Exercices du cours de Théorie de l'Information et Codage
cours 2 du 22 février 2011.

1. Soit une source discrète sans mémoire d'entropie $H(U)$. On considère un encodage des suites de k lettres de la source dans des mots code de longueur n dans un alphabet de taille D. La fonction d'encodage doit être injective et P_E est la probabilité de l'événement: une suite de la source ne correspond à aucun mot code (la fonction d'encodage n'étant pas surjective). On a vu au premier cours que pour tout $\delta > 0$, si $n/k \geq \frac{H(U)+\delta}{\log D}$, alors on peut trouver une fonction d'encodage avec $P_E \to 0$ quand $k \to \infty$. Montrer que si $n/k \leq \frac{H(U)-\delta}{\log D}$ alors quelque soit l'encodage, on a $P_E \to 1$ quand $k \to \infty$.

2. **Un test pour les codes non-ambigus.** Le but de cet exercice est de donner un algorithme qui permet de vérifier si un code est non-ambigu. Voici un exemple d'un code binaire ambigu:
$$C = \{1, 011, 01110, 1110, 10011\}. \tag{1}$$
Le mot $w = 011101110011$ a deux factorisations:
$$w = (01110)(1110)(011) = (011)(1)(011)(10011).$$
L'alphabet D-aire est noté $\mathcal{D}$. L'ensemble des mots sur $\mathcal{D}$ est noté $\mathcal{D}^*$. Pour $x, y \in \mathcal{D}^*$, on définit:
$$x^{-1}y = \{z \in \mathcal{D}^*;\ xz = y\} \text{ et, } xy^{-1} = \{z \in \mathcal{D}^*;\ x = zy\}.$$
Pour des ensembles X, Y de $\mathcal{D}^*$, on étend ces définitions comme suit:
$$X^{-1}Y = \cup_{x\in X} \cup_{y \in Y} x^{-1}y \text{ et, } XY^{-1} = \cup_{x\in X} \cup_{y \in Y} xy^{-1}.$$
Les puissances de X sont définies par $X^0 = \{e\}$ où e est le mot vide, $X^1 = X$ et $X^{n+1} = XX^n = \{xy,\ x \in X, y \in X^n\}$, pour $n \geq 1$.
On voit un code D-aire C comme un sous-ensemble de $\mathcal{D}^+ = \mathcal{D}^* - e$. On définit alors
$$\begin{aligned} U_1 &= C^{-1}C - e, \\ U_{n+1} &= C^{-1}U_n \cup U_n^{-1}C, \text{ pour } n \geq 1. \end{aligned}$$
Nous allons montrer le théorème suivant:

Théorème 0.1 *Le code $C \subset \mathcal{D}^+$ est un code non-ambigu si et seulement si aucun des ensembles U_n définis ci-dessus ne contient le mot vide.*

1) Ecrire U_1, U_2, U_3 pour le code binaire donné par (1). Que vaut U_1 pour un code instantané? Que valent les U_n pour l'exemple de code binaire vu en cours: $\{10, 00, 11, 110\}$?
2) Montrer par induction sur k que: pour tout $n \geq 1$ et $k \in \{1, \ldots, n\}$, on a $e \in U_n$ ssi il existe un mot $u \in U_k$ et des entiers $i, j \geq 0$ tels que:
$$uC^i \cap C^j \neq \emptyset \text{ et } i + j + k = n. \tag{2}$$
3) En déduire le Théorème 0.1.