Feuille d'exercices n°3 de l'option d'informatique.

Jules César a utilisé pendant la guerre des Gaules un chiffrage de substitution pour communiquer avec ses troupes. Ce chiffrage consisté à substituer les lettres de l'alphabet par d'autres lettres obtenue par décalage de l'alphabet. Pour le décalage suivant :

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

le texte «la guerre des gaules» devient alors «NC IWGTTG FGU ICWNGU». Pour connaître le code utilisé, il suffit de connaître la lettre qui correspond à la lettre **a** (**C** dans le cas de l'exemple précédent). Par convention, on écrira le texte codé en lettres majuscules et le texte en clair en minuscules.

1. Étant donné le code, écrire la fonction de cryptage correspondant au chiffre de César (compte tenu de la convention donnée précédemment).
exemple :

```
#encode ‘c‘ "la guerre des gaules";;
- : string = "NC IWGTTG FGU ICWNGU"
```

2. Faire de même pour la fonction de déchiffrage.
exemple :

```
#decode ‘c‘  "NC IWGTTG FGU ICWNGU";;
- : string = "la guerre des gaules"
```

3. Dans le cadre du contre espionnage, on intercepte un message chiffré et on en cherche la signification. Si on sait que le message a été codé avec le chiffre de César, on peut évidemment essayer les 26 codes possibles et retrouver le texte original parmi les 26 textes obtenus (c'est celui qui aura un sens). On veut obtenir une méthode automatique qui nous donnera la clef et le message original. Pour cela, on va applique une méthode de cryptanalyse, c'est-à-dire de décodage d'un message sans en connaître la clef. Le principe de cryptanalyse que nous allons appliquer repose sur le fait que pour une langue donné la fréquence d'apparition de chacune des lettres de l'alphabet n'apparaissent pas avec la même fréquence. Ainsi en français, le lettre **e** est beaucoup plus fréquente que la lettre **z**.

a) À partir du texte de référence, établir le tableau de pourcentage d'apparition de chacune des lettres de l'alphabet (on ne tient pas compte des accents qui ont été supprimé dans le texte de référence).

b) Faire de même avec le texte crypté.

c) L'idée pour trouver la clef, est de faire «tourner» le tableau f_c des fréquences d'apparition des lettres du texte chiffré pour le faire coïncider le mieux possible avec le tableau f_r des fréquences d'apparition de référence.

i Écrire une fonction «distance» qui calcule $d[i] = \sum_{j=0}^{25} \left| f_r[j] - f_c[(j+i) \bmod 26] \right|$

$d \rightarrow$ | d_0 | d_1 | d_2 | $\ldots$ | d_{25} |

$f_r \rightarrow$ | $f_r[0]$ | $f_r[1]$ | $f_r[2]$ | $\ldots$ | $f_r[25]$ | ← `tableau des fréquences de référence`

$f_c \rightarrow$ | $f_c[0]$ | $f_c[1]$ | $f_c[2]$ | $\ldots$ | $f_c[25]$ | ← `tableau des fréquences du texte chiffré`

ii Écrire une fonction qui trouve i_0 tel que $d[i_0] = \min\left\{ d[i] / i \in [0,25] \right\}$

iii Écrire une fonction qui à l'aide des fonctions précédentes décode automatiquement le texte codé.