

An example of codebreaking using statistical analysis

This is an example of an actual codebreaking. I will describe the process as it happened, and this document was written while the cipher was being broken. A text was taken from the start of a novel, and encrypted using a random monoalphabetic substitution. This gave the following cipher text.

```
STFIT QIFGU VFGTB EIGTV AZDQI ZQIFI MOQIN TQIQA TILDM GMYAE
TQVGU FVIZN JITAZ QVGLI MFZVZ PTQMT WMJWM FGQDM GMXXM FIZTL
JZATT AMFFV SITVL LISIZ VZPQI DMGZA TDQAL LJNVG OAZOI FTINA
TILIP FMWUF AWQVW YIGXI MKVZP MFAAW AZLJV UZATZ AVGJF IXLJX
MVNDM GXFAN EOINU AFTQI IZBEV FIFMT TQIAU UVOIG ATQMT TQIEZ
NIFGT MZNVZ PTQIJ GQAEI NWIIT MTCQI GTIFF MTQIF TQMZM TLVSI
FXAAL FIWMV ZINTA TQMTI HTIZT GAEZN TQIGM WIGIO FITXF VZOVX
LIQAD ISIFT QMTQM NXFAW XTINS TFITQ IFZAT MYGAL ETILJ TANIG
VFIWM JWMFG QGXFI GIZOI MTTQI NAOKT QMTQM NLINQ VWTQE GTAXA
GTXAZ IUAFM UIDQA EFGQV GIZRA JWIZT AUVTZ ADAXI FMTIN TAWMK
IQVWU IILQI OAELN GTVLL DMVTD VTQAE TNVGM XXAVZ TWIZT TQIJD
AELNN VZITA PITQI FMTTQ IDAFG TMZND VTQML LFIGX IOTTA NIMFA
LNWMJ WMFGQ VUZAT ISIZU AFTQM TWMTT IFTAQ VWGIL UTQIF IDMGL
VTTLI UIMFT QMTVZ TQIGI BEILT QIJGQ AELNZ TGIII ZAEPQ AUIMO
QATQI FTQIX FVZOV XLIIQ MSIRE GTWIZ TVAZI NMGAX IFMTV ZPQMN
YIIZD VTQTQ IWAGT ZIDLJ NVGIW YMFKI NAUTQ ITDAW IZDQA LLJVZ
GTVZO TVSIT QIUFE VTAUM GQMFY GIZGI TQMTN ILVPQ TUELM GVTDA
ELNYI TAUVZ NQVWG ILULA AKVZP MUTIF GAWEO QGIXM FMTVA ZVZTA
QVGOA WFMNI GUMOI QVGYE GVZIG GDAEL NYIMT FVULI YEZPL INGQA
ELNQI GVWXL JMFEM ZPIUA FTQVG OAEZT IZMZO ITAXF IGIPT VTGIL
UTATQ IZIMF VZPGT IMWIF MGTQI UVFGT ZATIA UEEFA XIMVH INDVT
QISIF JTQVZ PDMGT QIMXX FIQIZ GVAZM LFIMN JAZST FITQI FGXMF
TTQMT VTDAE LNMTY IGTTQ FAEPQ AETXF ASITQ IZATI AUVEF AXIVZ
BEVTI MGEUU VOVIZ TNIPF II
```

First we do some letter counts. This is easily done; we cut and paste the text into a webpage that counts letters and get the following percentages:

I=13%, T=11%, A,Q=7%,

F,M,V,Z=6%, G=5%, L=4%, N=3%,

D,E,U,W,X=2%, J,O,P,S=1%, with the rest less than 1%.

We also do the same for bigrams. Here we have to be careful as the spaces may mess things up, so we remove them. The total number of the most popular pairs was as follows, where in brackets I give the count for the reversed pair. This will be useful as **er** and **re** are both common pairs, so it can help to find these two letters.

TQ=48 (2), QI=35 (6), IF=20 (17), IZ=22 (7), VZ=20 (3),
with the rest being less than 20.

We also look for doubled letters: TT=11, II=8, LL=5, AA,FF,XX=3, EE, UU=2.

The distribution of letters looks similar to that for English, so we suspect a monoalphabetic substitution (but I knew this already). An obvious guess for the most common letter is I=e. Also, the most common pair suggests TQ=th. We put these in. Note that we use CAPITAL letters for ciphertext and lowercase letters for plaintext. We get

```

StFet heFGU VFGtB EeGtV AZDhe ZheFe MOheN thehA teLDM GMYAE
thVGU FVeZN JetAZ hVGLe MFZVZ PthMt WMJWM FGhDM GMXXM FeZtL
JZAAt AMFFV SetVL LeSeZ VZPhe DMGZA tDhAL LJNVG OAZOe FteNA
teLeP FMWUF AWhVW YeGXe MKVZP MFAAW AZLJV UZAtZ AVGJF eXLJX
MVNDM GXFAN EOeNU AFthe eZBEV FeFMt theAU UVOeG AthMt theEZ
NeFGt MZNVZ PtheJ GhAEL NWeet MtChe GteFF MtheF thMZM tLVSe
FXAAL FeWMV ZeNtA thMte HteZt GAEZN theGM WeGeO FetXF VZOVX
LeHAD eSeFt hMthM NXFAW XteNS tFeth eFZAt MYGAL EteLJ tANeG
VFeWM JWMFG hGXFe GeZOe Mtthe NAOkt hMthM NLeNh VWthE GtAXA
GtXAZ eUAFM UeDhA EFGhV GeZRA JWeZt AUVtZ ADAXe FMteN tAWMK
ehVWU eeLhe OAELN GtVLL DMVtD VthAE tNVGM XXAVZ tWeZt theJD
AELNN VZetA Pethe FMtth eDAFG tMZND VthML LFeGX eOttA NeMFA
LNWMJ WMFGh VUZAt eSeZU AFthM tWMtt eFtAh VWGeL UtheF eDMGL
VttLe UeMft hMtVZ theGe BEeLt heJGh AELNZ tGeee ZAEPH AUeMO
hAthe FtheX FVZOV XLeeh MSeRE GtWeZ tVAZe NMGAX eFmTV ZPhMN
YeeZD Vthth eWAGt ZeDLJ NVGeW YMFKe NAUth etDAW eZDhA LLJVZ
GtVZO tVSet heUFE VtAUM GhMFX GeZGe thMtN eLVPh tUELM GvtDA
ELNYe tAUVZ NhVWG eLULA AKVZP MUteF GAWEO hGeXM FMtVA ZVZtA
hVGOA WFMNe GUMoE hVGYE GVZeG GDAEL NYeMt FVULe YEZPL eNGhA
ELNhe GVWXL JMFfM ZPeUA FthVG OAEZt eZMZO etAXF eGeZt VtGeL
UtAth eZeMF VZPGt eMWeF MGthe UVFGt ZAtEA UEEFA XeMVH eNDVt
heSeF JthVZ PDMGt heMXX FeheZ GVAZM LFeMN JAZSt Fethe FGXMF
tthMt VtDAE LNMtY eGtth FAEPH AeTXF ASeth eZAtE AUeEF AXeVZ
BEVte MGEUU VOVeZ tNePF ee

```

In plaintext the pairs er and re should both be popular; the only example in our list which is popular both ways round is IF, so we try F=r.

```

Stret herGU VrGtB EeGtV AZDhe Zhere MOheN thehA teLDM GMYAE
thVGU rVeZN JetAZ hVGLe MrZVZ PthMt WMJWM rGhDM GMXXM reZtL
JZAAt AMrrV SetVL LeSeZ VZPhe DMGZA tDhAL LJNVG OAZOe rteNA
teLeP rMWUr AWhVW YeGXe MKVZP MrAAW AZLJV UZAtZ AVGJr eXLJX
MVNDM GXrAN EOeNU Arthe eZBEV rerMt theAU UVOeG AthMt theEZ
NerGt MZNVZ PtheJ GhAEL NWeet MtChe Gterr Mther thMZM tLVSe
rXAAL reWMV ZeNtA thMte HteZt GAEZN theGM WeGeO retXr VZOVX
LeHAD eSert hMthM NXrAW XteNS treth erZAt MYGAL EteLJ tANeG
VreWM JWMrG hGXre GeZOe Mtthe NAOkt hMthM NLeNh VWthE GtAXA
GtXAZ eUARm UeDhA ErGhV GeZRA JWeZt AUVtZ ADAXe rMteN tAWMK
ehVWU eeLhe OAELN GtVLL DMVtD VthAE tNVGM XXAVZ tWeZt theJD
AELNN VZetA Pethe rMtth eDarG tMZND VthML LreGX eOttA NeMrA

```

LNWMJ WMrGh VUZAt eSeZU ArthM tWMtt ertAh VWGeL Uther eDMGL
VttLe UeMrt hMtVZ theGe BEeLt heJGh AELNZ tGeee ZAEPH AUeMO
hAthe rtheX rVZOV XLeeh MSeRE GtWeZ tVAZe NMGAX erMtV ZPhMN
YeeZD Vthth eWAGt ZeDLJ NVGeW YMrKe NAUth etDAW eZDhA LLJVZ
GtVZO tVSet heUrE VtAUM GhMrX GeZGe thMtN eLVPh tUELM GvtDA
ELNYe tAUVZ NhVWG eLULA AKVZP MUter GAWE0 hGeXM rMtVA ZVZtA
hVGOA WrMNe GUM0e hVGYE GVZeG GDAEL NYeMt rVULe YEZPL eNGhA
ELNhe GVWXL JMrrM ZPeUA rthVG OAEZt eZMZO etAXr eGeZt VtGeL
UtAth eZeMr VZPGt eMWer MGthe UVrGt ZAtEA UEErA XeMVH eNDVt
heSer JthVZ PDMGt heMXX reheZ GVAZM LreMN JAZSt rethe rGXMr
tthMt VtDAE LNMtY eGtth rAEPH AEtXr ASeth eZate AUEEr AXeVZ
BEVte MGEUU VOVeZ tNePr ee

The beginning looks odd? But we continue anyway. We might try IZ=ed. But Z=6% and d=4% in frequency so not sure [in fact it turned out later that this guess would be wrong!] There are no common vowel pairs in our bigrams, so V may be a vowel?

Now we look through the text for aid. We notice that thM occurs a lot, usually followed by t, and “that” is a common word, so maybe M=a? We can try this:

Stret herGU VrGtB EeGtV AZDhe Zhere aOheN thehA teLDa GaYAE
thVGU rVeZN JetAZ hVGLe arZVZ Pthat WaJWa rGhDa GaXXa reZtL
JZAtt AarrV SetVL LeSeZ VZPhe DaGZA tDhAL LJNVG OAZOe rteNA
teLeP raWUr AWhVW YeGXe aKVZP arAAW AZLJV UZAtZ AVGJr eXLJX
aVNDa GXrAN EOeNU Arthe eZBEV rerat theAU UV0eG Athat theEZ
NerGt aZNVZ PtheJ GhAEL Nweet atChe Gterr ather thaZa tLVSe
rXAAL reWaV ZeNtA thate HteZt GAEZN theGa WeGe0 retXr VZOVX
LeHAD eSert hatha NXrAW XteNS treth erZAt aYGAL EteLJ tANeG
VreWa JWarG hGXre GeZOe atthe NAOkt hatha NLeNh VWthE GtAXA
GtXAZ eUara UeDhA ErGhV GeZRA JWeZt AUVtZ ADAXe rateN tAWaK
ehVWU eeLhe OAELN GtVLL DaVtD VthAE tNVGa XXAVZ tWeZt theJD
AELNN VZetA Pethe ratth eDARg taZND VthaL LreGX eOttA NearA
LNWaJ WarGh VUZAt eSeZU Artha tWatt ertAh VWGeL Uther eDaGL
VttLe Ueart hatVZ theGe BEeLt heJGh AELNZ tGeee ZAEPH AUea0
hAthe rtheX rVZOV XLeeh aSeRE GtWeZ tVAZe NaGAX eratV ZPhaN
YeeZD Vthth eWAGt ZeDLJ NVGeW YarKe NAUth etDAW eZDhA LLJVZ
GtVZO tVSet heUrE VtAUa GharX GeZGe thatN eLVPh tUELa GvtDA
ELNYe tAUVZ NhVWG eLULA AKVZP aUter GAWE0 hGeXa ratVA ZVZtA
hVGOA WraNe GUa0e hVGYE GVZeG GDAEL NYeat rVULe YEZPL eNGhA
ELNhe GVWXL Jarra ZPeUA rthVG OAEZt eZaZO etAXr eGeZt VtGeL
UtAth eZear VZPGt eaWer aGthe UVrGt ZAtEA UEErA XeaVH eNDVt
heSer JthVZ PDaGt heaXX reheZ GVAZa LreaN JAZSt rethe rGXar
tthat VtDAE LNatY eGtth rAEPH AEtXr ASeth eZate AUEEr AXeVZ
BEVte aGEUU VOVeZ tNePr ee

Look at line 6: Nweet atChe Gterr ather thaZa tLVSe. This looks like it might be Nweet at CheGter rather thaZ at.... which would suggest Z=n. This fits with IZ=en

being a popular bigram (albeit not on my list), and also with Z=6% and n=6.7%, so we try try this.

```

Stret herGU VrGtB EeGtV AnDhe nhere aOheN thehA teLDa GaYAE
thVGU rVenN JetAn hVGLe arnVn Pthat WaJWa rGhDa GaXXa rentL
JnAtt AarrV SetVL LeSen VnPhe DaGnA tDhAL LJNVG OAnOe rteNA
teLeP raWUr AWhVW YeGXe aKVnP arAAW AnLJV UnAtn AVGJr eXLJX
aVNDa GXrAN EOeNU Arthe enBEV rerat theAU UVOeG Athat theEn
NerGt anNVn PtheJ GhAEL Nweet atChe Gterr ather thana tLVSe
rXAAL reWaV neNtA thate Htent GAEnN theGa WeGeO retXr VnOVX
LehAD eSert hatha NXrAW XteNS treth ernAt aYGAL EteLJ tANeG
VreWa JWarG hGXre GenOe atthe NAOkt hatha NLeNh VWthE GtAXA
GtXAn eUara UeDhA ErGhV GenRA JWent AUVtn ADAXe rateN tAWaK
ehVWU eeLhe OAELN GtVLL DaVtD VthAE tNVGa XXAVn tWent theJD
AELNN VnetA Pethe ratth eDARg tanND VthaL LreGX eOttA NearA
LNWaJ WarGh VUnAt eSenU Artha tWatt ertAh VWGeL Uther eDaGL
VttLe Ueart hatVn theGe BEeLt heJGh AELNn tGeee nAEPH AUeaO
hAthe rtheX rVnOV XLeeh aSeRE GtWen tVane NaGAX eratV nPhaN
YeenD Vthth eWAGt neDLJ NVGeW YarKe NAUth etDAW enDhA LLJVn
GtVnO tVSet heUrE VtAUa GharX GenGe thatN eLVPh tUELa GVtDA
ELNYe tAUVn NhVWG eLULA AKVnP aUter GAWE0 hGeXa ratVA nVntA
hVGOA WraNe GUaOe hVGYE GVneG GDAEL NYeat rVULe YEnPL eNGhA
ELNhe GVWXL Jarra nPeUA rthVG OAEnt enanO etAXr eGent VtGeL
UtAth enear VnPGt eaWer aGthe UVrGt nAteA UEErA XeaVH eNDVt
heSer JthVn PDaGt heaXX rehen GVAna LreaN JAnSt rethe rGXar
tthat VtDAE LNatY eGtth rAEPH AEtXr ASeth enAte AUEEr AXeVn
BEVte aGEUU VOven tNePr ee

```

Now the most common code-letter left is A, which also appears doubled 3 times. The common plain-letters left are o,i,s, so probably A=o or A=s. But in the top line we have aOheN thehA teLDa and the middle of this seems to make no sense if A=s. So we try A=o. If this is the case then V is the next most common letter and we already think this is a vowel, so try V=i.

```

Stret herGU irGtB EeGti onDhe nhere aOheN theho teLDa GaYoE
thiGU rienN Jeton hiGLe arnin Pthat WaJWa rGhDa GaXXa rentL
Jnott oarri SetiL LeSen inPhe DaGno tDhoL LJNiG OonOe rteNo
teLeP raWUr oWhiW YeGXe aKinP arooW onLJi Unotn oiGJr eXLJX
aiNDa GXroN EOeNU orthe enBEi rerat theoU UiOeG othat theEn
NerGt anNin PtheJ GhoEL Nweet atChe Gterr ather thana tLiSe
rXool reWai neNto thate Htent GoEnN theGa WeGeO retXr inOiX
LehOd eSert hatha NXrow XteNS treth ernot aYGoL EteLJ toNeG
ireWa JWarG hGXre GenOe atthe NoOKt hatha NLeNh iWthE GtoXo
GtXon eUora UeDho ErGhi GenRo JWent oUitn oDoXe rateN toWaK
ehiWU eeLhe OoELN GtiLL DaitD ithoE tNiGa XXoin tWent theJD

```

```

oELNN ineto Pethe ratth eDorG tanND ithaL LreGX eOtto Nearo
LNWaJ WarGh iUnot eSenU ortha tWatt ertoh iWGeL Uther eDaGL
ittLe Ueart hatin theGe BEeLt heJGh oELNn tGeee noEPH oUeaO
hothe rtheX rinOi XLeeh aSeRE GtWen tione NaGoX erati nPhaN
YeenD ithth eWoGt neDLJ NiGeW YarKe NoUth etDoW enDho LLJin
GtinO tiSet heUrE itoUa GharX GenGe thatN eLiPh tUELa GitDo
ELNYe toUin NhiWG eLULo oKinP aUter GoWEO hGeXa ratio ninto
hiGoo WraNe GUaOe hiGYE GineG GDoEL NYeat riULe YEnPL eNGho
ELNhe GiWXL Jarra nPeUo rthiG OoEnt enanO etoXr eGent itGeL
Utoth enear inPGt eaWer aGthe UirGt noteo UEEro XeaiH eNDit
heSer Jthin PDaGt heaXX rehen Giona LreaN JonSt rethe rGXar
tthat itDoE LNatY eGtth roEPH oEtXr oSeth enote oUEEr oXein
BEite aGEUU iOien tNePr ee

```

Look at EeGti onDhe nhere aOheN in the first line. It looks like here is one word, and that tion forms part of another word. No obvious word continues ...tion.hen so we probably have a word Dhen. We cannot have D=t, so the next guess would be D=w. Looking through we see that this would give who four times in the text, and does not seem obviously wrong anywhere else. So we try this.

```

Stret herGU irGtB EeGti onwhe nhere aOheN theho teLwa GaYoE
thiGU rienN Jeton hiGLe arnin Pthat WaJWa rGhwa GaXXa rentL
Jnott oarri SetiL LeSen inPhe waGno twhoL LJNiG OonOe rteNo
teLeP raWUr oWhiW YeGXe aKinP arooW onLJi Unotn oiGJr eXLJX
aiNwa GXroN EOeNU orthe enBEi rerat theoU UiOeG othat theEn
NerGt anNin PtheJ GhoEL Nweet atChe Gterr ather thana tLiSe
rXool reWai neNto thate Htent GoEnN theGa WeGeO retXr inOiX
Lehow eSert hatha NXrow XteNS treth ernet aYGoL EteLJ toNeG
ireWa JWarG hGXre GenOe atthe NoOKt hatha NLeNh iWthE GtoXo
GtXon eUora Uewho ErGhi GenRo JWent oUitn owoXe rateN toWaK
ehiWU eeLhe OoELN GtiLL waitw ithoE tNiGa XXoin tWent theJw
oELNN ineto Pethe ratth eworG tanNw ithaL LreGX eOtto Nearo
LNWaJ WarGh iUnot eSenU ortha tWatt ertoh iWGeL Uther ewaGL
ittLe Ueart hatin theGe BEeLt heJGh oELNn tGeee noEPH oUeaO
hothe rtheX rinOi XLeeh aSeRE GtWen tione NaGoX erati nPhaN
Yeenw ithth eWoGt newLJ NiGeW YarKe NoUth etwoW enwho LLJin
GtinO tiSet heUrE itoUa GharX GenGe thatN eLiPh tUELa Gitwo
ELNYe toUin NhiWG eLULo oKinP aUter GoWEO hGeXa ratio ninto
hiGoo WraNe GUaOe hiGYE GineG GwoEL NYeat riULe YEnPL eNGho
ELNhe GiWXL Jarra nPeUo rthiG OoEnt enanO etoXr eGent itGeL
Utoth enear inPGt eaWer aGthe UirGt noteo UEEro XeaiH eNwit
heSer Jthin PwaGt heaXX rehen Giona LreaN JonSt rethe rGXar
tthat itwoE LNatY eGtth roEPH oEtXr oSeth enote oUEEr oXein
BEite aGEUU iOien tNePr ee

```

In line 1 we have aOheN theho telwa GaYoE. the middle looks like the hotel was, and LL=5 fits with ll being common. Looking at the remaining common letters we have G=5% and all others 3% or less, and we still do not have s=6.1% Trying L=1 and G=s gives

```

Stret hersU irstB Eesti onwhe nhere aOheN theho telwa saYoE
thisU rienN Jeton hisle arnin Pthat WaJWa rshwa saXXa rentl
Jnott oarri Setil leSen inPhe wasno twhol lJNis OonOe rteNo
teleP raWUr oWhiW YesXe aKinP arooW onlJi Unotn oisJr eXlJX
aiNwa sXroN EOeNU orthe enBEi rerat theoU UiOes othat theEn
Nerst anNin PtheJ shoEl Nweet atChe sterr ather thana tliSe
rXool reWai neNto thate Htent soEnN thesa WeseO retXr inOiX
lehow eSert hatha NXroW XteNS treth ernet aYsol EtelJ toNes
ireWa JWars hsXre senOe atthe NoOKt hatha NleNh iWthE stoXo
stXon eUora Uewho Ershi senRo JWent oUitn owoXe rateN toWaK
ehiWU eelhe OoElN still waitw ithoE tNisa XXoin tWent theJw
oElNN ineto Pethe ratth ewors tanNw ithal lresX eOtto Nearo
lNwaJ Warsh iUnot eSenU ortha tWatt ertoh iWsel Uther ewasl
ittle Ueart hatin these BEelt heJsh oElNn tseee noEPH oUeaO
hothe rtheX rinOi Xleeh aSeRE stWen tione NasoX erati nPhaN
Yeenw ithth eWost newlJ NiseW YarKe NoUth etwoW enwho llJin
stinO tiSet heUrE itoUa sharX sense thatN eliPh tUEla sitwo
ElNye toUin NhiWs elUlo oKinP aUter soWEO hseXa ratio ninto
hisOo WraNe sUaOe hisYE sines swoEl NYeat riUle YEnPl eNsho
ElNhe siWXl Jarra nPeUo rthis OoEnt enanO etoXr esent itsel
Utoth enear inPst eaWer asthe Uirst noteo UEEro XeaiH eNwit
heSer Jthin Pwast heaXX rehen siona lreaN JonSt rethe rsXar
tthat itwoE lNatY estth roEPH oEtXr oSeth enote oUEEr oXein
BEite aseUU iOien tNePr ee

```

Lots of this looks like real words now, with a few odd bits. The beginning is odd, but ignoring the first S for now we have Stret hersU irstB Eesti onwhe nhere aOheN theho telwa s which looks like Strethers first question when he reached the hotel was... So let us try U=f, B=q, E=u, O=c, N=d. We now have so many letters that we can try to write the words out.

Strether's first question when he reached the hotel was aYouthis friend Jetonhis learninP that WaJWarshwasaxXarentlJ not to arriSe till eSeninP he was not whollJ disconcerted ...

This looks very good. aYouthis will be about his and learninP will be learning, arriSe will be arrive and this all fits with eSeninP for evening. Also whollJ will be wholly. This still leaves a few unclear bits, but we will now try Y=b, P=g, S=v, J=y. Now the text starts

vtrether's first question when he reached the hotel was about his friend yet on his learning that WayWarsh was axXarently not to arrive till evening he

was not wholly disconcerted otelegramwfromwhiwbesXeaKingarooW only if not noisy reXlyXaid was Xroduced for the enquirer at the office so that the understanding they should Weet at Chester rather than at liverXool reWained to that eHtent sound the saWe secret XrinciXle however that had XroWXtedvtrether not absolutely...

The beginning is odd, but the rest looks good. It must be that aXXarently is apparently and Xroduced is produced and reWained is remained and eHtent is extent. Trying X=p, W=m, H=x we get

vtrethers first question when he reached the hotel was about his friend yet on his learning that maymarsh was apparently not to arrive till evening he was not wholly disconcerted o telegram from him bespeaking a room only if not noisy reply paid was produced for the enquirer at the office so that the understanding they should meet at Chester rather than at liverpool remained to that extent sound the same secret principle however that had prompted vtrether not absolutely to desire maymarshs presence at the dock that had led him thus to postpone for a few hours his enRoyment of it now operated to make him feel he could still wait without disappointment they would dine together at the worst and with all respect to dear old maymarsh if not even for that matter to himself there was little fear that in the sequel they shouldnt see enough of each other the principle e have Rust mentioned as operating had been with the most newly disembarked of the two men wholly instinctive the fruit of a sharp sense that delightful as it would be to find himself looking after so much separation in to his comrades face his business would be a trifle bungled should he simply arrange for this countenance to present itself to the nearing steamer as the first note of uurope aixed with everything was the apprehension already on vtrethers part that it would at best throughout prove the note of uurope in quite a sufficient degree

It is now easy to see what the passage should be, except that there are some obvious problems. vtrether occurs three times and maymarsh twice, and there is also o telegram and e have and uurope twice and aixed.

The first two seem to be names, and uurope could be Europe — and the mystery is resolved. It seems that the encryption did not deal with capital letters properly! Indeed, looking at the cipher text, it appears that they were left unaltered, so we should have Strether and Waymarsh and A telegram and I have and Europe and Mixed. The last few letter substitutions are obvious and we find that we have the opening paragraph of “The Ambassadors” by Henry James. The cipher was given by

abcdefghijklmnopqrstuvwxyz
 MYONIUPQVRKLWZAXBFGTESDHJC

Note that this process was not automatic. We had to work by trial and error, and some of our guesses would have been wrong. The final letter count was

e=13%, t=11%, o,h=7%,
r,a,i,n=6%, s=5%, l=4%, d=3%,
w,u,f,m,p=2%, y,c,g,v=1%, with the rest less than 1%.

which is not quite the predicted order (h is higher than expected for example). But the statistical data plus some common sense was good enough to decode this — even though there were some errors in the original text! These errors were not planted deliberately, but turned out to be a side-effect of the webpage used to encrypt the message. However, this is something that does happen in real life; for example the text might contain spelling errors or miscoding due to human error. Thus when we are guessing possible solutions, we have to allow a certain amount for the possibility of error in the original text.