Exercices et **problèmes** de **cryptographie**

Exercices et problèmes de cryptographie

Damien Vergnaud
Professeur à Sorbonne Université

Préface de **Jacques Stern**
Professeur à l'École normale supérieure

4e édition

DUNOD

Illustration de couverture : © faithie/Shutterstock

11 rue Paul Bert, 92240 Malakoff
www.dunod.com
ISBN 978-2-10-085284-0

PRÉFACE

« Pour devenir habile en quelque profession que ce soit, il faut le concours de la nature, de l'étude et de l'exercice ». Cette maxime d'Aristote semble bien mal s'appliquer à la cryptologie tant l'exercice y est absent. Il existe de multiples ouvrages de référence de qualité mais, pour la plupart, ils sollicitent très peu l'initiative des étudiants. Et même ceux – rares – qui sont accompagnés d'un véritable choix de problèmes à résoudre, par exemple sous forme d'un livre compagnon, ne couvrent pas totalement une discipline qui connaît une évolution rapide. C'est donc un réel manque que vient combler le recueil que propose Damien Vergnaud.

Le livre que j'ai le plaisir de présenter est issu d'un vrai travail de terrain puisqu'il est le résultat de plusieurs années d'enseignement de la cryptologie à l'Ecole normale supérieure. À l'évidence, l'auteur a beaucoup de talent pour éveiller l'intérêt des étudiants et les conduire, pas à pas, à s'approprier les concepts et les méthodes de la science du secret. Beaucoup de culture également, puisque les sujets choisis sont extrêmement variés à l'image d'une science qui emprunte à l'algèbre, à la théorie des probabilités, à l'algorithmique, à la théorie de l'information. D'ailleurs, ils débordent largement le cadre strict de la cryptographie. Ce talent et cette culture conduisent à un choix d'exercices qui ne demandent pas simplement à l'étudiant de faire des gammes mais lui proposent de s'attaquer à de véritables compositions : ici un effort raisonnable de programmation illustre des cryptanalyses célèbres comme celle de l'Enigma ou celle du programme Venona qui a permis l'interception de communications où les services russes mettaient incorrectement en oeuvre le chiffrement jetable ; là une invitation à « mettre la main à la pâte » permet d'entrer de plain pied dans les méthodes modernes de cryptanalyse – différentielle et linéaire – des algorithmes conventionnels tels que le DES ou l'AES ; là encore, une initiation progressive aux méthodes de factorisation d'entiers, intiment liées à la sécurité du RSA est proposée.

Présenter un tel ouvrage comme un simple livre d'exercices est le reflet de la modestie de son auteur. Certes, il permet la pratique nécessaire à l'acquisition des éléments essentiels de la cryptologie. Mais il va au-delà de cet objectif : chaque chapitre inclut une présentation qui est un véritable cours d'introduction et l'ensemble constitue de fait une forme d'ouvrage d'enseignement avancé fondé sur la pratique. En d'autres termes, le lecteur qui va au terme de tous les exercices proposés est

déjà un véritable spécialiste, capable de se confronter aux multiples concepts que la cryptologie moderne a développés ces trente dernières années. À un moment où la cryptologie est au cœur de la société de l'information, de l'internet aux moyens de paiement en passant par les téléphones portables, une telle expertise est indispensable et il faut souhaiter au livre de Damien Vergnaud des lecteurs à la fois nombreux et actifs.

Jacques Stern, Professeur à l'Ecole normale supérieure

TABLE DES MATIÈRES

AVANT-PROPOS

La cryptologie est un ensemble de techniques permettant d'assurer la sécurité des systèmes d'information. Cette discipline permet notamment de conserver aux données leur caractère de confidentialité, de contrôler leur accès ou d'authentifier des documents. L'utilisation de la cryptographie est de plus en plus répandue et les utilisateurs des systèmes cryptographiques doivent être en mesure non seulement de comprendre leur fonctionnement mais aussi d'en estimer la sécurité.

Cet ouvrage s'adresse aux étudiants de second cycle d'informatique ou de mathématiques. Il s'est développé à partir de textes de travaux dirigés et de travaux pratiques proposés à des étudiants du *Master Parisien de Recherche en Informatique* et aux élèves de première année de l'*École normale supérieure*. Il a été conçu pour aider à assimiler les connaissances d'un cours d'introduction à la cryptologie et à se préparer aux examens. Il présente les outils mathématiques et algorithmiques utiles en cryptographie et les fonctionnalités cryptographiques de base dans le cadre de la cryptographie symétrique et asymétrique.

Cet ouvrage est destiné directement aux étudiants de « master 1 » mais certains exercices pourront être abordés par un étudiant motivé de licence ayant un goût pour l'algorithmique dans ses aspects mathématiques et pratiques. À l'intention des étudiants plus avancés, nous avons inclus des énoncés plus difficiles qui sont alors signalés par une étoile ($\star$). Enfin, l'ouvrage sera utile aux enseignants de cryptologie qui y trouveront un support pour leurs travaux dirigés.

La cryptologie est liée à d'autres disciplines mathématiques et informatiques comme l'arithmétique, l'algèbre, l'algorithmique, ou la théorie de la complexité. Le bagage informatique et mathématique requis pour aborder ce livre est celui que l'on acquiert lors des deux premières années de licence ou en classes préparatoires scientifiques augmenté de quelques notions de théorie des nombres de niveau 3ème année. Ces notions plus avancées font l'objet de brefs rappels qui n'ont cependant pas pour ambition de remplacer un livre de cours.

Le but de cet ouvrage est de permettre à ceux qui le souhaitent de s'initier à la cryptographie par l'exemple. Il propose plus de 140 exercices et problèmes entièrement utilisés dans le cadre de travaux dirigés, de travaux pratiques ou d'examens. Ces exercices sont entièrement corrigés mais le lecteur ne tirera profit de ce livre que s'il cherche des solutions personnelles avant d'en étudier les corrections. L'étude de la cryptologie moderne ne peut se concevoir sans un ordinateur à portée de main et le

livre propose de nombreux exercices de programmation qui ont pour but notamment d'acquérir une pratique de la cryptanalyse. Ces exercices sont signalés par le symbole ⌨ alors que les exercices qui ne nécessitent pas de programmation sont indiqués par le symbole ✍. Les données numériques des exercices de programmation sont disponibles en ligne sur le site

`https://www.dunod.com/EAN/9782100784615`

pour épargner au lecteur la tâche fastidieuse consistant à recopier les énoncés des exercices avant de les traiter. Le lecteur trouvera également des exercices supplémentaires et des références complémentaires sur ce site.

Avant propos de la quatrième édition. Cette nouvelle édition a été inspirée par les nombreuses demandes et remarques que m'ont envoyées des étudiants et collègues, utilisateurs des éditions précédentes. Elle m'a donné l'occasion de modifier et réécrire des parties importantes du texte en suivant ces remarques sur le contenu, le style et l'organisation de l'ouvrage. En plus d'épurer le texte de ses inévitables erreurs typographiques et coquilles, j'ai simplifié et clarifié une grande partie des énoncés des exercices et de leurs solutions. J'ai notamment ajouté des questions intermédiaires pour simplifier la résolution de certains exercices et détaillé certains points techniques dans des solutions d'exercices complexes. J'ai supprimé certains exercices jugés trop difficiles et j'en ai également ajouté de nouveaux. Enfin, les compléments en ligne qui accompagnent l'ouvrage ont été enrichis de nombreux exercices supplémentaires et d'autres exercices et compléments de cours seront ajoutés progressivement.

Remerciements. J'adresse un chaleureux merci à CH. BOUILLAGUET, D. NACCACHE et J. STERN avec qui j'ai eu le plaisir d'enseigner la cryptologie à l'ENS et à Sorbonne Université. Ma gratitude va également aux étudiants et aux élèves normaliens qui ont testé, malgré eux, la majorité des exercices présentés dans ce livre. Ce texte doit beaucoup à des conversations de couloirs et je tiens également à remercier les doctorants, post-doctorants et membres permanents de l'équipe Cryptographie de l'ENS - et particulièrement P.-A. FOUQUE - et de l'équipe ALMASTY de Sorbonne Université pour toutes les discussions que nous avons pu avoir. Je tiens à remercier J. JEAN pour la création et la maintenance du dépôt *TikZ for Cryptographers* [46] à partir duquel certaines figures de ce livre ont été adaptées. Enfin, je voudrais remercier A. BAUER, J.-L. BLANC, O. BLAZY, G. CASTAGNOS, C. CHEVALIER, P.-A. FOUQUE, A. GUILLEVIC, A. HUCHET, M. JALLIER-LUNDGREN, L. KHATI, F. LAGUILLAUMIE, R. LESCUYER, N. LIONET, B. MARTIN, F. MARTINEZ, A. PASSELÈGUE, D. H. PHAN, A. VALENCIA, J. VERGNAUD et V. ZUCCA pour la rigueur et la pertinence de leurs nombreux commentaires.

NOTATIONS

Les conventions et notations suivantes sont utilisées dans cet ouvrage.

Ensembles. Nous utilisons les notations ensemblistes classiques : $\emptyset$ désigne l'ensemble vide ; $x \in A$ signifie que x est un élément de l'ensemble A. Pour deux ensembles A et B, $A \subseteq B$ indique que A est un sous-ensemble de B (alors que $A \subset B$ indique que A est un sous-ensemble strict de B). De plus, $A \cup B$ désigne la réunion de A et B, $A \cap B$ désigne l'intersection de A et B, $A \setminus B$ l'ensemble des éléments de A qui ne sont pas dans B et $A \times B$ le produit cartésien des ensembles A et B. Le cardinal d'un ensemble A est noté $\#A$.

Nous utilisons les notations classiques suivantes pour désigner certains ensembles :

$\mathbb{N}$	ensemble des entiers naturels
$\mathbb{P}$	ensemble des nombres premiers
$\mathbb{Z}$	anneau des entiers relatifs
$\mathbb{Q}$	corps des nombres rationnels
$\mathbb{R}$	corps des nombres réels
$\mathbb{C}$	corps des nombres complexes
$(\mathbb{Z}/n\mathbb{Z})$	anneau des résidus modulo un entier $n \geq 1$
$\mathbb{F}_q$	corps fini à q éléments
$\mathfrak{S}_A$	groupe de permutations de l'ensemble A
A^*	groupe des éléments inversibles d'un anneau A
$\mathcal{M}_{n,m}(A)$	ensemble des matrices carrées $n \times m$ à coefficients dans un anneau A
$\mathcal{M}_\ell(A)$	anneau des matrices carrées $\ell \times \ell$ à coefficients dans un anneau A
$A[X]$	anneau des polynômes à une indéterminée X à coefficients dans un anneau A

La lettre p désigne le plus souvent un nombre premier $p \in \mathbb{P}$ et nous notons $(p_n)_{n\geq 1}$ la suite croissante des nombres premiers (avec $p_1 = 2$, $p_2 = 3$, ...). Pour un polynôme $P \in A[X]$, nous notons $\deg P$ le degré de P.

La notation $\mathbb{G}$ désigne un groupe dont la loi est notée multiplicativement. L'élément neutre pour la multiplication dans $\mathbb{G}$ est noté $1_\mathbb{G}$. L'ordre d'un groupe $\mathbb{G}$ est noté $|\mathbb{G}| = \#\mathbb{G}$ et $\langle g \rangle$ désigne le sous-groupe de $\mathbb{G}$ engendré par $g \in \mathbb{G}$.

Pour un entier $n \geq 2$, un élément de l'anneau $(\mathbb{Z}/n\mathbb{Z})$ sera identifié avec l'unique membre de sa classe de l'ensemble $\{0, \ldots, n-1\}$. Pour un entier $a \in \mathbb{Z}$, $(a \bmod n)$

désignera reste de la division euclidienne de a par n (dans $\{0, \ldots, n-1\}$). Pour deux entiers $a, b \in \mathbb{Z}$, la relation $a \equiv b \pmod{n}$ indique que a est congru à b modulo n (c.-à-d. que n divise $a - b$) et $a \not\equiv b \pmod{n}$ que a n'est pas congru à b modulo n.

Fonctions. Nous notons $f : A \longrightarrow B$ pour indiquer que f est une fonction d'un ensemble A dans un ensemble B. Pour un sous-ensemble $A' \subseteq A$, nous notons $f(A') = \{f(a), a \in A'\} \subseteq B$. Pour un sous-ensemble $B' \subseteq B$, nous notons $f^{-1}(B') = \{a \in A, f(a) \in B'\} \subseteq A$.

La composition de fonctions est notée $\circ$. Pour une fonction $f : A \longrightarrow A$, f^0 est la fonction identité sur A et $f^{i+1} = f^i \circ f$ pour $i \in \mathbb{N}$.

Nous utilisons les notations classiques suivantes pour désigner certains fonctions :

$\lfloor x \rfloor$	partie entière par défaut de $x \in \mathbb{R}$ ($\lfloor x \rfloor \leq x < \lfloor x \rfloor + 1$)
$\lceil x \rceil$	partie entière par excès de $x \in \mathbb{R}$ ($\lceil x \rceil - 1 < x \leq \lceil x \rceil$)
$\ln(x)$	logarithme népérien de $x \in \mathbb{R}$ ($x > 0$)
$\log(x)$	logarithme en base 2 de $x \in \mathbb{R}$ ($x > 0$)
$\text{logdisc}_g(h)$	logarithme discret de $h \in \langle g \rangle$ en base $g \in \mathbb{G}$
$\pi(x)$	nombre de nombres premiers inférieurs ou égaux à x ($\#\{p \in \mathbb{P}, p \leq x\}$)
$\Psi(x, y)$	fonction de Dickman-De Bruijn ($\#\{n \in \mathbb{N}, n \leq x$ et n est y-friable$\}$)
$\binom{n}{m}$	coefficient binomial $\binom{n}{m} = \frac{n!}{m!(n-m)!}$ pour $0 \leq m \leq n$
$\left(\frac{x}{m}\right)$	symboles de Legendre et de Jacobi
$a \mid b$	a divise b
$a \nmid b$	a ne divise pas b
$\text{pgcd}(a, b)$	plus grand commun diviseur de $a, b \in \mathbb{Z}$
$\text{ppcm}(a, b)$	plus petit commun multiple de $a, b \in \mathbb{Z}$
$\mathbf{P}(E)$	probabilité d'un événement E
$\mathbb{E}(X)$	espérance d'une variable aléatoire X
$\varphi(n)$	fonction indicatrice d'Euler $\varphi(n) = \lvert(\mathbb{Z}/n\mathbb{Z})^*\rvert$

Chaînes binaires. Dans les chapitres 2,3 et 4, nous utilisons une fonte de type « machine à écrire » pour représenter la valeur d'un octet avec deux chiffres hexadécimaux : `00` $= 0$, `01` $= 1$, …, `0A` $= 10$, …, `10` $= 16$, …, `FF` $= 255$.

Nous utilisons les notations classiques suivantes sur les chaînes binaires :

$\{0,1\}^n$	ensemble des chaînes de binaires de longueur n
$\{0,1\}^*$	ensemble des chaînes de binaires de longueur finie
$\wedge$	et logique (bit-à-bit pour deux chaînes de même longueur)
$\vee$	ou logique (bit-à-bit pour deux chaînes de même longueur)
$\oplus$	« ou exclusif » (bit-à-bit pour deux chaînes de même longueur)
$\neg$	non logique (bit-à-bit)
$\|x\|$	longueur binaire d'une chaîne $x \in \{0,1\}^*$
$\overline{x}$	chaîne binaire complémentaire de x ($\overline{x} = \neg x$)
$x\|\|y$	concaténation des chaînes x et y
x^n	concaténation de la chaîne x n fois ($\underbrace{x\| \ldots \|x}_{n \text{ fois}}$)
$x[a..b]$	sous-chaîne de x formée des bits situés entre les positions a et b (inclues).
$\lll i$	rotation à gauche d'une chaîne de bits de i positions

Notations algorithmiques. Les algorithmes sont présentés sous forme de pseudo-code simple (notamment en s'affranchissant des problèmes de mémoire). Les entrées et les sorties sont toujours précisées. Les structures de contrôle classiques sont notées en gras (**tant que** condition **faire** instructions, **si** condition **alors** instructions **sinon** instructions, ...).

Les commentaires dans les algorithmes sont signalés par le symbole $\triangleright$. Le symbole $a \leftarrow b$ indique l'assignation algorithmique (c.-à-d. a prend la valeur de b) et le symbole $a \xleftarrow{\text{⚄⚄}} A$ l'assignation d'un élément tiré uniformément aléatoirement (c.-à-d. un élément est tiré uniformément aléatoirement dans l'ensemble A et la valeur obtenue est enregistrée dans a).

Dans ce livre, nous n'étudierons que des algorithmes classiques et pas les algorithmes conçus pour être exécutés sur un ordinateur *quantique* (c.-à-d. qui exploite les phénomènes de la mécanique quantique). La construction d'un ordinateur quantique à grande échelle semble lointaine en raisons de difficultés techniques mais un tel ordinateur présenterait un danger en cryptographie.
Des encadrés de ce type seront utilisés pour mentionner (sans détails) des résultats importants liés aux algorithmes quantiques en cryptographie.

CRYPTOGRAPHIE CLASSIQUE

1

La cryptologie est une science très ancienne : les hommes ont toujours eu besoin de dissimuler des informations et de transmettre des messages en toute confidentialité. Le terme cryptologie vient du grec *kruptos* (κρυπτός) signifiant *secret, caché* et de *logos* (λογός) signifiant *discours*. La cryptologie est donc la *science du secret*. Elle regroupe la cryptographie et la cryptanalyse : la première a pour but de concevoir des systèmes visant à assurer la sécurité des communications sur un canal de communication public et la seconde vise à trouver des failles dans ces systèmes.

La cryptographie est traditionnellement utilisée pour dissimuler des messages aux yeux de certains utilisateurs et le chiffrement des communications militaires a depuis l'Antiquité été une préoccupation majeure des diverses forces armées. Le *chiffrement* regroupe les techniques mises en œuvre pour brouiller la signification d'un message qui est matériellement visible. Le contenu du message ne doit alors être récupérable que par les personnes auxquelles le message est adressé. Le chiffrement fait appel à deux processus élémentaires de transformation du message pour satisfaire ces propriétés :

— la *substitution* qui consiste à remplacer, sans en modifier l'ordre, les symboles d'un texte clair par d'autres symboles,
— et la *transposition* qui repose sur le bouleversement de l'ordre des symboles (mais pas leur identité).

Dans ce chapitre, nous allons étudier des systèmes de chiffrement relativement simples qui ont été utilisés de l'Antiquité (*chiffrement de César* ou *scytale*) jusqu'au début du XX^{e} siècle (*chiffrement de Vernam*, *chiffrement de Hill*, *machine Enigma*).

1.1 Chiffrement par substitution monoalphabétique

Le *chiffrement par substitution* consiste à remplacer dans un message un ou plusieurs symboles par un ou plusieurs symboles (généralement du même alphabet) tout en conservant l'ordre de succession des symboles du message. Dans cette section, nous considérons le chiffrement par substitution *monoalphabétique* qui consiste à remplacer chaque symbole individuel du message par un autre symbole de l'alphabet. Nous allons étudier des techniques de cryptanalyse permettant d'attaquer un tel système.

Elles reposent sur l'analyse des fréquences des symboles utilisés dans le texte chiffré et utilisent le fait que, dans chaque langue, certains symboles ou combinaisons de symboles apparaissent plus fréquemment que d'autres. Les systèmes de chif-

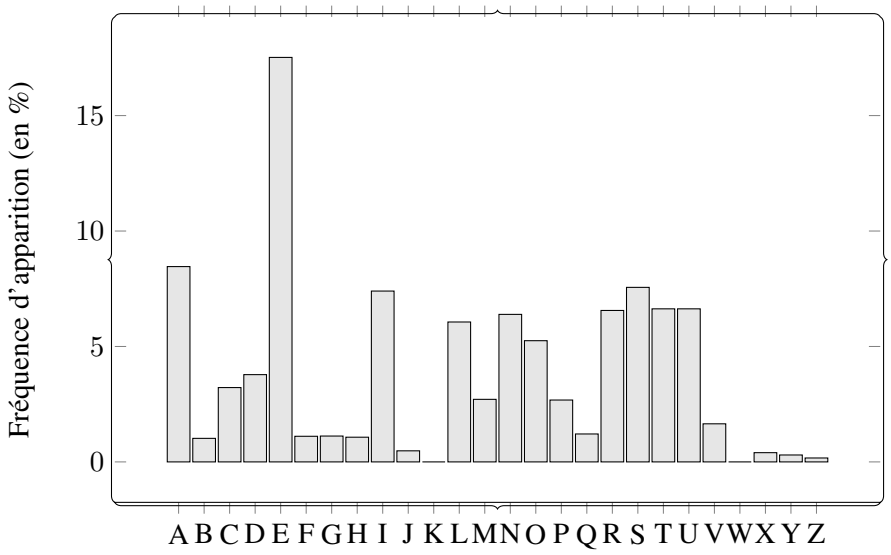

Figure 1.1 – Fréquence d'apparition des lettres en français

frement par substitution monoalphabétique conservent la répartition des fréquences et si le message chiffré est suffisamment long, la recherche d'un symbole ayant une fréquence élevée permettra parfois de retrouver tout ou partie du message clair associé.

La fréquence d'apparition des lettres varie bien évidemment en fonction de la langue et du type de texte considérés. Pour un texte rédigé en français, nous obtenons généralement les fréquences d'apparition (en pourcentage) proches des valeurs suivantes [1] (*cf.* Figure (1.1)) :

a	b	c	d	e	f	g	h	i	j	k	l	m
8,46	1,02	3,22	3,78	17,52	1,11	1,12	1,07	7,40	0,48	0,00	6,06	2,71
n	o	p	q	r	s	t	u	v	w	x	y	z
6,39	5,25	2,68	1,21	6,56	7,56	7,26	6,63	1,65	0,00	0,40	0,30	0,17

De même, certains couples de lettres (ou *bigrammes*) apparaissent plus souvent que d'autres dans une langue donnée. Les vingt bigrammes les plus fréquents [2] de la langue française sont : *es*, *le*, *re*, *de*, *en*, *et*, *ai*, *te*, *ou*, *nt*, *it*, *er*, *la*, *el*, *se*, *qu*, *on*, *ne*, *an* et *ur* (*cf.* Figure (1.2)).

Le chiffrement par substitution monoalphabétique le plus simple est le *chiffrement par décalage*, aussi connu sous le nom de chiffrement de César. Il consiste

1. Ces valeurs correspondent aux fréquences d'apparition des lettres dans le roman *Notre-Dame de Paris* de V. HUGO (en identifiant les lettres accentuées et les lettres non accentuées).

2. Les bigrammes sont classés du plus fréquent au moins fréquent dans le roman *Notre-Dame de Paris*. Les données sont disponibles dans les compléments en ligne du présent livre.

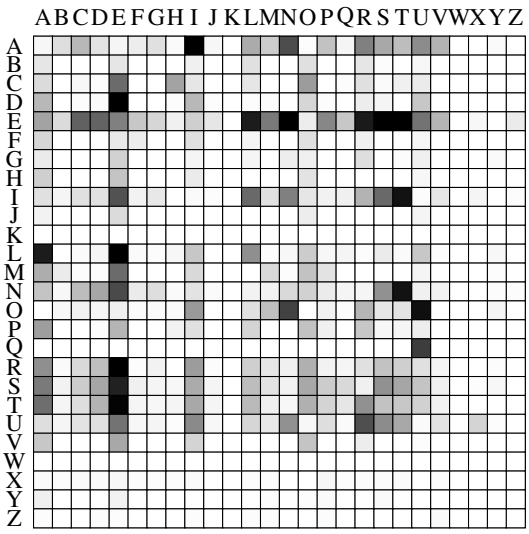

Figure 1.2 – Fréquence d'apparition des bigrammes en français (la fréquence du bigramme `XY` est symbolisée par l'intensité de la teinte de gris de la case à l'intersection de la ligne `X` et de la colonne `Y`)

simplement à décaler les lettres de l'alphabet d'un nombre de positions constant vers la droite ou la gauche. Par exemple, en décalant les lettres de 3 rangs vers la droite (comme le faisait J. CÉSAR), le texte clair *veni vidi vici* devient *yhql ylgl ylfl*.

Exercice 1.1 – Chiffrement de César

Décrypter le texte suivant qui a été obtenu en appliquant le chiffrement de César sur un texte en langue française dans lequel les espaces ont été supprimées :

```
vcfgrwqwoizcuwgtowbsobhgoizcuwgsghqseisqsghoixcifrvi
wxcifrstshsqcaasbhsghqseisjcigbsgojsndogeishobhrsgof
hwgobgjcigbsrsjsndogjcigacbhfsfibxcifcijfwsfgobgojcw
fzsgwbgwubsgrsjcgdfctsggwcbgdofzshcweiszsghhcbashwsf
```

Solution

Le chiffrement de César ne modifie donc pas la fréquence d'apparition des lettres. La lettre la plus fréquente dans un texte français étant le « e », le décalage entre la lettre la plus fréquente dans ce texte chiffré et la lettre « e » doit donc nous révéler la clé utilisée pour le chiffrement. La lettre qui apparaît le plus souvent dans le texte chiffré est le « s » avec 30 occurrences (puis vient la lettre « g » avec 26 occurrences). Le décalage des lettres utilisé lors du chiffrement est donc vraisemblablement de 14

rangs vers la droite dans l'alphabet (qui transforme le « e » en « s ») et, en effectuant l'opération inverse (c.-à-d. un décalage de 14 rangs vers la gauche), on obtient le message clair suivant :

```
horsdiciaulogisfaineantsaulogisestcequecestaujourdhu
ijourdefetecommentestcequevousnesavezpasquetantdesar
tisansvousnedevezpasvousmontrerunjourouvriersansavoi
rlesinsignesdevosprofessionsparletoiquelesttonmetier
```

En ajoutant les espaces et la ponctuation, nous reconnaissons la première réplique de la pièce *Jules César* écrit par W. SHAKESPEARE en 1599 (dans la traduction de É. MONTÉGUT) :

> Hors d'ici ! Au logis, fainéants, au logis ! Est-ce que c'est aujourd'hui jour de fête ? Comment ! Est-ce que vous ne savez pas qu'étant des artisans, vous ne devez pas vous montrer un jour ouvrier, sans avoir les insignes de vos professions ? – Parle, toi, quel est ton métier ?

Le chiffrement affine est un système de chiffrement par substitution mono-alphabétique. La clé consiste en un couple d'entiers $(a, b) \in (\mathbb{Z}/26\mathbb{Z})^* \times \mathbb{Z}/26\mathbb{Z}$. En remplaçant chaque lettre de l'alphabet par son rang (c.-à-d. la lettre « a » est remplacée par 0, la lettre « b » est remplacée par 1, ... et la lettre « z » est remplacée par 25), une lettre du texte clair de rang $i \in \{0, \ldots, 25\}$ est remplacée dans le chiffré par la lettre de rang $(a \cdot i + b \bmod 26)$. Puisque a est inversible dans $\mathbb{Z}/26\mathbb{Z}$, cette transformation est bien une permutation de $\mathbb{Z}/26\mathbb{Z}$ et la permutation inverse est définie par

$$j \longmapsto a^{-1}(j - b) \bmod 26 = \alpha \cdot j + \beta \bmod 26$$

avec $(\alpha, \beta) \in (\mathbb{Z}/26\mathbb{Z})^* \times \mathbb{Z}/26\mathbb{Z}$ avec $\alpha = a^{-1} \bmod 26$ et $\beta = -a^{-1} \cdot b \bmod 26$.

⌨ Exercice 1.2 – Chiffrement affine

Décrypter le texte suivant qui a été obtenu en appliquant le chiffrement affine sur un texte en langue française dans lequel les espaces ont été supprimées :

```
ntjmpumgxpqtstgqpgtxpnchumtputgfsftgthnngxnchumw
xootrtumhpyctgktjqtjchfooxujqhgztumxpotjxotfoqto
hrxumhzutwftgtopfmntjmpuatmfmshodpfrxpjjtqtghbxuj
```

Solution

Comme le chiffrement de César, le chiffrement affine ne modifie pas la fréquence d'apparition des lettres. La lettre la plus fréquente dans un texte français est le « e » et les lettres suivantes sont par ordre de fréquence le « a », le « i », le « s » et le « t »

(avec des fréquences très variables d'un texte à l'autre). La lettre qui apparaît le plus souvent dans le texte chiffré est le « t » avec 24 occurrences puis viennent les lettres « m », « p » et « u » avec 11 occurrences.
En supposant que la lettre « t » (de rang 19) correspond à la lettre « e » (de rang 4) et que la lettre « m » (de rang 12) correspond à l'une des lettres « a », « i », « s » ou « t », nous devons résoudre pour chaque choix un système linéaire à deux équations et deux inconnues dans $\mathbb{Z}/26\mathbb{Z}$ (pour la clé de déchiffrement $(\alpha, \beta) \in (\mathbb{Z}/26\mathbb{Z})^* \times \mathbb{Z}/26\mathbb{Z}$) de la forme :

$$\begin{cases} 19 \cdot \alpha + \beta & \equiv & 4 & \pmod{26} \\ 12 \cdot \alpha + \beta & \equiv & \ell & \pmod{26} \end{cases}$$

où ℓ est l'entier 0, 8, 18 ou 19 selon que la lettre testée pour « m » est « a », « i », « s » ou « t » (respectivement).
Nous obtenons

$$7 \cdot \alpha \equiv (4 - \ell) \pmod{26} \text{ et } \alpha = 15 \cdot (4 - \ell) \bmod 26,$$

puisque $7^{-1} \equiv 15 \pmod{26}$ et $\beta = 4 - \alpha \cdot 19 \bmod 26$. Les couples (α, β) obtenus sont testés en déchiffrant les premiers caractères du chiffré et nous obtenons les résultats suivants :

Lettre testée	« a »	« i »	« s »	« t »
ℓ	0	8	18	19
(α, β)	$(8, 8)$	$(18, 0)$	$(24, 16)$	$(9, 15)$
Début du « clair » associé	`iecaym`	`aegikw`	`qeysmc`	`cestun`

La clé à utiliser pour le déchiffrement est donc vraisemblablement le couple $(9, 15)$ et nous obtenons le message clair suivant :

```
cestuntroudeverdureouchanteuneriviereaccrochantf
ollementauxherbesdeshaillonsdargentoulesoleildel
amontagnefiereluitcestunpetitvalquimoussederayons
```

Il s'agit bien sûr du premier quatrain du sonnet *Le Dormeur du val* écrit par A. RIMBAUD en 1870.

Les systèmes de chiffrement par substitution monoalphabétique les plus généraux utilisent une permutation aléatoire des symboles de l'alphabet utilisé. Pour l'alphabet latin formé de 26 lettres, le nombre de clés possibles est égal à $26! \simeq 2^{88,4}$. Cependant, même si le nombre de clés rend toute recherche exhaustive impossible, les techniques d'analyse fréquentielle permettent de décrypter facilement un chiffré suffisamment long.

Exercice 1.3 – Substitution monoalphabétique

Le texte suivant résulte du chiffrement d'un texte français (traduit de l'anglais) par une substitution monoalphabétique.

```
v ubcfb osu ymoqsuu n cxqfj dqmfnu ub vjcfqu juz amqjmruz
zmsscfusb bquflu auoquz hfszbms zwfba ju wusbms qusbqu ncsz
ju vmo z uddmqvcfb n uxfbuq ju xusb wcoxcfz fj eczzc qcefnuwusb
jc emqbu xfbquu no ijmv nuz wcfzmsz nu jc xfvbmfqu ecz czzul
qcefnuwusb vueusncsb emoq uweuvauq kou z usrmoddqu us wuwu buwez
kou jof os bmoqifjjms nu emozzfuqu ub nu zciju
ju acjj zusbcfb ju vamo vofb ub ju xfuog bcefz c j osu nu zuz
ugbquwfbuz osu cddfvau nu vmojuoq bqme xczbu emoq vu nuejmfuwusb
fsbuqfuoq ubcfb vjmouu co woq ujju quequzusbcfb zfwejuwusb os
usmqwu xfzcru jcqru nu ejoz n os wubqu ju xfzcru n os amwwu n
usxfqms kocqcsbu vfsk csz c j uecfzzu wmozbcvau smfqu cog bqcfbz
cvvusbouz ub iucog
hfszbms zu nfqfruc xuqz j uzvcjfuq fj ubcfb fsobfju n uzzcpuq nu
equsnqu j czvuszuoq wuwu cog wufjjuoquz uemkouz fj dmsvbfmsscfb
qcquwusb cvboujjuwusb n cfjjuoqz ju vmoqcsb ujuvbqfkou ubcfb
vmoeu ncsz jc ymoqsuu v ubcfb osu nuz wuzoquz n uvmsmwfu eqfzuz
us xou nu jc zuwcfsu nu jc acfsu
```

Décrypter ce texte.

Solution

Le nombre d'occurrences de chaque caractère du texte chiffré est donné dans le tableau suivant :

a	b	c	d	e	f	g	h	i	j	k	l	m
10	59	60	8	22	60	5	2	4	48	6	2	38
n	o	p	q	r	s	t	u	v	w	x	y	z
29	49	1	53	6	57	0	157	27	28	13	2	57

Il est vraisemblable que le caractère « u » représente le caractère « **e** » dans le texte clair (nous utiliserons une fonte grasse pour indiquer les lettres appartenant au texte clair). Les autres lettres les plus fréquentes sont le « c » et le « f » mais avec le même nombre d'occurrences et il est difficile de décider à ce stade quels caractères elles représentent.

Les bigrammes les plus fréquents de la langue française sont *es*, *le*, *re*, *de*, *en*, *et*, *ai*, *te*, *ou*, *nt*, *it*, *er* et *la* (*cf.* Figure (1.2)). Les bigrammes commençant par « u » dans le texte chiffré ne sont pas assez fréquents pour décider quel caractère correspond à la lettre « **s** » dans le texte clair. Les bigrammes les plus fréquents du texte chiffré finissant par « u » sont « ju » (18 occurrences) et « nu » (15 occurrences), ce qui suggère que « j » représente « **l** » et « n » représente « **d** ». Ces trois substitutions donnent, pour le premier paragraphe, le texte suivant :

```
v ebcfb ose ymoqsee d cxqfl dqmfde eb vlcfqe lez amqlmrez
zmsscfesb bqefle aeoqez hfszbms zwfba le wesbms qesbqe dcsz
le vmo z eddmqvcfb d exfbeq le xesb wcoxcfz fl eczzc qcefdewesb
lc emqbe xfbqee do ilmv dez wcfzmsz de lc xfvbmfqe ecz czzel
qcefdewesb veeesdcsb emoq eweevaeq koe z esrmoddqe es wewe bewez
koe lof os bmoqifllms de emozzfeqe eb de zcile
```

Le mot « **d**o » suggère que le caractère « o » représente « **u** ». Avec cette déduction, le mot « os**e** » devient « **u**s**e** » qui suggère que le caractère « s » représente « **n** ». Le premier paragraphe du texte devient alors :

```
v ebcfb une ymuqnee d cxqfl dqmfde eb vlcfqe lez amqlmrez
zmnncfenb bqefle aeuqez hfnzbmn zwfba le wenbmn qenbqe dcnz
le vmu z eddmqvcfb d exfbeq le xenb wcuxcfz fl eczzc qcefdewenb
lc emqbe xfbqee du ilmv dez wcfzmnz de lc xfvbmfqe ecz czzel
qcefdewenb veeendcnb emuq eweevaeq kue z enrmuddqe en wewe bewez
kue luf un bmuqifllmn de emuzzfeqe eb de zcile
```

Le mot « k**ue** » qui apparaît deux fois suggère que la lettre « k » représente « **q** ». Les mots « qcef**de**w**en**b » et « **e**b » suggèrent que la lettre « b » représente « **t** ». Les mots « **lu**f » et « f**l** » suggèrent que la lettre « f » représente « **i** » et nous obtenons :

```
v etcit une ymuqnee d cxqil dqmide et vlciqe lez amqlmrez
zmnncient tqeile aeuqez hinztmn zwita le wentmn qentqe dcnz
le vmu z eddmqvcit d exiteq le xent wcuxciz il eczzc qceidewent
lc emqte xitqee du ilmv dez wcizmnz de lc xivtmiqe ecz czzel
qceidewent veeendcnt emuq eweevaeq que z enrmuddqe en wewe tewez
que lui un tmuqiillmn de emuzzieqe et de zcile
```

Le mot « **et**c**it** » suggère que la lettre « c » représente « **a** ». Les mots « w**e**w**e** » et « qce**ide**w**ent** » suggèrent que la lettre « w » représente « **m** ». Le premier paragraphe du texte chiffré devient alors :

```
v etait une ymuqnee d axqil dqmide et vlaiqe lez amqlmrez
zmnnaient tqeile aeuqez hinztmn zmita le mentmn qentqe danz
le vmu z eddmqvait d exiteq le xent mauxaiz il eazza qaeidement
la emqte xitqee du ilmv dez maizmnz de la xivtmiqe eaz azzel
qaeidement veeendant emuq emeevaeq que z enrmuddqe en meme temez
que lui un tmuqiillmn de emuzzieqe et de zaile
```

Le mot « **ent**m**n** » suggère que la lettre « m » représente « **o** ». Le mot « **dan**z » suggère que la lettre « z » représente « **s** ». Le mot « **mau**x**ai**z » suggère que la lettre « x » représente « **v** ». Avec cette déduction, le mot « **e**x**ite**q » suggère que la lettre « q » représente « **r** ». L'expression « **en meme tem**ez » suggère que « e » représente « **p** ». Nous avons obtenu à ce stade la table de correspondance suivante :

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

et le texte suivant :

v **etait une** yournee **d avril** droide **et** vlaire **les** aorlores
sonnaient treile aeures hinston **smita le menton rentre dans**
le vou **s eddorvait d eviter le vent mauvais il passa rapidement**
la porte vitree du ilov **des maisons de la** vivtoire **pas** assel
rapidement vependant **pour empevaer que s enrouddre en meme temps**
que lui un touriillon **de poussiere et de sa**ile

le aall **sentait le** vaou **vuit et le vieu**g **tapis a l une de ses**
egtremites **une a**ddivae **de** vouleur **trop vaste pour** ve **deploiement**
interieur etait vlouee **au mur elle representait simplement un**
enorme visare **larr**e **de plus d un metre le visa**re **d un** aomme **d**
environ quarante vinq **ans a l epaisse moustav**ae **noire au**g **traits**
avventues **et** ieaug

hinston **se dirirea vers l esv**alier **il etait inutile d essa**per **de**
prendre l asvenseur **meme aux meilleures epoques il** donvtionnait
rarement avtuellement **d ailleurs le** vourant **elevtrique etait**
voupe **dans la** yournee v **etait une des mesures d e**vonomie **prises**
en vue de la semaine de la aaine

En terminant l'analyse de façon similaire, nous obtenons la table de correspondance :

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

et le texte complet (soit les premiers paragraphes du roman *1984* de G. ORWELL publié en 1949 dans la traduction de A. AUDIBERTI).

1.2 Chiffrement par substitution polyalphabétique

La vulnérabilité des systèmes de chiffrement par substitution monoalphabétique a poussé les cryptologues à développer à partir du XV^e siècle des systèmes plus élaborés qui utilisent plusieurs alphabets de chiffrement. Le chiffrement par substitution *polyalphabétique* consiste à remplacer un symbole du texte clair par un autre qui dépend de l'état du cryptosystème. Cette modification de la substitution dépend alors de la position du symbole dans le message.

L'utilisation de plusieurs alphabets pour effectuer le chiffrement a pour conséquence que des occurrences différentes d'une même lettre du texte clair ne sont pas toujours chiffrées par la même lettre et des occurrences différentes d'une même lettre dans le texte chiffré ne correspondent pas nécessairement à la même lettre du clair. En particulier, une cryptanalyse par analyse fréquentielle d'un chiffrement par substitution polyalphabétique doit utiliser des techniques plus évoluées que pour des

substitutions monoalphabétiques.

Le chiffrement de Vigenère est un système de substitution polyalphabétique créé par G. B. BELLASO en 1553 et popularisé par B. DE VIGENÈRE en 1586 dans son *Traité des chiffres*. Ce procédé de chiffrement repose sur l'utilisation périodique de plusieurs alphabets de substitution déterminés par la clé (en général un mot). Pour pouvoir chiffrer un texte clair, à chaque caractère nous associons une lettre de la clé pour effectuer le décalage correspondant comme dans le chiffrement de César. Ainsi le texte clair « `vigenere` » chiffré avec la clé « `cle` » devient le chiffré « `xtkgyitp` ». En effet, la lettre `v` chiffrée avec la lettre `c` est décalée de deux positions, la lettre `i` est chiffrée avec la lettre `l` et la lettre `g` chiffrée avec la lettre `e`. Ensuite, la lettre `e` est chiffrée avec la lettre `c` et ainsi de suite de façon périodique.

En particulier, si la longueur de la clé ℓ est connue, retrouver le texte clair à partir du texte chiffré c peut se faire en appliquant une cryptanalyse du chiffrement de César (comme dans l'exercice (1.1)) pour chaque sous-chiffré c_i de c formé uniquement des lettres dont les positions sont congrues à i modulo ℓ (pour $i \in \{0, \ldots, \ell - 1\}$). La difficulté pour le cryptanalyste consiste donc à retrouver la longueur de la clé.

La première méthode pour déterminer la longueur de la clé est connue sous le nom de *test de Kasiski* (d'après F. W. KASISKI). Elle repose sur le fait que si deux groupes de lettres (ou *polygrammes*) du chiffré sont égaux alors il s'agit probablement du même polygramme dans le texte clair chiffré avec la même partie de la clé. La taille de l'intervalle qui sépare ces deux polygrammes identiques dans le chiffré sera donc, dans la majorité des cas, un multiple de la longueur de la clé. S'il y a plusieurs répétitions de polygrammes, le plus grand diviseur commun des distances les séparant est très probablement la taille de clé.

Exercice 1.4 – Chiffrement de Vigenère - test de Kasiski

Le texte suivant a été obtenu en appliquant le chiffrement de Vigenère sur un texte en langue française dans lequel les espaces ont été supprimées :

```
  zbpuevpuqsdlzgllksousvpasfpddggaqwptdgptzweemqzrdjtddefek
  eferdprrcyndgluaowcnbptzzzrbvpssfpashpncotemhaeqrferdlrlw
  wertlussfikgoeuswotfdgqsyasrlnrzppdhtticfrciwurhcezrpmhtp
  uwiyenamrdbzyzwelzucamrptzqseqcfgdrfrhrpatsepzgfnaffisbpv
dblisrplzgnemswaqoxpdseehbeeksdptdttqsdddgxurwnidbdddplncsd
```

Utiliser le test de Kasiski pour déterminer la longueur de la clé utilisée et décrypter ce texte.

Solution

En cherchant les répétitions de chaînes de trois ou quatre caractères dans le texte chiffré, nous obtenons par exemple la répétition des motifs `ferd`, `pas`, `ptz` et `ddd`

à distance 48, 68, 40 puis 156 et 12.

```
  zbpuevpuqsdlzgllksousv pas fpddggaqwptdg ptz weemqzrdjtddefek
  e ferd prrcyndgluaowcnb ptz zzrbvpssf pas hpncotemhaeqr ferd lrlw
  wertlussfikgoeuswotfdgqsyasrlnrzppdhtticfrciwurhcezrpmhtp
  uwiyenamrdbzyzwelzucamr ptz qseqcfgdrfrhrpatsepzgfnaffisbpv
dblisrplzgnemswaqoxpdseehbeeksdptdttqs ddd gxurwnidb ddd plncsd
```

Le plus grand commun diviseur de ces distances est l'entier 4 et nous en déduisons que la clé est probablement de longueur 4. Nous obtenons les quatre sous-chiffrés c_0, c_1, c_2 et c_3 formés des 72 (ou 71 pour c_3) caractères de rang congru à 0, 1, 2 ou 3 modulo 4 (respectivement) :

$$c_0 = \texttt{zeqzkssdqdzmddk...dszmqdhktqdrddc}$$
$$c_1 = \texttt{bvsgsvfgwgwqjee...brgsosbsdsgwbps}$$
$$c_2 = \texttt{ppdloppgppeztff...lpnwxeedtdxndld}$$
$$c_3 = \texttt{uulluadatterdee...vileapeeptduidn}$$

La lettre la plus fréquente dans le chiffré c_0 (*resp.* c_1, *resp.* c_2, *resp.* c_3) est le `d` (*resp.* le `s`, *resp.* le `p`, *resp.* le `e`). En supposant que ces lettres représentent la lettre « e » dans le texte clair, nous en déduisons que la clé utilisée pour le premier chiffré (*resp.* le second chiffré, *resp.* le troisième chiffré, *resp.* le quatrième chiffré) est probablement le `z` (*resp.* le `o`, *resp.* le `l`, *resp.* le `a`). En déchiffrant avec la clé `zola`, nous obtenons ainsi le texte clair :

```
  aneufheureslasalledutheatredesvarietesetaitencorevidequel
  quespersonnesaubalconetalorchestreattendaientperduesparmi
  lesfauteuilsdeveloursgrenatdanslepetitjourdulustreademife
  uxuneombrenoyaitlagrandetacherougedurideauetpasunbruitnev
enaitdelascenelarampeeteintelespupitresdesmusiciensdebandes
```

En ajoutant les espaces et la ponctuation, nous retrouvons le texte :

> Á neuf heures, la salle du théâtre des Variétés était encore vide. Quelques personnes, au balcon et à l'orchestre, attendaient, perdues parmi les fauteuils de velours grenat, dans le petit jour du lustre à demi-feux. Une ombre noyait la grande tache rouge du rideau ; et pas un bruit ne venait de la scène, la rampe éteinte, les pupitres des musiciens débandés.

qui débute le roman *Nana* écrit en 1880 par É. ZOLA.

Une méthode plus générale pour déterminer la longueur de la clé dans un système de chiffrement polyalphabétique a été proposée par W. FRIEDMAN en 1920. Elle repose sur le calcul de l'indice de coïncidence qui permet de mesurer la probabilité

de tirer deux lettres identiques en les choisissant aléatoirement dans un texte donné. L'indice se calcule par la formule suivante :

$$I = \sum_{i=0}^{25} \frac{n_i(n_i - 1)}{n(n-1)}$$

où n_i est le nombre de lettres de rang i dans le texte chiffré (pour $i \in \{0, \ldots, 25\}$) et n la longueur du texte chiffré. Dans le cas d'un texte aléatoire (c.-à-d. où les lettres sont tirées uniformément aléatoires dans l'alphabet $\{a, b, \ldots, z\}$), l'indice de coïncidence est proche de $1/26 \simeq 0{,}0385$. Dans un langage structuré, l'indice de coïncidence ne dépend que de la distribution des lettres de l'alphabet et n'est pas modifié par une substitution monoalphabétique. En français, l'indice de coïncidence vaut environ $0{,}078$ et il est donc suffisamment éloigné de celui d'un texte aléatoire pour permettre de distinguer un texte chiffré par une substitution monoalphabétique d'un texte aléatoire. Pour un chiffré produit par le chiffrement de Vigenère, si étant donné un entier ℓ, chaque sous-chiffré, formé uniquement des lettres dont les positions sont congrues à une valeur fixée modulo ℓ, a un indice de coïncidence proche de $0{,}078$, alors ℓ est vraisemblablement un multiple de la longueur de la clé utilisée.

Exercice 1.5 – Chiffrement de Vigenère - indice de coïncidence

Le texte suivant a été obtenu en appliquant le chiffrement de Vigenère sur un texte en langue française dans lequel les espaces ont été supprimées :

```
ufzbdemltfnlfgmoneefrttkophfeiuplbfxbtrmltfczfyg
ipzjygblyyjivigpfpffveptmfmfavjyxbymfcisptpyhjeu
vppmpemwejeiopjgpxbweqpgipwpygilrelmmciqcmtpioei
   nzmhyejeiuditpwqlhstpmpwslgpcrjpwqlvmpwfw
```

Utiliser l'indice de coïncidence pour déterminer la longueur de la clé utilisée et décrypter ce texte.

Solution

Dans une première étape de cryptanalyse, nous calculons l'indice de coïncidence moyen obtenu à partir des sous-chiffrés en fonction de la longueur de la clé.

Longueur	1	2	3	4	5	6	7	8
Indice	0,0534	0,0518	0,0816	0,0494	0,0492	0,0792	0,0529	0,0485

Nous en déduisons que la clé est probablement de longueur 3 ce qui donne les cinq sous-chiffrés c_i pour $i \in \{0, 1, 2\}$ formés des caractères dont le rang est congru à i

modulo 3 :

$$c_0 = \texttt{ubmffoetofubbmf...onhjutqsmspjqmf}$$
$$c_1 = \texttt{fdlngnftpepftlc...ezyedpltplcplpw}$$
$$c_2 = \texttt{zetlmerkhilxrtz...iimeiiwhpwgrwvw}$$

La lettre la plus fréquente dans le chiffré c_i est le `f` pour $i = 0$, le `p` pour $i = 1$ et le `i` pour $i = 2$. En supposant que ces lettres représentent la lettre « e » dans le texte clair, nous en déduisons que la clé utilisée pour le chiffré c_i pour $i \in \{0, 1, 2\}$ est le `b`, le `l` et le `e` (respectivement). Nous obtenons ainsi le texte clair

```
tuvasalapechevincadunsignedetetehautainlaperven
chevincaauxyeuxcouleurdepluieprintaniererepondi
tquelleallaiteneffetalapechesonchandailreprisee
 ntemoignaitetsesespadrillesracorniesparlesel
```

En ajoutant les espaces et la ponctuation, nous retrouvons le texte

> –Tu vas à la pêche, Vinca ?
> D'un signe de tête hautain, la Pervenche Vinca aux yeux couleur de pluie printanière, répondit qu'elle allait, en effet, à la pêche. Son chandail reprisé en témoignait, et ses espadrilles racornies par le sel.

qui débute le roman *Le Blé en herbe* écrit en 1923 par COLETTE.

Une faiblesse du chiffrement de Vigenère est son usage répété de la clé qui permet aux cryptanalystes d'en déterminer la longueur (comme nous venons de le voir dans les deux exercices précédents). Un système de chiffrement est dit *autoclave* s'il emploie une partie du texte clair comme clé de chiffrement. Dans le chiffrement autoclave de Vigenère, les décalages des lettres sont déterminés initialement par une clé (en général encore un mot), puis par le texte clair lorsque les lettres de la clé sont épuisées. Ainsi le texte clair « `vigenere` » chiffré avec la clé « `cle` » devient le chiffré « `xtkzvkvr` ». Comme dans le chiffrement de Vigenère, le début du texte `vig` est chiffré avec le mot `cle` mais le reste du texte `enere` est chiffré avec le mot `vigen`. Le déchiffrement s'effectue séquentiellement à l'aide de la clé puis du début du texte clair reconstitué. Ce procédé est donc particulièrement sensible aux erreurs de transmission du message chiffré.

Exercice 1.6 – Chiffrement autoclave de Vigenère

Le texte suivant a été obtenu en appliquant le chiffrement autoclave de Vigenère sur un texte en langue française (traduit de l'anglais) dans lequel les espaces ont été supprimées :