs didinclude, though in an
    imperfect fashion, the two elements of secrecy andtransformation that
    comprise the essential attributes of the science. andso cryptology was
    born. in its first 3,000 years, it did not grow steadily. cryptology
    aroseindependently in many places, and in most of them it died the deaths
     ofits civilizations. in other places, it survived, embedded in a
    literature,and from this the neYt generation could climb to higher levels
    .but progress was slow and EerLy. more was lost than retained. much of
    thehistory of cryptology of this time is a patchworL, a crazy quilt
    ofunrelated items, sprouting, flourishing, withering. only toward
    thewestern renaissance does the accreting Lnowledge begin to build up
```

```
amomentum. the story of cryptology during these years is, in other words,
eYactly the story of manLind. china, the only high civilization of
antiquity to use ideographicwriting, seems never to have developed much
real cryptography perhaps for that reason. in one case Lnown for military
 purposes, the11th-century compilation, wu-ching tsung-yao ("essentials
from militaryclassics"), recommended a true if small code. to a list of
40 plainteYtitems, ranging from requests for bows and arrows to the
report of avictory, the correspondents would assign the first 40
ideograms of apoem. then, when a lieutenant wished, for eYample, to
request morearrows, he was to write the corresponding ideogram at a
specified placeon an ordinary dispatch and stamp his seal on it.in china'
s great neighbor to the west, india, whose civilizationliLewise developed
 early and to high estate, several forms of secretcommunications were
Lnown and, a pparently, practiced. the artha-sastra, a classic worL on
statecraft attributed to Lautilya, in describingthe espionage service of
india as practically riddling the country withs pies, recommended that
the officers of the institutes of  spionage givetheir spies their
assignments by secret writing.perhaps most interesting to cryptologists,
amateur orprofessional, is that vatsyayana's famous teYtbooL of erotics,
the Lamasutra,lists secret writing as one of the 64 arts, or yogas, that
womenshould Lnow and practice. the fourth great civilization of antiquity
, themesopo-tamian, rather paralleled egypt early in its
cryptographicevolution, but then surpassed it. thus, in the last period
of cuneiformwriting, in colophons written at uruL (in present-day iraq)
under theseleucid Lings in the last few score years before the christian
era,occasional scribes converted their names into numbers.
theenciphermentif such it bemay have been only for amusement or toshow
off.
```

Listing 9: Round 8 — after R→w, S→l, A→b, B→q, M→z, K→v

Step 11 — Final clean-up from residual word shapes ("likewise")

A remaining token `liLewise` clearly corresponds to the English word `likewise`, which fixes

$$\texttt{L} \rightarrow \texttt{k}.$$

Using this and the last unresolved pairs from context, we also finalize:

$$\texttt{Y} \rightarrow \texttt{j}, \qquad \texttt{E} \rightarrow \texttt{x}.$$

Known mappings so far (finalized):

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

```
the addition of secrecy to the transformations producedcryptography. true,
    it was more of a game than anything elseit soughtto delay comprehension
    for only the shortest possible time, not thelongestand the cryptanalysis
    was, likewise, xust a puzzle. egypt's wasthus a quasi cryptology in
    contrast to the deadly serious science of today.yet great things have
    small beginnings, and these hieroglyphs didinclude, though in an
    imperfect fashion, the two elements of secrecy andtransformation that
    comprise the essential attributes of the science. andso cryptology was
    born. in its first 3,000 years, it did not grow steadily. cryptology
    aroseindependently in many places, and in most of them it died the deaths
     ofits civilizations. in other places, it survived, embedded in a
    literature,and from this the nejt generation could climb to higher levels
    .but progress was slow and xerky. more was lost than retained. much of
    thehistory of cryptology of this time is a patchwork, a crazy quilt
    ofunrelated items, sprouting, flourishing, withering. only toward
    thewestern renaissance does the accreting knowledge begin to build up
    amomentum. the story of cryptology during these years is, in other words,
    ejactly the story of mankind. china, the only high civilization of
    antiquity to use ideographicwriting, seems never to have developed much
    real cryptography perhaps for that reason. in one case known for military
     purposes, the11th-century compilation, wu-ching tsung-yao ("essentials
    from militaryclassics"), recommended a true if small code. to a list of
    40 plaintejtitems, ranging from requests for bows and arrows to the
    report of avictory, the correspondents would assign the first 40
    ideograms of apoem. then, when a lieutenant wished, for ejample, to
    request morearrows, he was to write the corresponding ideogram at a
    specified placeon an ordinary dispatch and stamp his seal on it.in china'
    s great neighbor to the west, india, whose civilizationlikewise developed
     early and to high estate, several forms of secretcommunications were
    known and, a pparently, practiced. the artha-sastra, a classic work on
    statecraft attributed to kautilya, in describingthe espionage service of
    india as practically riddling the country withs pies, recommended that
    the officers of the institutes of  spionage givetheir spies their
    assignments by secret writing.perhaps most interesting to cryptologists,
    amateur orprofessional, is that vatsyayana's famous tejtbook of erotics,
    the kamasutra,lists secret writing as one of the 64 arts, or yogas, that
    womenshould know and practice. the fourth great civilization of antiquity
    , themesopo-tamian, rather paralleled egypt early in its
    cryptographicevolution, but then surpassed it. thus, in the last period
```

```
of cuneiformwriting, in colophons written at uruk (in present-day iraq)
under theseleucid kings in the last few score years before the christian
era,occasional scribes converted their names into numbers.
theenciphermentif such it bemay have been only for amusement or toshow
off.
```

Listing 10: Round 9 — after L→k, Y→j, E→x

Recovered key alphabet

The final substitution key (cipher alphabet aligned with plaintext) is shown below.

V	W	X	N	P	T	I	G	Q	H	S	O	U	F	Z	D	J	C	R	A	K	L	B	M	Y	E
273	250	201	195	188	185	169	165	111	89	84	79	66	64	61	59	59	56	42	21	20	15	7	7	5	2
11.0	10.1	8.1	7.9	7.6	7.5	6.8	6.7	4.5	3.6	3.4	3.2	2.7	2.6	2.5	2.4	2.4	2.3	1.7	0.8	0.8	0.6	0.3	0.3	0.2	0.1
e	t	i	o	s	a	r	n	h	c	l	d	p	y	m	u	g	f	w	b	v	k	q	z	j	x

Figure 5: Recovered substitution key.

Conclusion

In conclusion, this laboratory work provided a comprehensive understanding of how monoalphabetic substitution ciphers can be analyzed and broken using frequency-based techniques. By computing the statistical distribution of symbols in the ciphertext and comparing it with the known frequency of letters in the English language, we were able to identify consistent patterns that led to the gradual reconstruction of the plaintext. The process also demonstrated the importance of linguistic intuition—recognizing common digraphs, trigraphs, and typical word structures played a crucial role in refining the decryption.

This exercise illustrated both the power and the limitations of classical ciphers. Although monoalphabetic substitution once represented an essential step in the historical development of cryptography, its deterministic nature makes it highly vulnerable to statistical analysis. The experiment emphasized that true security cannot rely solely on secrecy of the algorithm but must also depend on the strength and variability of the key. Modern encryption systems therefore adopt complex mathematical transformations and large key spaces to resist such analytical attacks.

Git repository: `https://github.com/AlexandruRudoi/CS_Labs/tree/main/Lab_2`

Bibliography

1. Course materials, *Cryptography and Security*, UTM – FCIM, 2025.

2. Crypto Corner, Frequency Analysis: Breaking the Code.