Capítulo 3

Criptografia não digital

3.1 Esteganografia

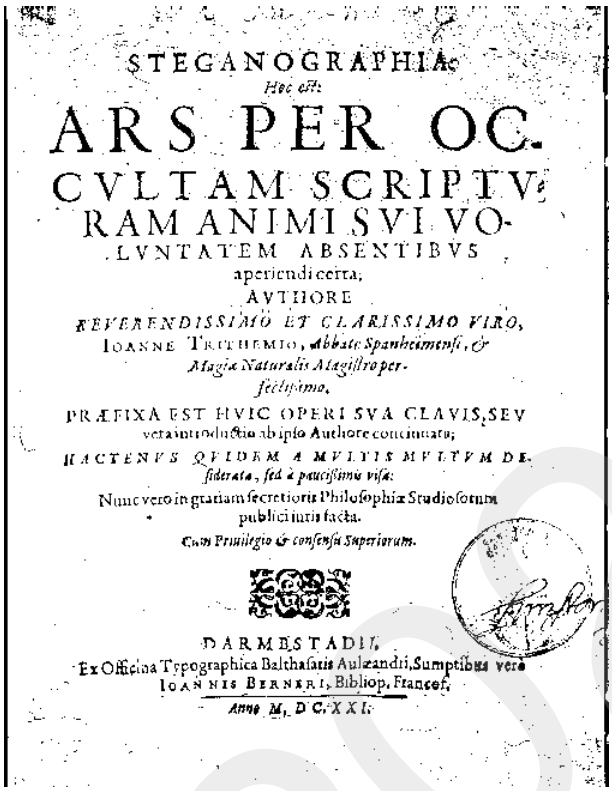
Para poder enviar uma mensagem de forma a que o seu conteúdo não seja legível para terceiros, a criptografia oculta o método (a cifra) com que foi codificada a mensagem, ou oculta simplesmente a chave que foi utilizada, sendo a cifra pública. Apesar de não dever ser fácil ler o conteúdo da mensagem, é evidente que de uma mensagem se trata, e que esta está cifrada para ocultar o seu conteúdo. Mas esta não é a única solução. A esteganografia tenta ocultar o próprio facto de que se está a transmitir informação. Tenta-se que a própria transmissão da informação não seja evidente, ou então dissimula-se a verdadeira mensagem no suporte de uma outra mensagem sem importância. O nome de esteganografia que provém está ligado a **Johannes Trithemius** que publica um livro que aparentemente versa a Magia Negra, mas que de facto consiste num tratado sobre criptografia. Os exemplos ao longo da História são muitos:

- **Escrita no crânio de um escravo.** Segundo Heródoto, quando Histiaeus quis comunicar ao seu genro Aristagoras de Mileto que estava na altura de se revoltar contra a ocupação persa, tatuou essa mensagem no crânio de um escravo de confiança, a quem previamente havia rapado o cabelo, e enviou-o através das linha inimigas depois de esperar que o cabelo tivesse voltado a crescer. [Kah67, Sal90].

- **Mensagens engolidas** Os chineses faziam transportar as mensagens mais sensíveis por mensageiros, que as engoliam depois de as colocar em pequenas bolas de cera [Gai39].

- **Escrita em ovos cozidos.** Giovanni Porta (1535–1615) descreve a receita de uma solução de alumen e vinagre que escrita na casca de um ovo cozido, penetra através da casca e pode ser lida somente depois desta removida [Sin99].

Draft 2021.1



STEGANOGRAPHIAE
Hoc eſt:
ARS PER OC-
CVLTAM SCRIPTV-
RAM ANIMI SVI VO-
LVNTATEM ABSENTIBVS
aperiendi certa;
AVTHORE
REVERENDISSIMO ET CLARISSIMO VIRO,
IOANNE TRITHEMIO, Abbate Spanheimenſi, &
Magiæ Naturalis Magiſtro per-
fectiſſimo.
PRÆFIXA EST HVIC OPERI SVA CLAVIS, SEV
vera introductio ab ipſo Authore concinnata;
HACTENVS QVIDEM A MVLTIS MVLTVM DE-
ſiderata, ſed à pauciſſimis viſa:
Nunc vero in gratiam ſecretioris Philoſophiæ Studioſorum
publici iuris facta.
Cum Priuilegio & conſenſu Superiorum.
DARMESTADII,
Ex Officina Typographica Balthaſaris Aulæandri, Sumptibus vero
IOANNIS BERNERI, Bibliop. Francof.
Anno M. DC. XXI.

Figura 3.1: "Steganographia"de Johannes Trithemius

- **Micro-perfuração** Para evitar pagar as altas taxas postais para a correspondência privada e passar a pagar a taxa de envio de materiais impressos (muito mais baixa), generalizou-se em Inglaterra, durante algum tempo no século passado, a utilização de um estranho meio de comunicação: Num qualquer jornal, produziam-se com o auxílio de uma agulha, pequenos furos sobre alguma letras do texto impresso, por forma a que essas letras escolhidas formassem a mensagem que se queria enviar! [Gai39]

- **A "escrita simpática".** Escrita de uma mensagem com sumo de um citrino (normalmente limão) nas entrelinhas de uma outra mensagem. Quando o papel é submetido ao calor a mensagem até aí invisível aparece em tons de castanho. Martin Gardner [Gar72] descreve para além desta algumas outras receitas de tintas "invisíveis".

- **Microponto.** Um microfilme do formato de um ponto final de uma máquina de escrever pode transportar informação assim dissimulada numa carta absolutamente inócua. Este método foi usado desde 1870. Durante a segunda guerra mundial foi usado com regularidade pelas embaixadas alemãs na América Latina [Kah67].

- **Grelhas de leitura.** O cardeal Richelieu (1585–1642) usava na comunicação com os seus agentes, uma grelha de papel que sobreposta a uma mensagem escrita, ocultava

Draft 2021.1

Figura 3.2: Giovanni (Giambattista) della Porta

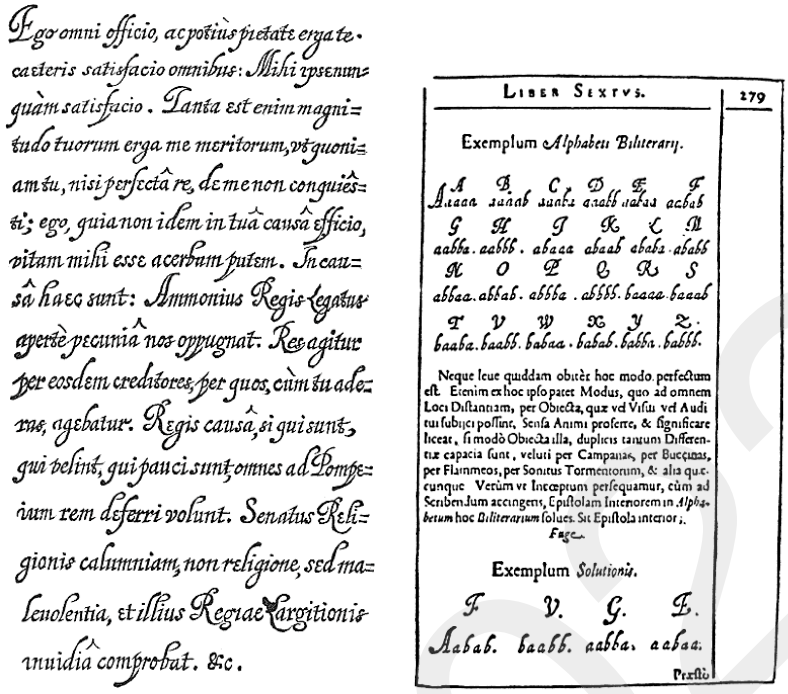
todas as palavras excepto aquelas que constituíam verdadeiramente a informação a transmitir [Gai39].

- **Variação de tipos impressos.** Francis Bacon (1561–1626) colocava uma mensagem oculta num texto impresso, fazendo variar entre duas fontes muito semelhantes as diversas letras de um texto [Gai39, Kah67].

Para todos este métodos não há muito a dizer de como os descobrir, a criptanálise não pode dar grandes ajudas! Mas a utilização da esteganografia assenta sempre na convicção que terceiros não suspeitam que informação, para além da evidente, está a ser enviada. No momento em que uma forte suspeita se instala estes métodos revelam-se quase sempre desastrosos. O principal problema, contudo, é a possibilidade da descoberta fortuita da informação secreta, sem que nenhum esforço tenha sido desenvolvido nesse sentido: o ovo parte-se por acidente, a mensagem com "escrita simpática" é aquecida por uma lâmpada, o mensageiro tem um problema de intestinos, etc ...

A esteganografia não é, no entanto, uma coisa do passado. As imagens digitalizadas que "circulam" na Internet constituem um campo fácil de utilização de processo esteganográficos. Podemos então adaptar o método usado por Francis Bacon (pag.59) às novas tecnologias e usar dois tons muito próximos de branco (por exemplo) em vez de duas fontes de caracteres. A alteração das imagens é imperceptível, e como se trata de um meio digital, tanto o processo de codificação como de descodificação, automatizados não necessitam um esforço excessivo.

Draft 2021.1



Ego omni officio, ac potiùs pietate erga te, cæteris satisfacio omnibus: Mihi ipsenunquàm satisfacio. Tanta est enim magnitudo tuorum erga me meritorum, vt quoniam tu, nisi perfectâ re, de me non conquiêsti; ego, quia non idem in tuâ causâ efficio, vitam mihi esse acerbum putem. In causâ hæc sunt: Ammonius Regis Legatus apertè pecuniâ nos oppugnat. Res agitur per eosdem creditores, per quos, cùm tu aderas, agebatur. Regis causâ, si qui sunt, qui velint, qui pauci sunt, omnes ad Pompeium rem deferri volunt. Senatus Religionis calumniam, non religione, sed malevolentia, et illius Regiæ Largitionis invidiâ comprobat. &c.

LIBER SEXTVS. 279

Exemplum *Alphabeti Biliterarij.*

A	B	C	D	E	F
aaaaa	aaaab	aaaba	aaabb	aabaa	aabab
G	H	I	K	L	M
aabba	aabbb	abaaa	abaab	ababa	ababb
N	O	P	Q	R	S
abbaa	abbab	abbba	abbbb	baaaa	baaab
T	V	W	X	Y	Z
baaba	baabb	babaa	babab	babba	babbb

Neque leue quiddam obiter hoc modo perfectum est. Etenim ex hoc ipso patet Modus, quo ad omnem Loci Distantiam, per Obiecta, quæ vel Visui vel Auditui subijci possint, Sensa Animi proferre, & significare liceat, si modò Obiecta illa, duplicis tantum Differentiæ capacia sunt, veluti per Campanas, per Buccinas, per Flammeos, per Sonitus Tormentorum, & alia quæcunque. Verùm vt Inceptum persequamur, cùm ad Scribendum accingens, Epistolam Intenorem in *Alphabetum hoc Biliterarium* solues. Sit Epistola interior;.

Fuge.

Exemplum *Solutionis.*

F	V.	G.	E.
Aabab.	baabb.	aabba.	aabaa.

Præstò

Figura 3.3: Mensagem cifrada por Bacon e explicação do método usado.

3.2 Cifras monoalfabéticas

O mais simples dos sistemas criptográficos consiste na substituição de cada uma das letras que formam uma mensagem pelas letras correspondentes de um outro alfabeto, a estas cifras chamamos **cifras monoalfabéticas**. O alfabeto correspondente ao texto cifrado não tem que usar os símbolos do alfabeto original como é exemplo a cifra usada por Conan Doyle em *"The Adventure of a Dancing Man"* (uma das aventuras do Sherlock Holmes)

Figura 3.4: Criptograma da aventura de Sherlock Holmes

Cifras deste género não são só produto literário. Carlos Magno usava também uma cifra deste género:

Draft 2021.1

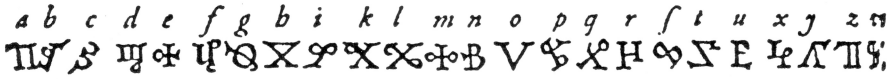


Figura 3.5: Cifra de Carlos Magno.

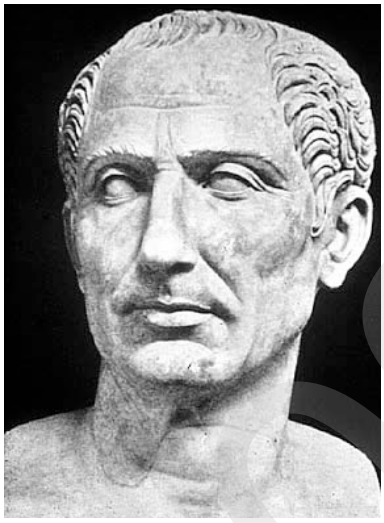


Figura 3.6: Júlio César

3.2.1 A cifra de César

As cifras mais usuais, e aquelas que vamos analisar com mais detalhe, usam no entanto os mesmos caracteres para o texto cifrado e para o original. A mais simples cifra monoalfabética deste tipo é chamada **cifra de César**, que apesar da sua inocente simplicidade, foi (como o nome indica) usada por Júlio César (100–44BC). Como qualquer cifra monoalfabética, fica caracterizada pela correspondência entre o alfabeto original e cifrado. Como convenção vamos usar caracteres minúsculos para representar o texto original e caracteres maiúsculos para representar texto cifrado.

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
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

O alfabeto cifrado é o resultado da aplicação de um deslocamento de três letras para a esquerda ao alfabeto original.

Para se cifrar um texto (por exemplo "`veni vidi vinci`"), basta substituir cada um dos caracteres do alfabeto de cima pelo correspondente caracter do alfabeto de baixo. O resultado é:

```
YHQL YLGL YLFL.
```


Draft 2022.1

Para decifrar basta executar o processo de forma inversa, i.e., substituir cada letra do alfabeto de baixo pela correspondente letra do alfabeto de cima.

Nada nos obriga a usar exactamente esta cifra (com um *shift* de três letras), podemos obter uma cifra diferente escolhendo um qualquer número de deslocamentos (*shifts*) que constituem a chave da cifra. Portanto, com uma cifra deste género temos um universo de **25 chaves**! O que não é propriamente um dissuasor de uma "ataque de força bruta"!

Exercício 3.1 *Qual a mensagem contida neste criptograma que usa a cifra de César exactamente com uma chave obtida por um deslocamento de 3 posições:*

FXLGDGR FRP RV LGRV GH PDUFR.

Não faz sequer sentido falar em criptanálise deste tipo de cifras. Basta tentar decifrar o criptograma com as 25 chaves possíveis[1], para imediatamente reconhecermos a chave usada.

Exercício 3.2 *Quais as mensagens contidas nos seguintes criptogramas que usam a cifra de César com chaves desconhecidas:*

i) `M BC BIUJMU JZCBCA UMC NQTPW`

ii) `SGHSGFCAOBCGGOCZCIQCG`

3.2.2 Cifra monoalfabética genérica

Se se optar por ter uma chave que não seja necessariamente obtida por deslocamento do alfabeto original, então temos um universo de chaves um pouco maior:

$$26! - 1 = 403291461126605635583999999 \approx 4 \times 10^{27}$$

o que afasta definitivamente a hipótese de um ataque de "força-bruta"! Claro que a chave passa a ser mais difícil de decorar, mas podemos tentar obviar esse problema usando a seguinte técnica:

1. Escolhemos uma palavra ou frase que vai servir de chave (senha), por exemplo `Gomes Teixeira`.
2. Começamos por retirar as ocorrências repetidas de cada letra, a partir da segunda: `gomestixra`.
3. Adicionamos à palavra assim obtida as restantes letras do alfabeto, pela respectiva ordem. Obtemos assim a cifra:

[1] E para este ataque basta decifrar um pedaço de uma dezena de caracteres para termos imediatamente a percepção se acertamos na chave ou não.

Draft 2021.1

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
G	O	M	E	S	T	I	X	R	A	B	C	D	F	H	J	K	L	N	P	Q	U	V	W	Y	Z

Apesar deste método não ser perfeito[2] vai permitir que não tenhamos que manter escrita a chave de um canal de comunicações, diminuindo o risco da descoberta da chave por meios não criptográficos[3].

Exercício 3.3 *Usando o método anteriormente descrito para completar as chaves, cifra as seguintes mensagens com as respectivas chaves:*

i) **Chave:** *Pedro Nunes*

Mensagem: *"Que sombras nocturnas são lançadas pela Lua na sua carruagem célere?"*[4]

ii) **Chave:** *Anastácio da Cunha*

Mensagem: *"Copado, alto, gentil PinheiroManso"*[5]

3.2.3 Criptanálise de uma cifra monoalfabética

Consideremos o seguinte criptograma:

```
IWLPY SITSB SDSID QHQJY RLWKS TSADR IWNMR AQIYS TDSNU SMRWA
SMUSC SDQZP JMRHP JDSJS RIQIS LQDQR JDQHR RWDQW ANSDR DSQIY
TSDSD QWEWS YTRCP JRYQI XTRYQ IMRIQ KRPQJ XSTTS KSTIQ IPNQJ
MPRIS QAQIQ TSHQN JSHSN QYSDR RWYTR NSDRN RXRWA SLSYQ XSHST
TQWSQ IYTSD SRIUR AQJIB SQJMU STMSD RICQN RIMUW HPIMR IEWQM
SPSAD QIDQR SNHRT QMQTC TRMWT STSAS LTPXR BWJYR SRIYT RJMRI
QILQN YRIDR ICPJU QPTRI IRDRP ITSCS ZPYRI IQDQP VSTSA KPMST
SLTPY STCQD TSTPJ DRDRI URAQJ IEWQK WXPSA SMUWH S
```

Comecemos por notar que foram eliminados todos os espaços, capitalização de palavras e sinais de pontuação. A razão é simples... se tal não tivesse sido feito muita informação sobre o tamanho das palavras, as letras que ocorrem no início ou do fim das mesmas, etc.., seria uma informação valiosa para simplificar o trabalho do criptanalista. A razão porque os criptogramas se apresentam tradicionalmente com os caracteres agrupados em grupos de cinco tem que ver somente com limitações históricas das transmissões telegráficas.

Procedendo a uma contagem dos caracteres obtemos o seguinte resultado:

[2] Para além de diminuir drasticamente o espaço combinatório das possíveis cifras, para grande parte das palavras (chaves) deixa inalteradas uns quantos caracteres no fim do alfabeto. A solução pode ser melhorada se em vez de colocarmos as restantes letras pela sua ordem natural o fizermos por ordem inversa.

[3] Como por exemplo, bisbilhotando nos cadernos de apontamentos ou na carteira de um dos interlocutores.

[4] Primeira estrofe do poema de António Pinheiro em louvor do "Livro sobre os Crepúsculos" de Pedro Nunes.

[5] Primeira estrofe de um poema de Anastácio da Cunha.

Draft 2021.1

A	B	C	D	E	F	G	H	I	J	K	L	M
12	3	7	25	3	0	0	9	35	16	5	7	16
3.0	0.7	1.7	6.3	0.7	0.0	0.0	2.3	8.9	4.0	1.2	1.7	4.0
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
11	0	19	41	44	57	30	8	1	18	6	16	2
2.8	0.0	4.8	10.4	11.2	14.5	7.6	2.0	0.2	4.6	1.5	4.0	0.5

Tabela 3.1: Ocorrências dos diversos caracteres no criptograma anterior, e respectivas percentagens.

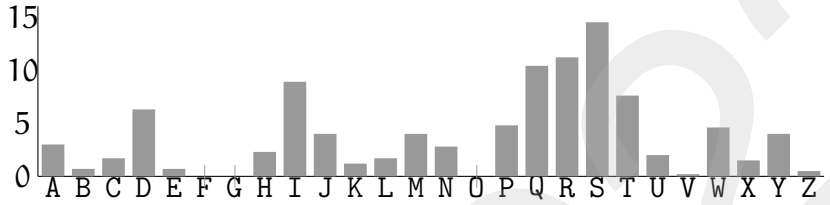


Figura 3.7: Histograma das frequências de ocorrência do criptograma

Como o que ocorre numa cifra monoalfabética é a substituição de cada carácter por um outro de um alfabeto, a frequência relativa das ocorrências vai ser preservada pelo processo de cifra. Portanto se soubermos qual é a língua em que está expressa a mensagem original, podemos comparar as frequências relativas de ocorrência de cada carácter do criptograma com as frequências relativas de ocorrência na língua de origem.[6]

a	b	c	d	e	f	g	h	i	j	k	l	m
13.8	0.9	4.5	5.6	12.0	1.0	1.2	0.6	7.0	0.3	0.0	2.8	4.1
n	o	p	q	r	s	t	u	v	w	x	y	z
5.3	10.8	2.9	0.8	6.9	7.8	4.9	3.8	1.3	0.0	0.2	0.0	0.3

Tabela 3.2: Percentagens de frequência dos caracteres no Português actual.

Os caracteres mais frequentes no português, são então o a, e e o, com 13.8%, 12.0% e 10.8%, respectivamente. No texto do criptograma os caracteres com percentagens comparáveis são o S (14.5%), o R (11.2%) e o Q (10.4%). Portanto temos muito provavelmente que

[6] Para conhecer valores das frequências típicas noutras línguas europeias, assim como alguns outros valores estatísticos sobre a língua inglesa, basta ver o apêndice do livro de Gaines [Gai39]. Os resultados enfermam, para fins práticos, do facto de se referirem à realidade linguística da primeira metade do sec.XX. Estes valores assim como outros actuais, em especial sobre a língua portuguesa podem ser consultados no Apêndice D

Draft 2021.1

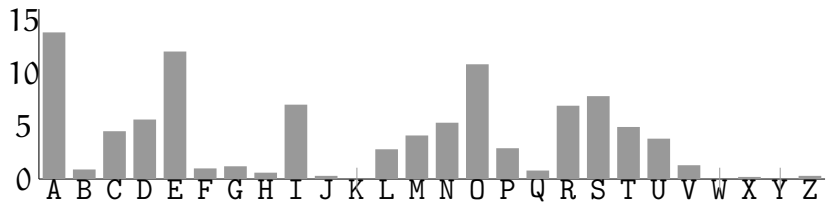


Figura 3.8: Histograma das frequências de ocorrência dos caracteres nos textos estudados.

$$\{\mathtt{E}(\mathtt{a}), \mathtt{E}(\mathtt{e}), \mathtt{E}(\mathtt{o})\} = \{\mathtt{S}, \mathtt{R}, \mathtt{Q}\}$$

Infelizmente no português, as três letras mais frequentes são todas vogais[7] pelo que provavelmente vamos ter que tentar todas as possibilidades (que são só 6).

Podemos apostar que neste criptograma a ordem da frequência destes caracteres se mantinha (o que não é nada garantido) e portanto supor que

$$\mathtt{E}(\mathtt{a}) = \mathtt{S} \quad \mathtt{E}(\mathtt{e}) = \mathtt{R} \quad \mathtt{E}(\mathtt{o}) = \mathtt{Q}$$

As letras que têm uma ocorrência média superior a 4% no português, são

$$\mathtt{c}, \mathtt{d}, \mathtt{i}, \mathtt{m}, \mathtt{n}, \mathtt{r}, \mathtt{s}, \mathtt{t}$$

As 8 letras mais frequentes no criptograma são:

$$\mathtt{D}, \mathtt{I}, \mathtt{J}, \mathtt{M}, \mathtt{P}, \mathtt{T}, \mathtt{W}, \mathtt{Y}$$

Podemos-nos sentir tentados a atribuir

$$\mathtt{E}(\mathtt{i}) = \mathtt{I}$$

dado ser a letra que resta com maior frequência. Mas se observarmos, imediatamente ao lado de I aparece predominantemente um pequeno grupo de letras:

Este tipo de distribuição de caracteres à esquerda e direita é mais compatível com uma consoante (que tem predominantemente na sua vizinhança vogais) do que com uma vogal (que tem predominantemente na sua vizinhança consoantes). Assim tomando a consoante mais frequente, provavelmente temos

$$\mathtt{E}(\mathtt{s}) = \mathtt{I}.$$

Existem duas ocorrências de II (que também podendo ser produto da concatenação das palavras) que é compatível com a existência de ss na língua portuguesa.

Draft 2021.1

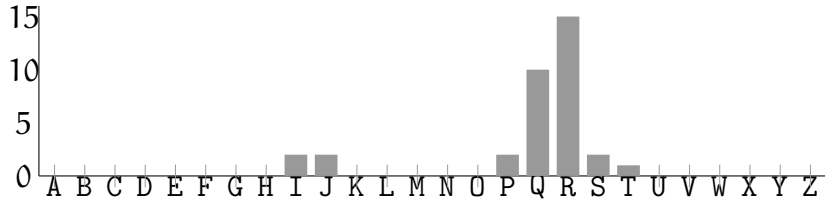


Figura 3.9: Histograma de frequência de caracteres à esquerda de `I`

Figura 3.10: Histograma de frequência de caracteres à direita de `I`

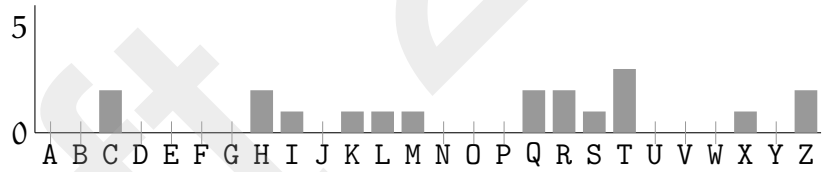


Figura 3.11: Histograma de frequência de caracteres à esquerda de `P`

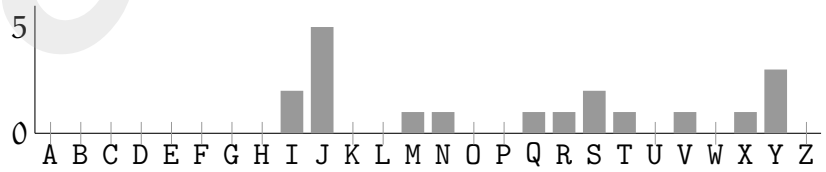


Figura 3.12: Histograma de frequência de caracteres à direita de `P`

Draft 2021.1

Se fizermos o mesmo tipo de análise para as ocorrências na vizinhança de P obtemos:

Esta distribuição muito mais homogénea, especialmente quanto à vizinhança à esquerda, indica que provavelmente corresponde a uma vogal. Então provavelmente teremos

$$E(\mathtt{i}) = \mathtt{P}.$$

O facto de J ocorrer relativamente frequente à direita de P (que por hipótese representa i) leva a suspeitar que se trate da substituição de n. Vejamos que caracteres ocorrem à esquerda de J.

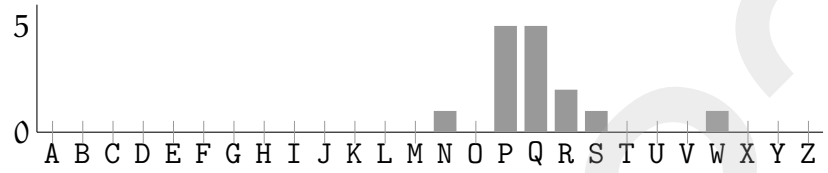


Figura 3.13: Histograma de frequência de caracteres à esquerda de J

O que é compatível com uma consoante (P,Q, R e S correspondem a vogais) e com as ocorrências de n. Para além disso

$$P_{\mathtt{n}} = 5.3\% \text{ e } P_{\mathtt{J}} = 4.0\%.$$

Do conjunto de letras mais frequentes que ainda restam T (7.6%) e r (6.9%), são as de frequência mais elevada e portanto mais prováveis de estarem relacionadas, e reforçando esta suspeita existe o facto de ocorrerem dois pares TT. Suponhamos portanto

$$E(\mathtt{r}) = \mathtt{T}.$$

Procedendo da mesma forma como fizemos para P podemos concluir que, muito provavelmente, W representa uma vogal, e a que resta é u.

$$E(\mathtt{W}) = \mathtt{u}.$$

Falta-nos associar D (6.3%), M (4.0%) e Y (4.0%) a c (4.5%), d (5.6%), m (4.1%) e t (4.9%). Parece que

$$E(\mathtt{d}) = \mathtt{D}$$

Observemos que existem 7 ocorrências de TR e somente 1 de TQ. Isto corresponde, segundo as nossas suposições a muito mais ocorrências de re do que de ro o que não é natural na língua portuguesa. Provavelmente escolhemos mal os valores associados a Q e R. Refaçamos essas escolhas tomando agora

$$E(\mathtt{e}) = \mathtt{Q} \quad E(\mathtt{o}) = \mathtt{R}.$$

Recapitulando temos até agora a seguinte chave "descoberta":

[7] Contráriamente ao que se passa no inglês onde as letras mais frequentes são e, t e a.

Draft 2021

```
abcdefghijklmnopqrstuvwxyz
S  DQ   P    JR  TI W
```

O que corresponde à seguinte decifração parcial:

```
IWLPY SITSB SDSID QHQJY RLWKS TSADR IWNMR AQIYS TDSNU SMRWA
su i  asra  adasd e en  o u a ra do su  o  es a rda   a ou

SMUSC SDQZP JMRHP JDSJS RIQIS LQDQR JDQHR RWDQW ANSDR DSQIY
a  a  ade i n o i ndana osesa  edeo nde o oudeu   ado daes

TSDSD QWEWS YTRCP JRYQI XTRYQ IMRIQ KRPQJ XSTTS KSTIQ IPNQJ
radad eu ua  ro i no es  ro e s ose  oien  arra  arse si en

MPRIS QAQIQ TSHQN JSHSN QYSDR RWYTR NSDRN RXRWA SLSYQ XSHST
 iosa e ese ra e  na a  e ado ou ro  ado  o ou  a a e  a ar

TQWSQ IYTSD SRIUR AQJIB SQJMU STMSD RICQN RIMUW HPIMR IEWQM
reuae s rad aos o  ens  aen   ar ad os e  os  u  is o s ue

SPSAD QIDQR SNHRT QMQTC TRMWT STSAS LTPXR BWJYR SRIYT RJMRI
aia d esdeo a  or e er  ro ur ara a  ri o  un o aos r on os

QILQN YRIDR ICPJU QPTRI IRDRP ITSCS ZPYRI IQDQP VSTSA KPMST
es e   osdo s in  eiros sodoi sra a  i os sedei  ara   i ar

SLTPY STCQD TSTPJ DRDRI URAQJ IEWQK WXPSA SMUWH S
a ri  ar ed rarin dodos  o en s ue  u ia  a  u  a
```

A simples observação desta semi-tradução e as letras que nos faltam atribuir levam-nos (depois de algumas tentativas falhadas[8]) a concluir que

$$\mathtt{E}(\mathtt{t}) = \mathtt{Y} \quad \mathtt{E}(\mathtt{m}) = \mathtt{A}.$$

Agora, simplesmente pelo sentido, consegue-se completar o texto, que se lê então:

"Súbitas rajadas de vento bufaram do Sul. Com estardalhaço, uma chapa de zinco, vinda não se sabe de onde, voou de um lado da estrada, deu quatro pinotes grotescos e foi engarrafar-se, silenciosa e miserável, na valeta do outro lado. Logo uma bátega varreu a estrada. Os homens, já encharcados pelos chuviscos que caíam desde o alvorecer, procuraram abrigo junto aos troncos esbeltos dos pinheiros. Só dois rapazitos se deixaram ficar a britar pedra, rindo dos homens que fugiam à chuva."

[8] Para as quais é de muita utilidade o modo `decipher` do GNU-Emacs [Sta02].. (ver Apêndice F)

Draft 2021.1

Figura 3.14: al-Kindī

Este tipo de ataque a uma cifra monoalfabética é conhecido desde há muito. Um livro redescoberto em 1987 do matemático árabe *Abū Yūsūf Ya'qūb ibn Is-hāq ibn as-Sabbāh ibn 'omrān ibn Ismail al-Kindī* (801-873) [9] com o título

"Um manuscrito sobre como decifrar mensagens criptográficas" discute pormenorizadamente o ataque a uma cifra mono-alfabética com recurso a tabelas de frequência de caracteres. Apesar disso, na Europa, só no século XVI é que as cifras monoalfabéticas caiem em desuso pela evidência da sua fragilidade.

A cifra do Kama-Sutra

O *Kama-Sutra*, famoso livro de escritos eróticos de Vatsyayana, inclui a Criptografia ("a escrita secreta") como uma das 64 artes, mais propriamente a número 45. Com o fito de permitir às mulheres poder escrever mensagens sem que olhos indiscretos possam descrutinar o seu sentido, é descrita uma cifra simples, mas bastante mais segura que a contemporânea cifra de César (pag. 62), o que nos leva a crer que ou os segredos de alcova eram bem mais sérios do que os segredos militares do romanos, ou a literacia na Índia, era muito mais avançada e o simples facto de uma mensagem ser escrita a tornava inacessível para a maioria dos militares romanos. A cifra descrita no *Kama-Sutra* é uma simples cifra monoalfabética em que a permutação de substituição monoalfabética é uma permutação composta somente por ciclos de tamanho 2. O normal é escrever a chave da cifra como um alfabeto, permutado, em duas linhas.

[9] De facto:

أبو يوسف يعقوب بن إسحاق الصباح الكندي

Draft 2022.1

Figura 3.15: Primeira página do texto de al-Kindī.

```
h a r u n l c i d e f g j
k m o p q s t v w x y z b
```

Os caracteres que constam em cada coluna constituem os pares da permutação, ou seja, cada caracter é cifrado pelo caracter corespondente na mesma coluna. Claramente, esta cifra é uma involução, ou seja se aplicarmos a cifra ao texto já cifrado, voltamos a obter o texto original.

Cmo exemplo, se cifrarmos, com a chave anterior. o texto

```
A Maria tem um cordeirinho
```

o resultado será

```
m amovm cxa pa trowxvovqkr.
```

"Remendos" às cifras mono-alfabéticas

A primeira tentativa de complicar a vida aos criptanalistas consiste na introdução de alguns símbolos que não representam nenhum caracter na mensagem inicial e que têm

Draft 2021.1

como simples propósito complicar a análise de frequências. Estes símbolos chamam-se usualmente **nulos** e eram aleatoriamente introduzidos nos criptogramas por forma a que, na contagem das ocorrências de cada carácter, se pudessem confundir com a ocorrência estatística de um outro, dificultando assim a decifração.

Uma solução mais radical consiste em usar não um único símbolo para cada carácter, mas sim um conjunto cujo cardinal seja directamente proporcional com a frequência relativa desse carácter, a chamada *cifra homofónica*. Suponhamos que vamos cifrar cada carácter por um par de dígitos, então como na língua portuguesa o carácter a corresponde a cerca de 14% do total de caracteres, vamos utilizar 14 diferentes símbolos para cifrar a. No momento de decifrar qualquer um desse símbolos é equivalente, e decifra-se da mesma forma como a. Um exemplo deste tipo de cifra é a apresentada pela tabela 3.3

a	b	c	d	e	f	g	h	i	j	k	l	m
32	04	14	81	16	89	22	53	06	19	87	00	63
50		60	21	17				18			96	38
85		25	56	09				07			23	12
35		68	78	44				70				11
76		30		03				84				
15				45				92				
10				46				71				
39				74								
37				57								
26				97								
67				42								
61				48								
73												

n	o	p	q	r	s	t	u	v	w	x	y	z
02	08	40	41	95	51	49	98	69	72	58	55	28
90	34	47		64	81	33	01					
95	05	93		65	62	31	99					
36	86			54	83	24	88					
29	77			59	27							
	79			66	91							
	52			13	20							
	43											
	94											

Tabela 3.3: Uma cifra monoalfabética que pretende ter uma distribuição de frequências "plana".

Quando se cifra um carácter para a qual existe mais do que um correspondente cifrado, deve ser escolhido *aleatoriamente* o elemento a utilizar. O texto cifrado deve ter, tendencialmente, uma estatística equidistribuída, o que complica muito um ataque nos termos em que foi descrito anteriormente.

Outra solução a que se começa a recorrer nos séculos XVI e XVII é aos chamados *nomenclators*. Livros de cifra que codificam não caracteres mas sim palavras inteiras. Neste caso dizemos que se trata de um código em vez de uma cifra. Os problemas deste tipo de solução são óbvios:

- Em vez de uma simples chave é necessário um verdadeiro dicionário de grande volume se se pretender um léxico suficientemente abrangente.
- Os *nomenclators* são de difícil e demorada geração, o que conduz a que se renovem menos frequentemente, e consequentemente, a descoberta de um destes livros pelo

Draft 2021

adversário tenha consequências, muito mais drásticas.

- Os *nomenclators* são muito mais difíceis de distribuir e esconder.

Para obviar este tipo de problemas alguns usaram técnicas mistas, em que algumas palavras (as mais frequentes) são codificadas e o resto das palavras são cifradas com uma cifra monoalfabética.

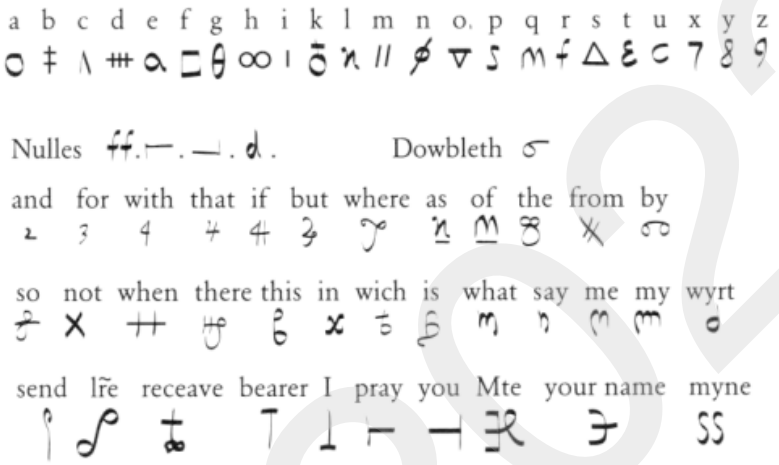


Figura 3.16: A cifra usada por Maria Stuart

A utilização mais famosa é a dos partidários de Maria Stuart. A utilização de um código simples e de uma cifra monoalfabética, facilmente quebrada pelos partidários da Rainha, levou à montagem de uma mensagem armadilhada que levou Maria Stuart à condenação por traição à coroa e daí ao cadafalso[10].

3.3 Cifras de transposição

Nas cifras apresentadas até agora cada letra do texto original é substituída por outra, no texto cifrado, mas não há lugar a nenhum deslocamento. As posições relativas dos símbolos no texto original são preservadas no texto cifrado. Se este princípio não for respeitado dizemos que se dá lugar a uma *transposição*. Algumas cifras assentam somente na aplicação deste princípio.

[10] Lição a tirar: mais vale escolher criteriosamente a cifra a usar antes de conspirar contra uma rainha como a Isabel I... O episódio histórico é descrito com algum pormenor por Singh [Sin99]

Draft 2021.1

3.3.1 Cifra de transposição por colunas.

A aplicação mais simples deste princípio consiste na escrita do texto a cifrar numa tabela com um número pré-determinado de colunas. Seja o número de colunas 7 e o texto da mensagem ‘‘`Comprar todo o chocolate de leite com avelãs`’’. Então se escrevermos os texto numa tabela com 7 colunas, e acrescentarmos os *nulos* suficientes para preencher a última linha[11], temos:

```
c o m p r a r
t o d o o c h
o c o l a t e
d e l e i t e
c o m a v e l
a s a b c d e
```

Agora transcrevemos o texto agrupando-o em colunas. Ou seja o texto cifrado resulta agora (escrito em pentagramas):

```
CTODC SOOCE OSMDO LMAPO LEABR OAIVC ACTTE DRHEE LE.
```

O processo de decifração é igualmente simples, bastando escrever o criptograma em colunas de tamanho 6, por forma a constituir um rectângulo de 7 colunas.

Apesar desta cifra não ser vulnerável a uma análise de frequências, um ataque corresponde à simples adivinhação do número de colunas, portanto bastante simples. O número de colunas nunca pode ser muito elevado, pois não deve ser comparável ao comprimento total da mensagem, sob pena de o criptograma não permutar suficientemente as letras da mensagem e tornar aparente o seu conteúdo.

Para tornar a cifra mais eficaz temos que tornar o espaço de chaves maior. Uma das alternativas mais usuais é a de usar uma palavra como chave da cifra. Por exemplo: `tablete`. A chave induz uma permutação de colunas da seguinte forma: atribuímos sucessivamente a cada letra da chave, inteiros consecutivos, por forma a preservar a ordem alfabética. Quando uma letra ocorre mais do que uma vez, atribui-se um inteiro menor à ocorrência mais à esquerda.

t	a	b	l	e	t	e
6	1	2	5	3	7	4

Estes inteiros, 6125374, podem ser então atribuídos a cada uma das colunas da tabela anterior:

[11] A este processo normalmente chama-se *“padding”*.

Draft 2021.1

6	1	2	5	3	7	4
c	o	m	p	r	a	r
t	o	d	o	o	c	h
o	c	o	l	a	t	e
d	e	l	e	i	t	e
c	o	m	a	v	e	l
a	s	a	b	c	d	e

Se agora transcrevermos o texto, coluna a coluna, respeitando a ordem dada pela sequência gerada pela chave, vem:

```
OOCEO SMDOL MAROA IVCRH EELEP OLEAB CTODC AACTT ED
```

3.3.2 Criptanálise de uma cifra de transposição por colunas

Suponhamos então que temos o seguinte criptograma:

```
UAAUE EISAI AODQE SJMNC DAUAR ROSSI AAAAM RIRNR NRCCP SNPTG
OAOTI E
```

Como o tamanho do criptograma é 56 podemos suspeitar que o número de colunas é 7, 8 ou 4, 2 é um número demasiado pequeno para colunas[12]. Se o número de colunas for 8 então temos as colunas

U	S	E	A	S	R	C	G
A	A	S	U	I	I	C	O
A	I	J	A	A	R	P	A
U	A	M	R	A	N	S	O
E	O	N	R	A	R	N	T
E	D	C	O	A	N	P	I
I	Q	D	S	M	R	T	E

Mas podemos observar que a última linha tem somente duas vogais e a segunda somente duas consoantes, distribuições bastante pouco prováveis (ver tabela D.2). Da mesma forma se o número de colunas for 4, ocorrem linhas somente com vogais, o que não é muito provável. Se tentarmos com 7 colunas, obtemos:

[12] Estamos a supor que a cifra é de tal forma que todas as colunas são do mesmo tamanho. Não sendo esse o caso, como vamos ver à frente, o ataque é mais custoso, mas não impossível.

Draft 2021.1

U	A	J	R	A	N	T
A	I	M	R	A	R	G
A	A	N	O	M	C	O
U	O	C	S	R	C	A
E	D	D	S	I	P	O
E	Q	A	I	R	S	T
I	E	U	A	N	N	I
S	S	A	A	R	P	E

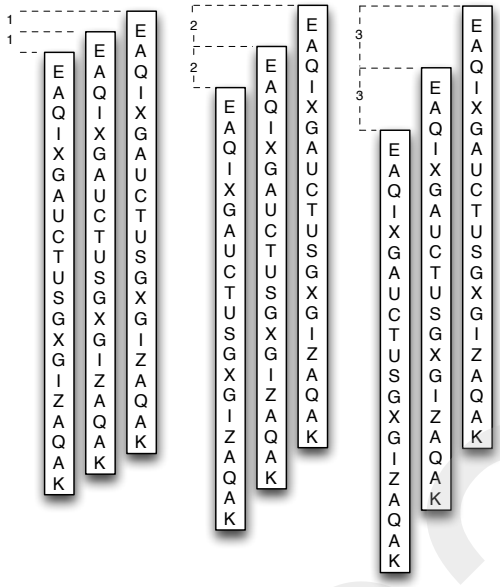
A distribuição de vogais e consoantes pelas diversas colunas parece razoável, pelo que nos resta tentar encontrar a ordem certa para as colunas. Notemos que a primeira linha tem que constituir palavras, assim como nas restante linhas as sequências de letras têm que ser passiveis de constituir palavras. Depois de algumas tentativas chegamos a

J U N T A R A
M A R G A R I
N A C O M O A
C U C A R S O
D E P O I S D
A E S T R I Q
U I N I N A E
A S P E R A S

3.3.3 Criptanálise de cifras de transposição com "fitas deslizantes"

Uma outra forma de atacar este tipo de cifra, assenta no conhecimento dos digrafos e trigrafos mais frequentes (vêr as tabelas D.6 e D.7). O método é facilmente ilustrável da seguinte forma. Suponhamos que temos diversas cópias do criptograma escritas em colunas de largura 1. Alinhemos estas colunas contendo o texto, e suponhamos que podemos fazer variar o seu alinhamento longitudinalmentecomo. Agora vamos alinha-las, primeiro com um deslocamento de uma posição, depois com um deslocamento de duas posições, etc...

Draft 2021.1



Quando o desfazamento das colunas coincidir com o número de colunas da tabela inicial, as frequências dos digrafos e trigrafos devem coincidir com as mesmas estatísticas na língua original. Claro que método funciona melhor, como todos os métodos baseados nas frequências médias dos textos orignais, com cpriptogramas de tamanho razoável.

3.3.4 Cifra de transposição por colunas sem *"padding"*

Se não "enchermos" a última coluna com *nulos*, como fizemos no exemplo da página 73, obtemos uma cifra que é mais difícil de atacar. Usemos a mesma chave. Escrevemos então o texto numa tabela de 7 colunas:

6	1	2	5	3	7	4
c	o	m	p	r	a	r
t	o	d	o	o	c	h
o	c	o	l	a	t	e
d	e	l	e	i	t	e
c	o	m	a	v	e	l
a	s					

Transcrevendo o texto, coluna a coluna, usando a ordem dada pela chave, temos:

```
OOCEO SMDOL MROAI VRHEE LPOLE ACTOD CAACT TE
```

Para decifrar só temos que fazer o percurso inverso, tendo em atenção que como o comprimento da mensagem é $37 = 7 \times 5 + 2$, as colunas 6 e 1 terão comprimento 6, enquanto as restantes terão comprimento 5.

Draft 2021.1

3.3.5 Criptanálise de uma cifra de transposição sem "padding"

Suponhamos que nos deparamos com o seguinte criptograma:

```
AHSUE OSETA SREAN TCTII TDTNR IQAHE TAEEA CJENO OEMMT DOODE
DCDIA NEOEC TILLE EEEMM IASAE DNSI
```

Sabemos que a cifra usada é de transposição de colunas, mas como suspeitamos que não a mensagem não preenche integralmente todas as linha da grelha, o seu tamanho não nos fornece pistas sobre o tamanho da chave. De qualquer forma, e por outros meios, conseguimos saber que a palavra "CHOCOLATES" faz parte da mensagem. Apesar de poder parecer artificial, este facto de conhecermos antecipadamente uma parte[13] (neste caso uma palavra) da mensagem, é bastante usual nas condições reais de aplicação da criptanálise. A grande maioria das vezes as mensagens seguem um formato normalizado, sabemos sobre que assunto vão versar, ou simplesmente contêm obrigatoriamente um preâmbulo de um formato conhecido. Como sabemos que a palavra "CHOCOLATES" está presente na mensagem podemos tentar ver se isso nos indica o tamanho da chave. Estudemos as possibilidades de a chave ter um comprimento entre 5 e 9.

- Suponhamos que a chave tem 5 letras. Então a palavra "CHOCOLATES" vai originar um conjunto de 5 digrafos que serão preservados pela cifra, seja qual for a posição em que se encontre

  ```
  CHOCO    .CHOC    ..CHO    ...CH    ....C
  LATES    OLATE    COLAT    OCOLA    HOCOL
  .....    S....    ES...    TES..    ATES.
  ```

  Os digrafos CL, HA, OT, CE e OS, estariam necessariamente presentes no criptograma. Quais as ocorrências destes padrões na mensagem intersectada?

CL 0	HA 0	OT 0	CE 0	OS 1

  A chave não pode portanto ter comprimento 5.

- Se a chave tiver comprimento 6, então o "*crib*" aparece como

  ```
  CHOCOL
  ATES..
  ```

[13] A um pedaço de texto que sabemos fazer parte de uma mensagem que tentamos quebrar é chamado usualmente "*crib*". Optamos por não tentar traduzir este termo usado normalmente na literatura, pois os seus diferentes significados em Inglês de "encaixe", "berço" e "cábula", são impossíveis são de muito difícil reprodução numa única palavra portuguesa.

Draft 2021.1

Os digrafos gerados têm as seguintes ocorrências no criptograma:

CA	0	HT	0	OE	2	CS	0

Pelo que a chave não pode ter comprimento 6.

- Se a chave tiver comprimento 7, então temos

```
CHOCOLA
TES....
```

Que corresponde à existência dos digrafos CT, HE e OS que têm as seguintes ocorrências no criptograma:

CT	2	HE	1	OS	1

A chave de comprimento 7 é compatível com o "crib"!!!

- Se a chave tiver comprimento 8, então temos

```
CHOCOLAT
ES......
```

Então os correspondentes digrafos têm as seguintes ocorrências:

CE	0	HS	1

Logo a chave não pode ter comprimento 8.

- Se a chave tiver comprimento 9, então temos

```
CHOCOLATE
S........
```

Mas o digrama CS não ocorre no criptograma, pelo que a chave não pode ter este comprimento.

Portanto o único comprimento de chave (entre 5 e 9) compatível com a existência do *crib* CHOCOLATES é 7. Muito provavelmente é este o número de colunas. Como o comprimento da mensagem é $79 = 7 \times 11 + 2$, temos que tentar descobrir quais as 2 colunas que são constituídas por 12 letras (as restantes têm 11). Temos que analisar as distâncias entre as

Draft 2022.1

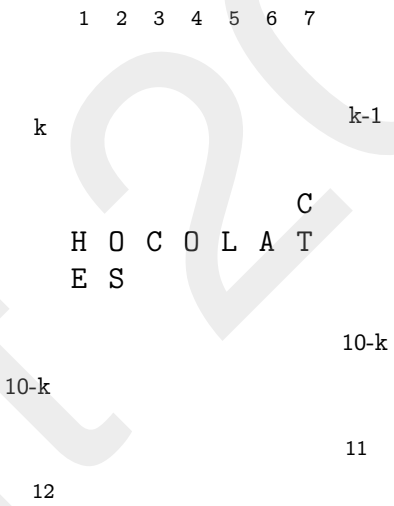
ocorrências no criptograma dos diversos digrafos gerados e tentar ver quais as compatíveis com as diversas possibilidades de colocação da primeira letra da palavra CHOCOLATES.

$$|\leftarrow 5 \rightarrow \texttt{OS} \leftarrow 9 \rightarrow \texttt{CT} \leftarrow 10 \rightarrow \texttt{HE} \leftarrow 29 \rightarrow \texttt{CT} \leftarrow 18 \rightarrow|$$

A palavra CHOCOLATES pode começar em qualquer das 7 colunas, pelo que temos outros casos a analisar:

```
.......  .......  .......  .......  .......  .......  .......
CHOCOLA  .CHOCOL  ..CHOCO  ...CHOC  ....CHO  .....CH  ......C
TES....  ATES...  LATES..  OLATES.  COLATES  OCOLATE  HOCOLAT
.......  .......  .......  .......  .......  S......  ES.....
.......  .......  .......  .......  .......  .......  .......
```

Primeiro observemos a que distâncias devemos esperar as diversas letras da palavra. Examinemos, por exemplo, a tabela correspondente ao último caso:



Quais são então os casos que temos que analisar?

Casos 3, 4, 5 e 6: A coluna que contém CT assim como a coluna que contém HE têm comprimento 11, pelo que a distância entre ambos os padrões não pode ser 9, 29 nem 69. Portanto não são compatíveis com o que foi observado.

Casos 1, 2 e 7: No caso 2 CT está numa coluna de tamanho 12 e HE e OS em colunas de tamanho 11. A distância entre OS e CT pode ser 9 e a distância entre CT e HE pode ser 10. Esta hipótese é compatível com os dados. O mesmos acontece para os casos 1 e 7.

Draft 2021.1

Caso 1: Colunas com tamanho 12 contêm os padrões CT e HE, e as restantes colunas são de tamanho 11. Então HE tem que distar 10 letras de C, O, L ou A. E o que acontece é exactamente isso:

$$\cdots \texttt{HE} \leftarrow 10 \rightarrow \texttt{O} \leftarrow 10 \rightarrow \texttt{C} \leftarrow 10 \rightarrow \texttt{L} \leftarrow 10 \rightarrow \texttt{A} \cdots$$

A solução seria então ordenar as colunas pela seguinte ordem:

3 1 2 5 4 6 7

Caso 2: CT encontra-se numa coluna com tamanho 12, assim como o A, os restantes estão em colunas de tamanho 11. Então a uma distância de 10 caracteres de HE pode encontrar-se o caracter A (o que não acontece) ou pode encontrar-se a uma distância 9 o caracter C, O ou L. O caracter O encontra-se em tal situação. Mas tem-se

$$\cdots \texttt{HE} \leftarrow 9 \rightarrow \texttt{O} \leftarrow 9 \rightarrow \texttt{ED} \cdots$$

Ora depois do caracter O teria que ocorrer o caracter A a uma distância 10, ou um dos caracteres C ou L a uma distância 9. Como tal não acontece, sabemos que este caso também não é compatível com a observação.

Caso 7: As colunas com tamanho 12 conteriam HE e OS e as restantes seriam de tamanho 11. Pelo que já vimos na análise do caso 1, sabemos que as restantes letras se encontram em posições compatíveis. A ordenação correspondente será então:

2 7 1 4 3 5 6

Para decifrar a mensagem basta agora tentar as duas possibilidades.

```
R N A O C O M      N A O C O M E
E R H O J E M      R H O J E M A
A I S D E C I      I S D E C I N
N Q U E N T A      Q U E N T A T
T A E D O I S      A E D O I S C
C H O C O L A      H O C O L A T
T E S D E L E      E S D E L E I
I T E I M E D      T E I M E D I
I A T A M E N      A T A M E N T
T E A N T E S      E A N T E S D
D E S E D E I      E S E D E I T
T A                A R
```


Draft 2021.1

Facilmente se conclui então que é a segunda hipótese a que decifra o criptograma.[14]

3.4 Cifras polialfabéticas

A fraqueza fundamental das cifras mono-alfabéticas assentam no facto de cada letra ser cifrada da mesma maneira ao longo de todo o texto. Para contornar este problema **Battista Alberti** (1404–1472) em 1460, ao serviço do Vaticano, cria uma nova cifra que não enferma dessa deficiência. Para que cada letra não seja sempre codificada da mesma forma, e portanto se torne vulnerável a um ataque de frequências, passam-se a utilizar **dois** alfabetos em vez de **um**:

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
F	Z	B	V	K	I	X	A	Y	M	L	P	L	S	D	H	J	O	R	G	N	Q	C	U	T	W
G	O	X	B	F	W	T	H	Q	I	L	A	P	Z	J	D	E	S	V	Y	C	R	K	U	H	N

e as letras passam a ser cifradas alternadamente com um ou o outro dos alfabetos.

A palavra `hello` é cifrada em `AFPAD`. É importante notar que enquanto o primeiro `l` é cifrado como `P` o segundo é cifrado como `A`. Portanto a curva de frequências dos diversos caracteres não vai ser preservado no texto cifrado. O problema é que todas as letras que ocorrem no texto original em posições de ordem ímpar (o mesmo acontecendo para as que ocorrem em posições de ordem par) são cifradas pela mesma cifra monoalfabética. Portanto, as estatísticas dos correspondentes cifrados de cada um destes grupos de letras, vão preservar a distribuição de ocorrências da mensagem original. O processo de quebra da cifra é um pouco mais trabalhoso, mas é basicamente equivalente ao das cifras monoalfabéticas.

3.4.1 A cifra de Vigenère

Apesar de **Alberti** ter descoberto o princípio que permitiria voltar a criptografia a levar a melhor à criptanálise, não levou essa descoberta até às últimas consequências. De facto se soubermos antecipadamente que a cifra usada é deste tipo podemos ainda fazer uma análise de frequências, dividindo o texto em em dois subtextos, um constituído pelas letras que ocupam as posições pares e outro com aquelas que ocupam as posições de ordem ímpar.

Esta ideia só é usada com sucesso pelos trabalhos de **Johannes Trithemius** (1462–1516), o de **Giovanni Porta** (1535–1615) e finalmente de **Blaise Vigenère** (1523–1596). Em vez de dois alfabetos consideremos uma tabela com os 26 possíveis alfabetos da cifra de **César**.

[14] Este criptograma ensina-nos, para além de uma muito evidente regra de prudência nos hábitos alimentares, que o uso destes métodos de cifra pode ser tão enfadonho, que depois de efectuadas todas as operações de cifra, ao operador, pode não restar força anímica suficiente para corrigir um simples erro ortográfico!

Draft 2021

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
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	A
C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	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

Tabela 3.4: A tabela de Vigenère

Draft 2021.1

Figura 3.17: Johannes Trithemius

Se para cifrar uma mensagem fosse utilizado somente um dos alfabetos, a cifra era *muito fraca...*

A cifra de **Vigenère** vai usar uma linha **diferente** para cifrar cada letra, dificultando muito uma análise de frequências. Para cifrar um texto começamos por escolher uma chave, por exemplo a palavra `chumbo`. Esta palavra não tem letras repetidas mas poderia ter sido escolhida uma em que isso acontecesse (haveria mesmo alguma vantagem para apresentar maior resistência a alguns tipos de ataques). Suponhamos que o texto a cifrar é:

"`Fujam enquanto é tempo!  Eles vêm ai!`".

Como sempre, e pelas razões óbvias, começamos por remover os caracteres *espaço* e todos os símbolos de pontuação. Fazemos corresponder a chave a cada um dos caracteres da mensagem (repetindo a chave):

```
chumbochumbochumbochumbochum
fujamenquantoetempoelesvemai
```

agora, para cada letra usamos a linha que começa com a letra da chave que que corresponde. Assim vem:

```
chumbochumbochumbochumbochum
fujamenquantoetempoelesvemai
HBDMNSPXOMOHQLNQNDQLFQTJGTUU
```

Draft 2022.1

Ataque de Babbage a uma cifra de Vigenère

O primeiro ataque sistemático, bem sucedido, à cifra de Vigenère, apesar desse facto só ter sido conhecido em meados do sec.XX, deve-se a **Charles Babbage** (1791–1871).

O método é facilmente ilustrável com um exemplo. Suponhamos que o texto:

```
A panela com asas tem arroz com grão.
```

é cifrado com a chave `xaile`.

```
xailexailexailexailexailexail
apanelacomasastemarrozcomgrao
XPIYIIAKZQXSIDXBMICVLZKZQDRIZ
```

A palavra `com` é cifrada como `KZQ` na primeira vez, assim como na segunda instância. Esta repetição prende-se com o facto de a distância entre as duas instâncias ser múltiplo do tamanho da chave.

Babbage observou que este simples facto poderia ser o suficiente para um ataque bem sucedido à cifra de Vigenère. Numa primeira fase examina-se o texto cifrado, para tentar encontrar sequências de caracteres que apareçam repetidas. Essas repetições, podem resultar de diferentes textos, cifrados com diferentes alfabetos, ou os mesmos textos, cifrados com a mesma parte da chave. Se forem usados padrões de tamanho considerável, a probabilidade do primeiro caso é diminuta. Podemos considerar somente padrões de tamanho pelo menos 4.

Consideremos então o seguinte texto cifrado:

```
WUBEF IQLZU RMVOF EHMYM WTIXC GTMPI FKRZU PMVOI RQMMW OZMPU
LMBNY VQQQM VMVJL EYMHF EFNZP SDLPP SDLPE VQMWC XYMDA VQEEF
IQCAY TQOWC XYMWM SEMEF CWYEY QETRL IQYCG MTWCW FBSWY FPLRX
TQYEE XMRUL UKSGW FPTLR QAERL UVPMV YQYCX TWFQL MTELS FJPQE
HMOZC IWCIW FPZSL MAEZI QVLQM ZVPPX AWCSM ZMORV GVVQS ZETRL
QZPBJ AZVQI YXEWW OICCG DWHQM MVOWS GNTJP FPPAY BIYBJ UTWRL
QKLLL MDPYV ACDCF QNZPI FPPKS DVPTI DGXMQ QVEBM QALKE ZMGCV
KUZKI ZBZLI UAMMV Z
```

Por exemplo a sequência `WCXYM` aparece repetida depois de 20 caracteres. Como 20 tem factores 1, 2, 4, 5, 10 e 20, são esses os diversos tamanhos possíveis de chave.

Podemos ver para os diversos padrões quais os correspondentes tamanhos de chave compatível:

Draft 2021

		1	2	3	4	5	6	8	10	12	15	19	20
EFIQ	95	x				x						x	
PSDLP	5	x				x							
WCXYM	20	x			x	x							x
ETRL	120	x	x	x	x	x	x	x	x	x	x		x

Se assumirmos que a chave tem tamanho 5, o passo seguinte é o descobrir quais as letras da chave $L_1L_2L_3L_4L_5$. Mas notemos então que o texto vai ser dividido em 5 partes:

$$\{a_{5n} \mid n \in \mathbb{N}\}, \{a_{5n+1} \mid n \in \mathbb{N}\}, ..., \{a_{5n+4} \mid n \in \mathbb{N}\}$$

que correspondem a cada uma das letras da chave, e que vão ser cifradas com um só alfabeto (e que neste caso é um alfabeto especialmente simples...)

3.4.2 Análise estatística de um criptograma

Dado um criptograma cifrado por uma cifra monoalfabética, é possível determinar com algum grau de certeza a sua língua de origem, sem ter que se descobrir o que lá está cifrado. Existe um conjunto de valores estatísticos que são preservados por este tipo de cifras e que são características das línguas de origem.

Definição e invariância de K

Dado um par de textos $T = (t_1, t_2, \ldots, t_N)$, $T' = (t'_1, t'_2, \ldots, t'_N)$ de igual comprimento N e sobre um léxico comum L, à frequência relativa de se encontrar nos dois textos o mesmo carácter na mesma posição chamamos **índice de coincidência** dos dois textos que representamos por $K(T, T')$. Portanto

$$K(T, T') = \sum_{i=1}^{N} \frac{\delta(t_i, t'_i)}{N}$$

em que

$$\delta(x, y) = \begin{cases} 1 & \text{se } x = y \\ 0 & \text{caso contário.} \end{cases}$$

É trivial observar que, $K(T, T') \leq 1$, com $K(T, T') = 1$ sse $T = T'$.

A observação empírica mostra que para textos suficientemente longos T e T′ escritos na mesma língua $\mathcal{S}$, os valores de $K(T, T')$ aproximam-se de uma constante $K_{\mathcal{S}}$.

De alguma forma o valor de $K_{\mathcal{S}}$ parece reflectir o grau de redundância da língua. Os valores da tabela 3.5 foram extraídos de [Bau97] excepto para o Português que foi obtido analisando 8.8MB de textos de jornal.

Draft 2021.1

$\mathcal{S}$	$K_{\mathcal{S}}$
Alemão	7.62%
Castelhano	7.75%
Francês	7.78%
Inglês	6.61%
Italiano	7.38%
Português	7.77%
Russo	5.28%

Tabela 3.5: Diversos valores indicativos de $K_{\mathcal{S}}$.

Teorema 1 *Para qualquer cifra polialfabética* E *(das que foram estudadas) e textos* T *e* T'*,*

$$K(T, T') = K(E(T), E(T')).$$

O valor espectável para $K(T, T')$ com T e T' textos de comprimento N, pode ser calculado a a partir das probabilidades p_i de ocorrência do caracter i numa dada posição nos textos. Portanto

$$K(T, T') \approx \sum_{i=a}^{z} p_i^2.$$

Definição e invariância de Ψ.

Seja T um texto de comprimento N, com $(m_i)_i$ frequências do diversos caracteres i em T. Definimos

$$\Psi(T) = \sum_{i=a}^{z} \frac{m_i^2}{N^2}.$$

Teorema 2 *Para qualquer cifra polialfabética* E *(das que foram estudadas) e texto* T *tem-se*

$$\Psi(T) = \Psi(E(T)).$$

Observemos que

$$\Psi(T) = \left(\sum_{i=a}^{z} \left(\frac{m_i}{N} - \frac{1}{26} \right)^2 \right) + \frac{1}{26},$$

Draft 2022.1

pois

$$\begin{aligned}\sum_{i=a}^{z}\left(\frac{m_i}{N}-\frac{1}{26}\right)^2 &= \sum_{i=a}^{z}\left(\left(\frac{m_i}{N}\right)^2-2\left(\frac{m_i}{N}\right)\left(\frac{1}{26}\right)+\left(\frac{1}{26}\right)^2\right)\\ &= \sum_{i=a}^{z}\left(\frac{m_i}{N}\right)^2-\frac{2}{26}\sum_{i=a}^{z}\left(\frac{m_i}{N}\right)+\sum_{i=a}^{z}\frac{1}{26^2}\\ &= \Psi(T)-\frac{2}{26}\times 1+26\times\frac{1}{26^2}\\ &= \Psi(T)-\frac{1}{26}\end{aligned}$$

Podemos portanto encarar $\Psi(T)$ como uma medida da "irregularidade" da distribuição de frequências dos diversos caracteres na mensagem T. Para uma mensagem com uma distribuição absolutamente equidistribuída pelas diversas letras temos

$$\Psi(T)=\sum_{i=a}^{z}\left(\frac{1}{26}-\frac{1}{26}\right)^2+\frac{1}{26}=\frac{1}{26}\approx 0.038$$

Como para uma cifra monoalfabética a distribuição de frequências é muito menos uniforme do que para uma cifra polialfabética (Porquê?), podemos dizer que

$$0.038\leq\Psi(T)\leq K_{\mathcal{S}}.$$

Em que $K_{\mathcal{S}}$ é a constante dependente da língua original da mensagem referida na tabela 3.5. $\Psi(T)$ serve portanto um indicador do número de alfabetos usados na cifra da Mensagem que deu origem a T. Quanto mais afastado de $K_{\mathcal{S}}$ maior o número de alfabetos usados.

O teste K de Friedman

Seja $T^{(\rho)}$ o texto T ao qual aplicamos um deslocamento de ρ letras.

Se ρ for um múltiplo do período da cifra, por exemplo $\rho = 2p$, com p o período da cifra. Temos então

$$T^{(\rho)}=(t_{2\rho+1},t_{2\rho+2},\ldots,t_N,t_1,\ldots,t_{2\rho}).$$

Se fizermos sobrepor os dois textos, T e $T^{(\rho)}$ temos

$$\begin{array}{cccccccc} t_1 & t_2 & \ldots & t_{N-2\rho} & t_{N-2\rho+1} & \ldots & t_N\\ t_{2\rho+1} & t_{2\rho+2} & \ldots & t_N & t_1 & \ldots & t_{2\rho}\end{array}$$

Cada um dos caracteres de ambos os textos foi cifrado com o mesmo alfabeto. Pelo que se estudarmos o comportamento de $K(T^{(\rho)},T)$ o período da cifra (τ) deve ser revelado com valores de $K(T^{(k\tau)},T)\approx K_{\mathcal{S}}$ para qualquer $k>0$.

Draft 2021.1

Exemplo de estimativa do período da cifra de um criptograma

Consideremos o seguinte criptograma C que sabemos ser de um texto original em português:

```
FPCFN BMHRV PMFCI MWSFF RRVGI VITNA UMZMG YVZSE IYMZE FPCCC
OILRQ AOIBB RHEOM IKWQS IVQNE JABHB WETPC QAEKU WPLYM SSBMP
DTRCE EWSPL SVYHI GGKLD RMKTN PULOE WHPMM OZBFI IMLXC QCGWE
XHSNF SFJEE LMNPI IWSSB WWDPV VFSIS TFMAW TSYEN SOLWR RVSPE
RFKBQ HRCNI FVNPF EULGF ZNILX CLBII PSLGQ XUDQA UNITT TFXBV
MITLT ZZNHH YOLCY CGIBV CBUVI GMGKK FORZO MMVQA ALTBT TVCEY
RXRRT QCTSG ZBZTZ GWARX YMVET LOAWT EWSTG AIQNI GXTPU DMDYT
XTZJF OBWWI UFSGM AWTSO LSFBR GGMEZ JFARQ HYTTN YWGIF EKYDR
WHXKE UARVW PYIEE ZEJXN VTEFP SCWPE VEQLB IITEH RFPVR QQIYV
QYTRZ BQHVG TRVVF IGSFZ MZVTS WITWE ZBRVF MPZUF RBKHG QXODI
EWXPJ LCFUH QTGQY TLANS GSRPD ZZRMM SGZNK MZJXD CAEEI FXKEV
LMVVG STPGX DFIIE GGCOO KWGEE MUEAV UGSWS UZSAW TSLHG DTRAH
TXVNT GRMKM LXGXT RUOIF QWTTR AEILI TGAJY HEGXQ LUKQY MSEEL
OUMAY GSXLL VVGIH PCEEI IYILU WPRUW FSYVG FUDIZ MTPIT AEWTI
FIQOA GMERT IUBUV ZQEJY GZBIQ TSNQQ FRZVU STWWM SKQGY BPQAO
IZVGT VFZCF AGEGS FPCFZ EIKHC DEXCA HTTCC TVLNT TVVTD RNEIG
XGLOD IEWXP JLCVA NVIIK IOKWC SKWGF LRLBG HRVTN LINVX GWAEI
IEHTM PEEID RRVEQ NIICE KBGCB UVWIE BEHLS KIEHH WTPLM IQSLR
QDPIW KMFSU XEJMF SIPCY TVTIS EXCLT IMVRT VJZJV LRQTR JL
```

A análise do valor de $\Psi(C) = 4.419$ mostra que a cifra usada não pode se monoalfabética pois este valor é muito inferior aos 7.77% esperados (ver tabela 3.5).

Podemos então proceder a um ataque de Babbage (também chamado método de Kapirski) para estimar o tamanho da chave.

Tomemos então os diversos padrões (de tamanho mínimo 4) e analisemos os seu divisores (só foram considerados os divisores menores que 17, pois não é natural que a chave tenha um tamanho superior):

Draft 2021

	2	3	6	5	7	9	10	11	12	13	14	15	17
AWTS	x				x						x		
XPJL	x	x	x			x			x				x
QAOI	x	x			x	x		x					
PJLC	x	x	x			x			x				x
ODIE	x	x	x			x			x				x
MAWT	x				x					x	x		
LBII		x				x							
IEWX	x	x	x			x			x				x
FPCF		x		x		x						x	x
EXXP	x	x	x			x			x				x
DIEW	x	x	x			x			x				x
CBUV		x				x							

Apesar de nos dar algumas boas pistas de qual o tamanho da chave, não é inteiramente claro. Tentemos então o teste de Friedman. Temos que calcular os diversos valores de $(K(C, C^{(n)})_n$, em que $C^{(n)}$ representa o texto C operado por n *shifts* para a esquerda $(C^{(n)} = C \lll n)$.

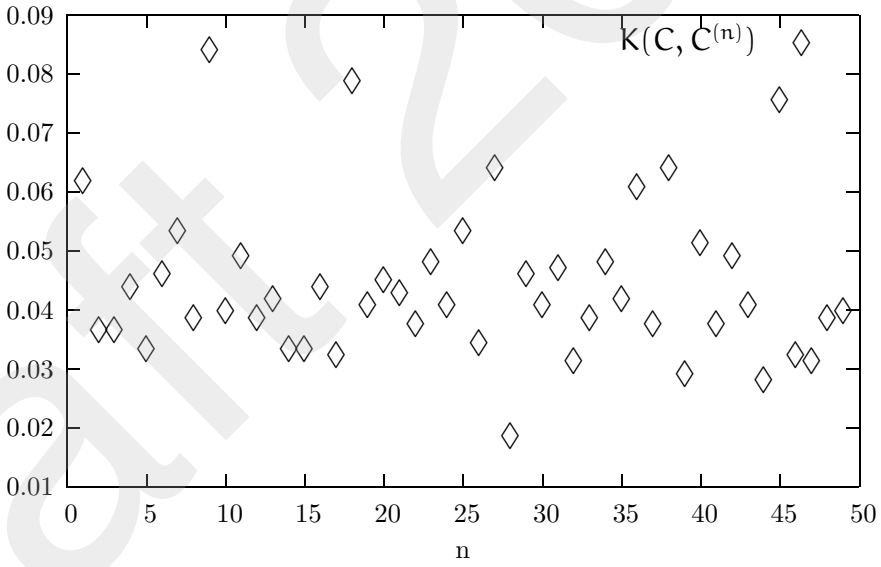


Figura 3.18: Valores de $K(C, C^{(n)})$ com $n = 1 \ldots 50$.

É fácil ver que para valores de n múltiplos de 9 o valor de $K(C, C^{(n)})$ é mais próximo de 7.77%. Isto indica que o período (e portanto o tamanho da chave) é 9.

Analisemos então os valores de $\Psi(C_i)$ para $i = 0 \ldots 8$, em que C_i é o texto que se obtém com os caracteres de ordem $9k + i$ de C.

Draft 2021.1

i	0	1	2	3	4	5	6	7	8
$\Psi(C_i)$	8.8%	9.1%	7.0%	9.4%	9.1%	6.7%	9.8%	8.5%	8.4%

Apesar das grandes variações, motivadas pelo tamanho relativamente pequeno do criptograma, todos os valores são compatíveis com uma cifra monoalfabética, e portanto confirmamos que o tamanho da chave é 9.

Podemos também tentar verificar se alguma das letras da chave são repetidas, calculando $\Psi(C_i + C_j)$. Por exemplo calculemos este valores com $j = 8$

i	0	1	2	3	4	5	6	7
$\Psi(C_i + C_8)$	5.5%	5.6%	5.7%	6.5%	6.3%	5.7%	**8.2%**	5.7%

Provavelmente a penúltima e última letra da chave são iguais...

Uma abordagem alternativa ao teste de Friedman.

Uma maneira alternativa de abordar o teste de Friedman é começar por definir o **índice de coincidência de um texto** como a probabilidade de duas letras escolhidas ao acaso, nesse texto, serem iguais. Este número pode ser estimado do seguinte modo: seja n o número de letras do texto (que se supõe ser suficientemente grande) e, para cada $\alpha \in \{A, B, C, ..., Z\}$, seja n_α o número de letras iguais a α. Designando por $\mathcal{I}$ o índice de coincidência do texto, tem-se então:

$$\begin{aligned}
\mathcal{I} &= \mathbf{P}(\text{“ duas letras escolhidas ao acaso serem iguais”}) \\
&= \sum_{\alpha=A}^{Z} \mathbf{P}(\text{“duas letras escolhidas ao acaso serem iguais a } \alpha\text{”}) \\
&= \sum_{\alpha=A}^{Z} \frac{\binom{n_\alpha}{2}}{\binom{n}{2}} = \sum_{\alpha=A}^{Z} \frac{n_\alpha^2(1-\frac{1}{n_\alpha})}{n^2(1-\frac{1}{n})} \approx \sum_{\alpha=A}^{Z} \left(\frac{n_\alpha}{n}\right)^2 = \sum_{\alpha=A}^{Z} f_\alpha^2,
\end{aligned}$$

onde f_α = frequência da letra α no texto.

Os índices de coincidência de um texto (suficientemente grande) dependem assim apenas das frequências das letras de cada língua, obtendo-se os valores da tabela 3.5 de tabelas análogas à tabela 3.2.3.

Para um texto completamente aleatório, em que a frequência de cada letra é aproximadamente igual a $\frac{1}{26}$, tem-se $\mathcal{I} = \frac{1}{26}$, e como

$$\sum_{\alpha=A}^{Z} f_\alpha^2 = \frac{1}{26} + \sum_{\alpha=A}^{Z} \left(f_\alpha - \frac{1}{26}\right)^2,$$

conclui-se que quanto “menos aleatório” for um texto, maior é o seu índice de coincidência. Em particular, dado um criptograma, quanto mais próximo de 0.078 for o seu índice de

Draft 2021.1

coincidência maior é a probabilidade de ser um texto em português que foi cifrado com uma cifra monoalfabética; quanto mais próximo de 0.038 for o seu índice de coincidência, maior a probabilidade de ter sido usada uma cifra polialfabética.

No caso de ter sido usada a cifra de Vigenère, o índice de coincidência pode ser usado para obter informação sobre o comprimento da chave, do modo que passamos a descrever. Seja ℓ o comprimento da chave, e suponhamos que a chave não contém letras repetidas (o caso geral fica como exercício...). Se imaginarmos o criptograma disposto em ℓ colunas,

$$\begin{array}{cccc} \alpha_1 & \alpha_2 & \cdots & \alpha_\ell \\ \alpha_{\ell+1} & \alpha_{\ell+2} & \cdots & \alpha_{2\ell} \\ \alpha_{2\ell+1} & \alpha_{2\ell+2} & \cdots & \alpha_{3\ell} \\ \vdots & \vdots & \cdots & \vdots, \end{array}$$

podemos calcular o seu índice de coincidência, $\mathcal{I}_C$, do seguinte modo. Como as colunas foram cifradas com uma cifra monoalfabética, nomeadamente uma cifra de César, o índice de coincidência em cada coluna é aproximadamente igual ao índice de coincidência da língua em que o texto foi escrito, e que designaremos por $\mathcal{I}_T$. Por seu lado, a probabilidade de duas letras que pertencem a colunas distintas serem iguais está mais perto de $\frac{1}{26}$. Como o número de pares de letras em que ambas estão na mesma coluna é aproximadamente $\ell\binom{\frac{n}{\ell}}{2} = \frac{n(n-\ell)}{2\ell}$, enquanto o número de pares de letras que estão em colunas distintas é aproximadamente $\binom{\ell}{2}\left(\frac{n}{\ell}\right)^2 = \frac{n(\ell-1)}{2\ell}$, resulta que o número esperado de pares de letras iguais é aproximadamente: $\mathcal{I}_T\frac{n(n-\ell)}{2\ell} + \frac{1}{26}\frac{n(\ell-1)}{2\ell}$. Assim,

$$\mathcal{I}_C \approx \frac{1}{\binom{n}{2}}\left(\mathcal{I}_T\frac{n(n-\ell)}{2\ell} + \frac{1}{26}\frac{n(\ell-1)}{2\ell}\right),$$

uma equação em que ℓ é a única incógnita. Resolvendo-a, obtém-se:

$$\ell \approx \frac{\left(\mathcal{I}_T - \frac{1}{26}\right)n}{\mathcal{I}_C(n-1) + \mathcal{I}_T - \frac{n}{26}} \approx \frac{\mathcal{I}_T - \frac{1}{26}}{\mathcal{I}_C - \frac{1}{26}}.$$

O índice de coincidência de um criptograma pode ser facilmente estimado "à mão" ...

Exemplo de ataque criptográfico à cifra de Vigenère.

Suponhamos que a seguinte mensagem foi interceptada:

```
IYXSB EBLAP IZGEA WBGHO QDNEC WBGHC IHFAQ SALTO RGXDO ZVWAH
EBVOB GEXTO IQXFW RVWAQ SZHOI XETCC MFTQI EYJUS VPHMC IFMAD
IQKAQ MASEB XNXME YRFEG IAMOS HRLCO RFHCC QBXSH IEBBS MEHMO
RFHEA WRKEB SFLOP VRLSO PGHSQ SZHEG XRLPW RUXIF SFTLH SFJUS
IZOEF HRXOW VBLEO KVMAA GBFOS WGTSO ZRLQI ITKIH EZXMP IOXDS
```


Draft 2021

```
METSR INSUZ IYXSB EBLAP IZJUS SFHNV SROIB LBXEG THFAS JRKMS
RGHBW GUBNV SNEAQ VRXSS HRGTC HRYOQ MAAOD SAMIO KHWOE YRYOG
WNTTF EIXSR IGNDC RHFPS VCXTI SZHVW QRGTC IYXSB EBLAP IZJUS
SFHNV SRMEZ ERVOF ICBNQ IYUAG ISNSH IPTPW XREAF GBXMC KVOAJ
MGKAZ TVGAQ YYHDS GNMER VNECC RGKAD SAMOG MAYOB MNFAG GNKAU
VRZAA ETBAE YRXRS XBKTO HRTLE YVFIG XNFAD EQHMI RQHDW WGTNH
IEHSO HBLVS RGHSW RSTNH IPTRO ZREAE YVGHS RGBSH EDNES GNUOR
EOHAS WCXRO RPTOI VBVAB IYTMO VSBMT PBKEH IQXEG TNWAQ LVFBO
WGBDC VCTSG SQXDO RPTCC PBFBW RNXAF PRJUW QCTSG EEHLO ZBTDC
VNIAF EETIC WYHCC QBMIJ EOTRQ SQXPF SNYEG XVOAO PGHFC VAHGS
VNWOF EPBSO SQHAH SZHRO HNKUZ XETSC QGXLS ZVLAC HRLEA FNKQI
IRFFC KHXTO SATSI TRKFW GVXLI RNKEZ IFGAC WNUEA RRFSC RUTME
YRHSC RUHCC QNGDO EIBDO UHXSS QCKEE YRNMV SZXMG SAAAC QHGDC
THEAS EITNQ EPHMC FBEAQ SYHRW HNXNH VRTSA EBLDS YZTCF MNGCO
```

e que sabemos (por espionagem, por exemplo...) que foi usada a cifra de Vigenère e que o texto original foi escrito em português. Com o uso de um modesto computador (ou com um lápis, algum papel e muita paciência...) é muito fácil descodificar a mensagem sem conhecer a chave. Em primeiro lugar, aplicando o teste de Babbage–Kasiski, obtém-se:

posições	padrão repetido	factorização
0 , 260	IYXSBEBLAPIZ	$260 - 0 = 2^2 \times 5 \times 13$
15 , 25	WBGH	$25 - 15 = 2 \times 5$
69 , 174	QSZH	$174 - 69 = 3 \times 5 \times 7$
91 , 911	PHMC	$911 - 91 = 2^2 \times 5 \times 41$
113 , 848	MEYR	$848 - 113 = 3 \times 5 \times 7^2$
129 , 149	ORFH	$149 - 129 = 2^2 \times 5$
132 , 712	HCCQB	$712 - 132 = 2^2 \times 5 \times 29$
169 , 739	OPGH	$739 - 169 = 2 \times 3 \times 5 \times 19$
260 , 385	IYXSBEBLAPIZJUSSFHNVSR	$385 - 260 = 5^3$
287 , 637	XEGT	$637 - 287 = 2 \times 5^2 \times 7$
299 , 559	SRGH	$559 - 299 = 2^2 \times 5 \times 13$
321 , 381	RGTC	$381 - 321 = 2^2 \times 3 \times 5$
334 , 479	DSAM	$479 - 334 = 5 \times 29$
429 , 569	HIPT	$569 - 429 = 2^2 \times 5 \times 7$
547 , 567	TNHI	$567 - 547 = 2^2 \times 5$
609 , 664	ORPT	$664 - 609 = 5 \times 11$
656 , 686	CTSG	$686 - 656 = 2 \times 3 \times 5$
712 , 857	HCCQ	$857 - 712 = 5 \times 29$
843 , 853	SCRU	$853 - 843 = 2 \times 5$

Draft 2021.1

indexcifra! de Vigenère

O máximo divisor comum das distâncias entre textos repetidos com pelo menos quatro letras é 5, e este é muito provavelmente o comprimento da chave usada (em geral, pode acontecer que o máximo divisor comum seja 1, mas se tomarmos o máximo divisor comum do maior número possível de distâncias apanha-se o tamanho de chave, se o texto for suficientemente longo, claro...). Dividindo agora o texto em subtextos que contêm as letras cujas posições correspondem às classes módulo 5, obtêm-se textos que foram cifrados com cifras de César. Uma análise às letras mais frequentes de cada um desses subtextos revela rapidamente a chave que era afinal menos secreta do que à partida se poderia julgar!

3.4.3 Vigenère Permutada

Uma forma de tornar a cifra de Vigenère mais segura é a de partir de uma tabela como a da Vigenère clássica (tabela 3.4) mas em que a primeira linha não é constituída por um alfabeto na sua ordem natural, mas sim por um alfabeto permutado. Isto corresponde à aplicação de uma cifra monoalfabética seguida de uma cifra de Vigenère. A chave da cifra passa a ser, não só a chave da cifra Vigenère como também a cifra monoalfabética que constitui a primeira linha da tabela de Vigenère que se vai usar.

Exemplo de criptanálise de uma Vigenère Permutada Suponhamos que temos o seguinte criptograma, que sabemos que resultou de uma Vigenère Permutada a um texto em português:

```
AARYL LVWVP KVYKN TKWBT KHRPV SYYME JUMPY AWBVI SKQAS YNYXF
MJZLO NEPFT TUTDZ EYAJB BVCKE DYLMP HIRCO NRIVV YPRUB ANWRM
MBXAW AKYBQ ZVMTL PNVXP TGPID OSWTT GVMKS MTPRT JHAHD MVHOE
IQAYG YWUMQ WVNKB ENUMT IGMKE ZNHZC HIAYO MWBRC ZGWDJ YWLXF
AXJLN STWVK AXRFF YFWGQ KCPDY GYHXK BAFAS YSPBC ZGMYD TNYXP
GKNUG WNAAI YBALN STWOM ALWKE AWLXP ZGBLE WOAVX WUNLV WOPFV
AIABY WUTRT DXRHW KHCOV AZYFL ZNRXN YQYCF ARPJK NYRIL OHMJC
NIRIY SNVWT ZKIDZ EYBXP IGVHY STTGM MJZAS YTWGZ ZCHPZ GKPMF
JGMDW AHHZZ GVKUD GOCCC SURKX KYWPZ HUVKW UUMVV MIADZ AEBPT
OIFPS KYYOT NAXVX KYHON YWRLO MLERT GILJZ MFBCQ ALMLG YIDXK
OGYKL DLIVU BUFOV MEWSC OOFVW SYHVN MXAID WYCXU EGILH WONON
EVCLO YTUVV DXPLJ YRUOE YJIAU DYLBT ZKCLL CVVMY MLBKI YWLVX
AUQLB POVYB AIZLE BYMTV NAXIY YRIYV MBAKE KYHAI YJLCO AUDMB
AIRKN YTIVU MARIY YTWCK YLACQ AWLVI KYPKZ ZNYXE MBAOV WYHGR
OCYYL SYLVX MILVL HKUXQ PBAYD ZKHZZ PYLDL KIWZK ALACW CHCMK
MLCUG AEPZT JKQOY OYAVP MIRVL JJWZC IGPVL PNIMK ABAYL BKUXV
```

Draft 2021

```
MAXVQ YKPGV YVPKZ EYGMU ALGZL ZNHPQ OPXYQ NXBGM BUFZE YOHMF
MCPDZ ZOHMV YWXCN ZNYYR KCQKD KLBFP YJZAE BVIKT WUQDZ BZMFF
NKCLN OYAZZ ZVCLJ NRTZT WVXMY KKLMZ ABADZ OFPTT WQFKW AWEFF
NVIUE ARWPI CUCAH YIPLK ASYIL ZNITC WUVKB ZOWPZ HCNPO NOWCC
JVJKI ZOYRC PWAIY BYTBK MGXDZ SNYOQ WBXML OYYJF BKROE AEPBT
JHAHL KHMOR KAXVL OVPPZ OPXYV BXIJQ APRFL BXMXV AKMHD GFBGT
IVYCD JYLVP ALJKE YEWGF MCAZQ ARLMA WGMMY OOAPC QCRPS YRCXZ
ASRVL OTBGT OBRYR WKAXU ZVMLL OQPKC OBXYO DSTPZ PHXYG ARCVN
OGQSY WUBRQ OLAUN EHPJR NWXYS YTTKQ IIAFL ZOYVE AJXCD GNHGZ
SVWKZ AREXL NSACS AOHXK TKPLS MRWXP BVFOX ARLYE QVPIV KEWTI
FKAAE ALPTG YWABY OREJE BVRBD PYHXQ HOQIY ZVCZC WIRPR BOLMM
NEXFY DLGVR WGKUV PFPTZ PALFQ MYMXN ZKQKS MHPJP YIWKN BTPHT
OAXVX KUTCK YLILX KUWPK MONKO ARIXM ALYCF ARWKT DXIKZ LVWMN
SOESL ZOWYE AGMHL SVPPC KCRPE KOYMQ PBAVY PRWFG MIPLO MWDMF
YOEAR MCMOR POMMV SJWBR QCRLZ YXYMF YJCAL DLIXN SKQKB YWUXG
DXYIL ZNLXP WVNKR YWHGZ JSAYQ WNH
```

Qualquer uma das técnicas descritas atrás pode ser utilizada para determinar o período da cifra usada. Comparemos, por exemplo, os valores obtidos para o indíce de coincidência por coluna, dividindo o texto em colunas e usando um número de colunas correspondente às sucessivas possibilidades do tamanho da chave. Usemos para isso valores entre 4 e 12. Os valores obtidos para as colunas são os que constam da tabela seguinte.

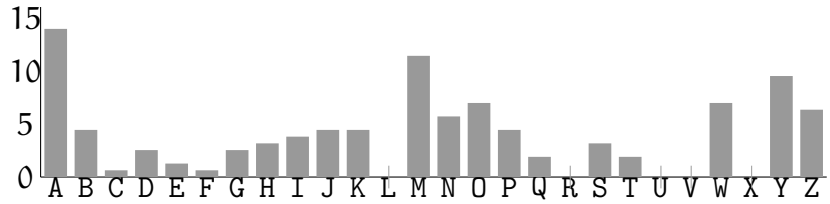
Período												
4	49	49	47	49								
5	72	52	60	56	50							
6	49	48	49	49	50	52						
7	47	46	44	46	44	50	48					
8	50	49	48	50	51	52	50	50				
9	48	47	51	43	47	49	50	45	52			
10	71	69	81	82	76	79	79	84	82	70		
11	50	56	47	45	54	50	48	48	49	47	44	
12	48	54	49	51	51	54	55	54	56	54	56	56

Os valores correspondentes à divisão do texto em 10 colunas são os mais altos o que aponta para que o período seja esse.

Estamos em condições de apontar, para cada uma das colunas, quais os prováveis valores cifrados das três mais frequentes letras na lingua portuguesa. Para tal basta analisar as estatísticas das ocorrências por caracter para cada coluna.

Para a primeira coluna, o gráfico das frequências por caracter tem o seguinte aspecto:

Draft 2021.1



Os valores mais altos, e portanto os mais prováveis de corresponderem a "a", "e" e "o", são "A", "M" e "Y". Porque o texto que consta da coluna, e aliás o texto que compõe a mensagem original, não contém os caracteres **exactamente** de acordo com a frequência média da sua ocorrência no português, é natural que a atribuição dos valores correspondentes a estes três caracteres seja passível de erro. Mas como veremos, e se automatizar-mos o processo que descrevemos a seguir, fazendo todas as tentativas de atribuição destas três letras, podemos eliminar quase todas as atribuições erróneas. Neste caso teríamos que testar (no máximo) 6^{10} atribuições, o que é possível fazer em tempo razoável.

Para cada coluna a atribuição correcta destes valores é a que se segue.

	0	1	2	3	4	5	6	7	8	9
a	A	V	R	K	L	A	Y	P	X	T
e	Y	G	A	L	D	Y	N	W	M	C
o	M	K	X	A	Y	M	O	B	V	Q

A primeira observação, é que a primeira e a sexta coluna partilham a mesma cifra monoalfabética, ou seja o primeiro e o sexto caracter da chave são iguais. Por outro lado, sabemos que todos as cifras monoalfabéticas correspondentes às colunas, correspondem a rotações sucessivas da primeira cifra, pelo que são todas rotações umas das outras. Isto implica que o número de caracteres que se encontram entre os caracteres correspondentes a "a" e "e" para a primeira coluna, vai ser o mesmo para todas as outras cifras da tabela.

Olhemos para a informação que dispomos para as colunas números 0 e 3.

```
   abcdefghijklmnopqrstuvwxyz
0  A   Y         M
3  K   L         A
```

O caracter correspondente a "o" na coluna 3 é o mesmo que corresponde a "a" na coluna 0. Como as duas cifras são rotações uma da outra, passamos a saber colocar "K" e "L" na cifra correspondente à coluna 0 assim como "Y" e "M" na da coluna 3. O resultado é:

```
   abcdefghijklmnopqrstuvwxyz
0  A   Y       K M L
3  K M L         A   Y
```

Draft 2021.1

Passamos a saber traduzir, na coluna 0, as letras "m" e "q". E essa tradução é compatível com a distribuição estatística de caracteres da coluna 0. Se tal não fosse, por exemplo se o caracter "K" ocorresse muito poucas vezes, isso significava que a atribuição que fizemos dos três primeiros caracteres estava errada, e deveríamos tentar outra. Se prosseguirmos este processo para as outra colunas obtemos a seguinte informação.

```
   abcdefghijklmnopqrstuvwxyz
0  A G Y   N X K M L O D R V
1  V A G Y   N X K M L O D R
2  R V A G Y   N X K M L O D
3  K M L O D R V A G Y   N X
4  L O D R V A G Y   N X K M
5  A G Y   N X K M L O D R V
6  Y   N X K M L O D R V A G
7  P   W         B
8  X K M L O D R V A G Y   N
9  T   C         Q
```

Podemos usar esta informação, para parcialmente traduzir o texto e tentar encontrar outras letras da cifra monoalfabética.

```
acasaqueoPmaiasTieBamHaPiSaYeEJUsPoaWooISoQoSeeYaFoJZeciEaF
aTUTiZEaAJBBaCaEuaLePHIaCcisIoVePaUBaeemMoBooWaiYBoZasTaPeV
aPTePIesSeTacasaSoTamaJHeHeouHiEIQesmeWUeoWamaBEeUeaIesaEZe
HZeHIescoWomeZeWiJeWLaFamJesSTeoKamaFFeFesomCPiocaHaKBcFoSe
SaBeZesseTeYaPcomUmWeAqIeBeesSTeiMasWaEaWLaPZeBeEWoAoXWUmei
WoaFVaIeBoWUTmaumaHWmHCiVaZiFaZeRaNeQiCFasaJKigaIasHMJeiIaI
oSeVWaZoIiZEaoaPIecHoSTTsMoJZoSeTesZZCHPZciaeFJesiWaHHZZcaq
UecoCCeSUaaumaePZHUcaWUUMoVoIeiZaEoPasIFPSmaYiaicomumaHiNeW
aecomEmacIuJZoFoCoassemeIDaKseiaaumIoUBUFgioEeSesuFmWSaHoNo
meIeWaCaUEeIeHWoNiNEaCeceTUoVumPeJesUiEeJIoUuaLBaZoCeaCuVeY
osBaIeWLoXaUQeBPoVuBaIZeEBaMTVicoIoesIuVoBeaEmaHqIeJuCcaUDe
BaIaaseTIoUocaIoeTeCKeseCQaWLoImgPaZZeYaEoBegiWaHsRsCisaSaL
oXoIumaHiUaoPBeseZiHZZPguiamIeZKaseCWCHCeKosCUmaEaZaJoQgosa
AoPoIamaJJeZeIePmaPeIeKaBesaBiUaVocomQeiasVeaPaZEaGeUasgZaZ
eHPosPosQigosMBUFZEeoHeFoCPiZZoHeVeWoCsZeYuRmCQaemmoFPeJZoE
BuIcaWUQiZBZMFFioCessaAZZZaCeJisTZaWaocomiLeZaBeiZsFaTaWQFa
WaWEFFiaIUEasePICUCoHeIagKaSiIaZeITeWUcaBZoePZHCmPcioeCeJaJ
aIZoYmePWeIoBaTBKoeoiZSeYioWBocasaYJFBoagEaEaBaJHeHamHMiRmc
omasuaPZsPosiBgIJoaPaFaBgMaVaosHecFosaIaiCeJaLoPasJaEeEesFo
```


Draft 2021.1

```
CeZQasLeAWescosoAPeQCaPSesCaZaSamasTosasBasgWiAaUZaseasQace
sBoscuSTPZPHosmasCoNseQSoWUomosseUsEHaJRiWosSeTTcoIIeFaZoYo
EaJoCeceHsZSaWaZasEaLiSeCSaoHaKToPeSoseaPBaFguasLuEQaPIimEe
TIFoeoEamaTGeWeBossEJEBaaBePaHaoHuQIoZuCZeWIaPgBoLeMiEoFoum
GoRWeqUiPFaTZPcuFQoaMaNZoQaSoHaJPeIWasBTaHascomumUTCKesIeum
UePKoumacasIaMasiCFasecaumIaZqueeNSuESaZoeuEaesHaSuaPemCaPE
moYeoPBemoPseFGoIPecoWDeFeuEogoCMiRPusciSJeBRQCaeZegYeFeJCo
aumIaNSoQaBeWUaGumiIaZeLaPWamageWHsZJSesQWeH
```

A decifração é bastante fácil de prosseguir a partir daqui e o texto que encontraríamos seria então:

"A casa que os Maias vieram habitar em Lisboa, no outono de 1875, era conhecida na vizinhança da rua de S. Francisco de Paula, e em todo o bairro das Janelas Verdes, pela casa do Ramalhete ou simplesmente o Ramalhete. Apesar deste fresco nome de vivenda campestre, o Ramalhete, sombrio casarão de paredes severas, com um renque de estreitas varandas de ferro no primeiro andar, e por cima uma tímida fila de janelinhas abrigadas à beira do telhado, tinha o aspecto tristonho de Residência Eclesiástica que competia a uma edificação do reinado da sr.ª D. Maria I: com uma sineta e com uma cruz no topo assemelhar-se-ia a um Colégio de Jesuítas. O nome de Ramalhete provinha de certo dum revestimento quadrado de azulejos fazendo painel no lugar heráldico do Escudo de Armas, que nunca chegara a ser colocado, e representando um grande ramo de girassóis atado por uma fita onde se distinguiam letras e números duma data. Longos anos o Ramalhete permanecera desabitado, com teias de aranha pelas grades dos postigos térreos, e cobrindo-se de tons de ruína. Em 1858 Monsenhor Buccarini, Núncio de S. Santidade, visitara-o com ideia de instalar lá a Nunciatura, seduzido pela gravidade clerical do edifício e pela paz dormente do bairro: e o interior do casarão agradara-lhe também, com a sua disposição apalaçada, os tectos apainelados, as paredes cobertas de frescos onde já desmaiavam as rosas das grinaldas e as faces dos Cupidinhos. Mas Monsenhor, com os seus hábitos de rico prelado romano, necessitava na sua vivenda os arvoredos e as águas dum jardim de luxo: e o Ramalhete possuía apenas, ao fundo dum terraço de tijolo, um pobre quintal inculto, abandonado ás ervas bravas, com um cipreste, um cedro, uma cascatasinha seca, um tanque entulhado, e uma estatua de mármore (onde Monsenhor reconheceu logo Vénus Citherêa) enegrecendo a um canto na lenta humidade das ramagens silvestres."

3.4.4 Vigenère Autokey

Uma variante da cifra de Vigenère permite evitar o tipo de ataques até agora apresentados para a cifra de Vigenère . A cifra é inteiramente igual, mas em vez de ser repetidamente

Draft 2021

usada a chave como guia de utilização da tabela de Vigenère, é usada a chave e seguidamente a própria mensagem [?].

Usando o mesmo exemplo da página 83 a mensagem

Fujam enquanto é tempo! Eles vêm aí!

cifra-se com usando a chave "chumbo", começando por escrever a chave sobre as primeiras letras da mensagem, e depois usando o próprio texto da mensagem como chave:

```
chumbofujamenquantoetempoele
fujamenquantoetempoelesvemai
```

O resto do processo é inteiramente igual ao descrito em 3.4.1, resultando no texto cifrado:

```
HBDMNSSKDAZXBUNEZICIEIEKSQLM
```

3.5 Cifras poligráficas

Uma solução para fugir aos ataques baseados na análise de frequências é a que foi primeiramente usada por Giovanni Porta em 1563. O seu método, em vez de cifrar uma letra de cada vez (uma cifra monográfica) passava fazer corresponder um símbolo a cada par de letras constituindo assim a primeira cifra poligráfica (digráfica de facto!)

3.5.1 A cifra de Playfair

Mas a cifra poligráfica mais conhecida á a chamada **cifra de Playfair**. A chave consiste num rectângulo de caracteres (normalmente 5×4 no caso das línguas latinas e em que se toma $\mathtt{u} \equiv \mathtt{v}$ e $\mathtt{i} \equiv \mathtt{j}$, ou 5×4 nas línguas anglosaxónicas e germânicas em que se toma $\mathtt{i} \equiv \mathtt{j}$).

R	S	N	Z	Q
T	A	P	E	G
X	M	V	F	W
I	K	D	U	O
C	L	B	Y	H

Para se cifrar uma mensagem procede-se da seguinte forma:

1. Primeiro divide-se a mensagem em pares, com o cuidado de nenhum par ter letras iguais (introduzindo um x se necessário) e acrescentando um x se necessário para completar o último par.

Draft 2022.1

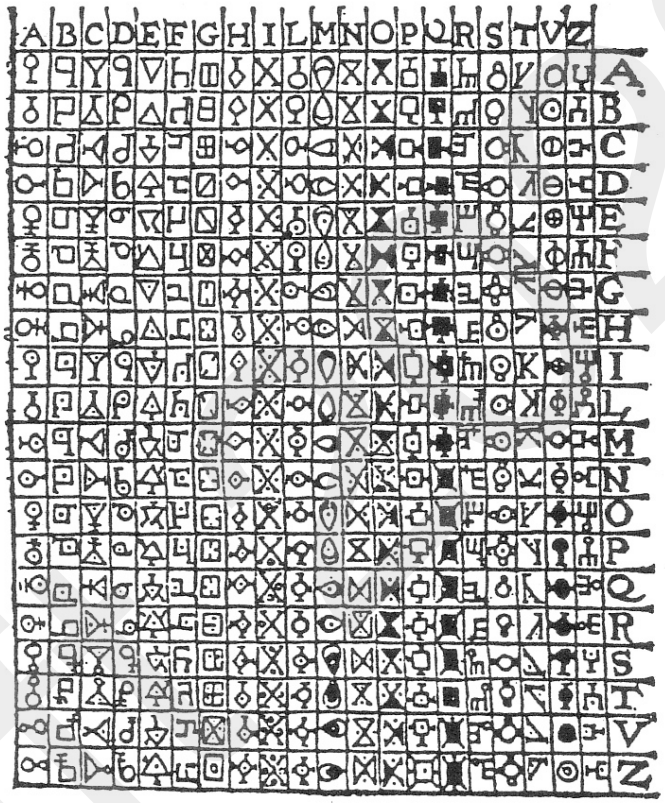


Figura 3.19: A cifra digráfica de Giovanni Porta

Draft 2.1

ag ue rx ra so te mv en ci do sx

2. Se as letras que compõem o par se encontram na mesma linha, são substituídas por aquelas que se encontram imediatamente à sua direita (podendo se necessário passar para o outro lado da tabela como se esta fosse um toro):

   Assim para ag, que se encontram na mesma linha da tabela,

   T A P . G

   resulta PT.

3. Se a letras se encontram na mesma coluna, toma-se as que se encontram imediatamente abaixo.

   Para ue temos

   .
   E
   F
   U
   Y

   pelo que o resultado é YF.

4. Se as letras se encontram em linhas e colunas diferentes, tomam-se as que na mesma linha pertencem à coluna da outra.

   Para so vem

   S . . Q
   . . . .
   . . . .
   K . . O

   logo o resultado é QK.

   O resultado da cifra seria então:

   PT YF TI ST QK AG VF PZ RC UI RM.

Como se pode ver, o primeiro a é cifrado num P e o segundo por um T, as duas ocorrências de r por um T e depois por um S. Uma simples análise de frequências como foi feita para as cifras monoalfabéticas não resulta para esta cifra.

Draft 2022.1

Exemplo de um ataque a uma cifra Playfair

Suponhamos que interceptamos o seguinte criptograma que sabemos cifrado por uma cifra de Playfair:

```
GC PE HB GJ NH EN CI TO QN AS MO BF TP HN AI NQ
BF IB ET OH TD ID AB JN HY BY PC KF QH KF MN EB
IB IB NK TJ IB SA BF IB CI TO IP OT EN DI QD JN
TQ FP FR BK AB CI TO QN DI G3 FK EN KT TO FK CK
OJ QO Q3 KF SO HN EF
```

Mas para além do texto do criptograma sabemos também algo sobre o seu conteúdo[15]:

- Uma das seguintes frases deve fazer parte do seu texto: "`resultado inesperado`" ou "`hipotese corroborada`".
- As frases do texto estão separadas por "`stop`" e a mensagem está terminada por um "`end`".

Três observações iniciais sobre a cifra de Playfair:

1. No texto cifrado nunca ocorre um par com as duas letras iguais (como aliás não pode ocorrer no texto original).
2. Nenhuma letra se tem a si própria como imagem cifrada.
3. Dois pares da forma `xy` e `yx` têm como imagens pares `ZW` e `WZ`.

Consideremos primeiro a hipótese de "`hipotese corroborada`" estar presente na mensagem. Se esta for a primeira frase da mensagem, então será finalizada obrigatoriamente com um `stop`, pelo que a sua divisão em pares de letras seria:

```
hi po te se co r@ ro bo ra da st op
```

onde "`@`" denota o carácter nulo cuja escolha, apesar de desconhecida, não é relevante para o que se segue. Observemos que o segundo par é "`po`", enquanto o décimo segundo é o seu inverso, "`op`", padrão que não ocorre no criptograma. Na hipótese de ser a última frase, esta está obrigatoriamente intercalada entre um "`stop`" e um "`end`", pelo que ocorre um dos dois casos seguintes:

[15] Apesar de este exemplo ser obviamente fantasiado, a suposição que conhecemos uma frase que tem que necessariamente se encontrar numa mensagem não é tão artificial assim. De facto os *cribs* (frases que sabemos que obrigatoriamente fazem parte do texto cifrado) são muito frequentes quando o alvo do ataque é um canal de comunicação há muito estudado. A maioria das mensagens cumprem protocolos, que têm frases fixas, as mensagens podem ser produto de um programa que tenha cabeçalhos fixos, etc... A História da criptografia está repleta de ataques que nunca seriam possíveis sem o conhecimento de *cribs*.

Draft 2021.1

```
st OP PO hi te se co r@ ro bo ra da en d@
?s to ph ip ot es ec OR RO bo ra da en d@
```

Ambas as hipóteses podem ser eliminadas porque nenhum dos padrões observados nos pares $(-13,-11)$ ou $(-7,-6)$, acima destacados, se verifica no criptograma. Resta pois verificar se o *crib* ocorre no meio do texto, ou seja, intercalado entre dois "stop", em cujo caso se teria uma das duas hipóteses:

```
ST OP hi PO te se co r@ ro bo ra da ST OP
?s TO ph ip OT es ec OR RO bo ra da st op
```

É fácil ver que os padrões assinalados não ocorrem no critpograma original. Conclui-se assim que a frase "resultado inesperado" tem que fazer parte da mensagem.

Pode-se proceder a uma análise inteiramente análoga à anterior sobre as cinco hipóteses seguintes, correspondentes às distintas formas com que a frase se pode apresentar no texto:

```
RE su lt ad oi ne sp ER ad os to p?
?s to pr ES ul ta DO in ES pe ra DO en d@
st op RE su lt AD oi ne sp ER AD oe nd
?s to pr ES ul ta DO in ES pe ra DO st op
st op RE su lt AD oi ne sp ER AD os to p?
```

Facilmente se verifica que a última hipótese é a única compatível com o criptograma, nos pares 7–20, pelo que ficamos a saber a seguinte correspondência entre o texto original e o texto cifrado:

```
CI TO QN AS MO BF TP HN AI NQ BF IB ET OH
st op re su lt ad oi ne sp er ad os to p?
```

Usando esta informação podemos começar a tentar reconstruir a chave (tabela) utilizada.

Comece-se por observar que o tamanho da tabela parece ser 4×7 ou 5×6, uma vez que ocorrem quase todos os caracteres do alfabeto (exactamente 21) e somente um algarismo, o 3. Como "op" é cifrado em "TO", os caracteres "o" e "p" têm de ocorrer na mesma linha ou coluna na chave, pois um par de caracteres que não esteja nestas condições é cifrado num par disjunto do primeiro. Como "o" é cifrado em "T" e "p" em "O", conclui-se que a chave contém "POT" numa "linha", seja ela uma linha ou uma coluna da tabela (considerada como inscrita num toro). Da mesma forma se conclui que se tem "ENH" e "OTE", assim como "USA". Daqui resulta que "POTENH" faz parte de uma "linha". Dos pares "TP", "oi" pode-se de seguida concluir que, dado que "T" e "P" pertencem a essa "linha", o "o" e o "i" também a ela pertencem, ocorrendo pois "IPOTENH" numa "linha". Fica assim eliminada a hipótese de a tabela ter dimensões 5×6, tendo obtido aquilo que muito provavelmente é uma "linha" inteira, "IPOTENH", assim como um pequeno bocado de uma outra, "USA".

Draft 2021.1

Do conhecimento da linha "IPOTENH" pode-se concluir que "OH" se decifra por "pn", o que determina o caracter desconhecido acima designado por "?".

Dos pares "QN", "re", uma vez que "N" e "e" se encontram na mesma "linha", conclui-se que "Q" e "r" devem pertencer a uma outra "linha", sendo as quatro letras vértices de um rectângulo. Isto permite concluir que "QN $\leftrightarrow$ RE" e "NQ $\leftrightarrow$ ER". Mais ainda, "Q" e "R" encontram-se lado a lado (num sentido tórico, relembre-se) na tabela que constitui a chave.

Usando os resultados obtidos, pode-se decifrar parcialmente o criptograma, obtendo:

```
GC it HB GJ en te st op re su lt ad oi ne sp er
ad os to pn TD ID AB JN HY BY PC KF QH KF MN EB
os os NK TJ os us ad os st op hi po te se QD JN
TQ FP FR BK AB st op re se G3 FK te KT op FK CK
OJ QO Q3 KF SO ne EF
```

Poder-se-ia agora suspeitar que o ínicio deve ser muito urgente e daí retirar mais informações sobre a chave.

Porém, usando o que acima se descreveu, e reparando ainda que do *crib* resultam as correspondências "sp $\rightarrow$ AI", "os $\rightarrow$ IB" e "st $\rightarrow$ CI" que mostram que "SABC" se encontra "debaixo" de "IPOT", tem-se que, a menos de uma transposição e de permutações circulares de linhas e colunas, a tabela que constitui a chave tem o aspecto seguinte:

```
I  P  O  T  E  N  H
.  .  .  .  .  .  .
S  A  B  C  .  .  U
.  .  .  .  .  .  .
.  .  .  .  Q  R  .
.  .  .  .  .  .  .
```

Note-se que, em princípio, a tabela tem apenas 4 linhas. As linhas constituídas por pontos significam apenas a eventual existência de uma linha; por outro lado, pelo que neste momento sabemos,"QR" poderá fazer parte de linha "SABC..U".

Há apenas mais duas correspondências obtidas a partir do crib que dão informação adicional: "ad $\rightarrow$ BF" e "lt $\rightarrow$ MO", que implicam que a chave é algo como:

```
I  P  O  T  E  N  H
.  .  .  .  .  .  .
S  A  B  C  D  E  U
.  .  .  .  .  .  .
.  .  .  .  Q  R  .
.  .  .  .  .  .  .
.  .  L  M  .  .  .
.  .  .  .  .  .  .
```


Draft 2021.1

O facto de aparecerem as letras "ABCDE" seguidas pode fazer suspeitar o uso de uma Playfair com "menomónica" que a simples mudança da última linha para primeira revela ser: "HIPOTENUSA".

Assim obtem-se a chave (em que o símbolo "?" não é possível determinar, mas que é irrelevante):

```
H I P O T E N
U S A B C D F
G J K L M Q R
V W X Y Z ? 3
```

com a qual se recupera imediatamente a mensagem original:

```
muito urgente stop
resultado inesperado stop
necessario voltar a gerar todos os primos usados stop
hipotese de riemann falsa stop
reservar tempo para miller rabin end
```

3.5.2 Outras cifras

Tabela de Políbio

Uma cifra muito simples criada pelo historiador Políbio (sec.III ac) consiste na substituição de cada letra do texto original pelas suas coordenadas numa tabela como a 3.6. A robustez criptográfica não é grande porque se soubermos reconstruir os pares de coordenadas que correspondem a cada letra, a cifra sofre de todas as debilidades das cifras monoalfabéticas (ver 3.2.3). Não se conhece nenhuma utilização histórica, apesar de ser bastante mais interessante que a cifra que Júlio César viria a usar nas suas campanhas (ver 3.2). A sua importância prende-se com as várias cifras que usam este princípio de codificação.

	1	2	3	4	5
1	W	O	P	D	V
2	X	A	H	I	N
3	G	E	L	C	U
4	Z	R	B	S	M
5	F	G	K	T	Y

Tabela 3.6: Tabela de Políbio 5×5

Draft 2022.1

A Bifid e a Trifid

Cifras que usam o princípio da cifra de Políbio, atribuídas a Delastelle (1840–1902) [Kah67] baseiam-se num processo chamado seriação (também chamado substituição fraccionada) e pode ser usado com alguma variações. Seguidamente apresentam-se duas variantes conhecidas deste método. A primeira usa a tabela de Políbio, que aqui vamos tomar como exemplo a tabela 3.6

A cifra Bifid Suponhamos que queremos cifrar a mensagem

```
nao misturar morcela com atum frio
```

Comecemos por escrever a mensagem em grupos de letras, por exemplo em pentagramas:

```
naomi stura rmorc elaco matum frio
```

A cada uma das letras da mensagem podemos associar um par de inteiros (entre 1 e 5) que corresponde à linha e coluna da ocorrência da letra na tabela 3.6. Por exemplo a `N` corresponde o par $(2, 5)$. Escrevemos então as coordenadas de cada letra dispondo-as na vertical:

```
naomi stura rmorc elaco matum frio
22142 45342 44143 33231 42534 5421
52254 44522 25224 23242 52455 1242
```

Seguidamente transformamos cada um dos grupos de coordenadas de cada pentagramas, concatenando primeiro as suas dua linhas de coordenadas. O resultado é a seguinte sequência:

```
22 14 25 22 54 45 34 24 45 22 44 14 32 52 24 33 23 12 32 42
42 53 45 24 55 54 21 12 42
```

que agora, podemos voltar a "recodificar" com o recurso à mesma tabela 3.6 obtendo:

```
22  14  25  22  54  45  34  24  45  22  44  14  32  52  24
 A   D   N   A   T   M   C   I   M   A   S   D   E   G   I

33  23  12  32  42  42  53  45  24  55  54  21  12  42
 L   H   O   E   R   R   K   M   I   Y   T   X   O   R
```

A cifra resultante é portanto:

```
ADNAT MCIMA SDEGI LHOER RKMIY TXOR
```

Apesar de bastante simples, esta cifra, cuja chave é constituída pela tabela e pelo comprimento dos grupos em que se vão transpor coordenadas, é bastante difícil de quebrar [MR07] e por isso, juntamente com a "Trifid" usada frequentemente pelos "amadores" de criptografia.

Draft 2022.1

Bifid de matriz conjugada A "Bifid" pode ser ligeiramente modificada para apresentar uma ainda maior robustez se forem usadas duas tabelas de Políbio em vez de uma só [ACA16]. A cifra é inteiramente igual à Bifid original, mas os dois momentos de codificação usam tabelas separadas. Para além da vantagem de duplicar o tamanho da chave, a complexidade a confusão (ver 6.0.3) obtida é aparentemente maior. Tomemos o exemplo usado para a Bifid com as tabelas de Políbio presentes na tabela 3.7 como chave.

	1	2	3	4	5
1	W	O	P	D	V
2	X	A	H	I	N
3	G	E	L	C	U
4	Z	R	B	S	M
5	F	G	K	T	Y

	1	2	3	4	5
1	Z	G	R	K	S
2	P	Q	A	H	L
3	I	O	E	T	Z
4	D	F	V	U	C
5	W	N	M	B	X

Tabela 3.7: Duas tabelas de Políbio para serem usadas como chave na Bifid de matriz conjugada

Como a primeira tabela é igual à usada com a Bifid, temos a mensagem codificada com a primeira tabela:

```
naomi stura rmorc elaco matum frio
22142 45342 44143 33231 42534 5421
52254 44522 25224 23242 52455 1242
```

Ou seja, reagrupando os códigos obtidos:

```
22 14 25 22 54 45 34 24 45 22 44 14 32 52 24 33 23 12 32 42
42 53 45 24 55 54 21 12 42
```

Agora, "recodificamos" a mensagem usando a segunda tabela, obtendo:

```
22  14  25  22  54  45  34  24  45  22  44  14  32  52  24
Q   K   L   Q   B   C   T   H   C   Q   U   T   O   N   H

33  23  12  32  42  42  53  45  24  55  54  21  12  42
E   A   G   O   F   F   M   C   H   X   B   P   G   F
```

A cifra resultante é portanto:

```
QKLQB CTHCQ UTONH EAGOF FMCHX BPGF
```

O processo de decifração é inteiramente igual ao da Bifid mas agora usando primeirio a segunda tabela e depois a primeira.

Draft 2021.1

Trifid Aqui vamos representar cada letra por uma sequência de três algarismos, usando a tabela:

W:111	M:121	Z:131	N:211	O:221	L:231	C:311	T:321	U:331
A:112	&:122	Y:132	E:212	V:222	P:232	X:312	J:322	G:332
K:113	B:123	H:133	Q:213	R:223	S:233	I:313	F:323	D:333

A mensagem usada na Bifid vem assim, agora codificada com este novo código, mas continuando a dispor os símbolos na vertical:

```
naomi stura rmorc elaco matum frio
21213 23321 21223 22132 11331 3232
11221 32321 22221 13112 21232 2212
12113 31132 31131 21211 12111 3331
```

Agora, e como foi feito para o Bifid transpomos os códigos obtidos concatenando cada uma das linhas de cada grupo. O resultado é a sequência

```
212 131 122 112 113 233 213 232 131 132 212 232 222 131
131 221 321 311 221 211 113 312 123 212 111 323 222 123
331
```

O que podemos voltar a traduzir usando o código anterior

```
212 131 122 112 113 233 213 232 131 132 212 232 222 131
 E   Z   &   A   K   S   Q   P   Z   Y   E   P   V   Z

131 221 321 311 221 211 113 312 123 212 111 323 222 123
 Z   O   T   C   O   N   K   X   B   E   W   F   V   B

331
 U
```

Ou seja o criptograma:

```
EZ&AK SQPZY EPVZZ OTCON KXBEW FVBU
```

Exercício 3.4 *Qual a texto cifrado com Bifid e com a chave*

E	*X*	*T*	*R*	*A*
K	*L*	*M*	*P*	*O*
H	*W*	*Z*	*Q*	*D*
G	*V*	*U*	*S*	*I*
F	*C*	*B*	*Y*	*N*

cujo cifra é:

```
KGMTF IFLWG DKYBS IETYO ILKAO DGN
```

De notar que nada obriga que os grupos de letras tenham tamanho 5.

Draft 2022.1

Substituição nilista

Assim chamada por ter sido utilizada pelos conspiradores nilistas russos por volta do fim do século XIX, é uma cifra de substituição que combina uma cifra de Políbio (página 104) com um processo de selecção de alfabeto similar ao da cifra de Vigenère.

Tomemos então uma tabela de Políbio, por exemplo a que figura na Tabela 3.6 (página 104), e uma qualquer palavra, por exemplo TURGENEV[16] A chave para esta cifra e pois constituída não só pela tabela de Políbio mas também pela palavra escolhida para cifrar.

Suponhamos que queremos então cifrar o texto:

```
organizar os camponeses educar o povo
```

Começamos por escrever o texto com a repetição da chave por cima. O texto cifrado é constituído pela soma (módulo 100) dos códigos de cada letra do texto original com o código da respectiva letra da chave. Todos os códigos obtidos através da aplicação da tabela de Políbio.

T	U	R	G	E	N	E	V	T	U	R	G	E	N	E	V
54	35	42	31	32	25	32	15	54	35	42	31	32	25	32	15
o	r	g	a	n	i	z	a	r	o	s	c	a	m	p	o
12	42	31	22	25	24	41	22	42	12	44	34	22	45	13	12
66	**77**	**73**	**53**	**57**	**49**	**73**	**37**	**96**	**47**	**86**	**65**	**54**	**70**	**45**	**27**

T	U	R	G	E	N	E	V	T	U	R	G	E	N	E	V
54	35	42	31	32	25	32	15	54	35	42	31	32	25	32	15
n	e	s	e	s	e	d	u	c	a	r	o	p	o	v	o
25	32	44	32	44	32	14	35	34	22	42	12	13	12	15	12
79	**67**	**86**	**63**	**76**	**57**	**46**	**50**	**88**	**57**	**84**	**43**	**45**	**37**	**47**	**27**

O texto cifrado é então:

```
66 77 73 53 57 49 73 37 96 47 86 65 54 70 45 27 79 67 86
63 76 57 46 50 88 57 84 43 45 37 47 27
```

[16] Ivan Turgenev é o autor de um romance "Pais e filhos" que tem como uma das principais personagens, Bazarov, que se descreve como um nilista.

Draft 2021.1