Sujet 3 : Comment communiquer avec Mars.

Nous sommes en 2097. En devançant de quelques années les projets américains, russes et chinois, un équipage européen composé d'une espagnole, un français, deux polonais et une italienne vient de se poser sur Mars. Voici le message qu'il vient de nous envoyer : "Noys pous somres boigré da uneee pistine roge avic das cannnnnnards blius". Mais comment peut-on faire pour communiquer correctement avec notre équipage ?

Une transmission sur une telle distance est obligatoirement parasitée (notamment à cause de diverses sources de perturbations électromagnétiques). Pourtant, il est primordial que les communications avec notre vaisseau soient rapides et fiables. Pour ce faire, notre équipe de MathEnJeans va développer un système qui permet de reconstituer les messages.

L'idée qui régit le projet n'est pas nouvelle et est présente dans toutes les langues. On part d'un alphabet, par exemple $A = \{a, b, c, d, \ldots, z\}$. Une suite finie de lettres forment un mot (par exemple $abbfty$ est un mot de longueur 6). Certains mots sont admis (par exemple *canard*) (ce sont ceux qui se trouvent dans le dictionnaire) et d'autre pas (par exemple *connnard*). Pour modéliser la situation on notera E l'ensemble de tous les mots et A l'ensemble des mots admissibles, c'est-à-dire ceux qui se trouvent dans le dictionnaire. Si la station terrestre reçoit un mot non-admissible il le corrigera en le remplaçant par le mot admissible le "plus proche". Ainsi le mot admissible le plus proche de *connnard* est *canard*, aussi la station remplacera *connnard* par *canard*.

Certains systèmes de correction de messages utilisent un dictionnaire complet (comme en français) mais :

- la correction est lente pour des messages instantanés,
- certains messages tels que les ondes (corrections des transmissions musicales) n'ont pas de dictionnaire rattaché.

Voici une autre façon de procéder. On se donne un alphabet (disons $A = \{a, b, c\}$). On dit que notre dictionnaire de base est l'ensemble de tous les mots de longueur 5 (par exemple *abcba*). La navette sur Mars envoie 4 fois à la suite le même mot, par exemple *abcbaabcbaabcbaabcba*. Dans la pratique, E est l'ensemble des mots de longueur 20 et A est l'ensemble des mots de longueur 20 de la forme $x_1x_2x_3x_4x_5x_1x_2x_3x_4x_5x_1x_2x_3x_4x_5x_1x_2x_3x_4x_5$. La *différence* entre deux mots $m = x_1 \cdots x_{20}$ et $m' = x'_1 \cdots x'_{20}$, notée $d(m, m')$, est le nombre de i tels que $x_i \neq x'_i$. Par exemple,

$$d(abcbaabcbaabcbaabcba, acbbaabcbaaccbaabcba) = 3.$$

Ainsi la station terrestre remplacera tout mot m' reçu par le mot $m \in A$ tel que $d(m, m')$ soit minimal.

Question 1. Soit m' un mot quelconque. Est-ce que le mot $m \in A$ tel que $d(m, m')$ est minimal est unique ? Soit m un mot émis et m' le mot reçu. Il est clair que, si $d(m, m')$ est grand, il est impossible de reconstituer m à partir de m'. Jusqu'à quelle valeur de $d(m, m')$ est-on certain de reconstituer m' à partir de m ? Pouvez-vous donner une méthode ou algorithme qui permette dans ce cas de reconstituer m à partir de m' ?

Question 2. Étudier les mêmes questions en faisant varier la taille de l'alphabet, la longueur des mots dans le dictionnaire et/ou le nombre de répétitions du mot.

Question 3. On remarque que, dans l'exemple de la question 1, A est l'ensemble des mots $x_1 \cdots x_{20}$ qui vérifient les équations $x_1 = x_6 = x_{11} = x_{16}$, $x_2 = x_7 = x_{12} = x_{17}$, $x_3 = x_8 = x_{13} = x_{18}$, $x_4 = x_9 = x_{14} = x_{19}$ et $x_5 = x_{10} = x_{15} = x_{20}$. Peut-on définir d'autres systèmes de correction (ensembles A et E) plus efficaces en utilisant d'autres équations ? La notion de "plus efficace" et la notion de "autres équations" sont à définir...