TD 5 - Chomsky et ambigüité

Exercice 1. *Trouvez la grand-mère*

Donner des grammaires algébriques engendrant les langages suivants.

1. L'ensemble des palindromes sur $\{a, b\}$ et son complémentaire.
2. L'ensemble des mots sur $\{a, b\}$ de longueur impaire.
3. L'ensemble des mots sur $\{a, b\}$ ayant le même nombre d'occurrences de a que de b.
4. L'ensemble des mots sur $\{a, b\}$ ayant deux fois plus de a que de b.
5. $\{w\#\overline{w}\#,\ w \in (a+b)^*\}$.
6. $\{w\#w' | w, w' \in (a+b)^* \text{ et } w \neq w'\}$.
7. L'ensemble des mots de $(a+b)^*$ qui ne sont pas de la forme ww.
   Indication : les mots qui ne sont pas de la forme ww et qui sont de longueur paire sont de la forme xy avec x et y de longueur impaire, et une autre condition sur x et y.

Exercice 2. *Forme normale de Chomsky*

Définition 1. *Une grammaire hors-contexte $G = (V, \Sigma, P, S)$ (Rappel : $V \cap \Sigma = \emptyset$) est sous* forme normale de Chomsky *si toutes ses règles sont de la forme*
i) $A \to BC$ *avec* $B, C \in V$
ii) $A \to a$ *avec* $a \in \Sigma$
iii) $S \to \epsilon$
De plus, si $S \to \epsilon$ est une règle de P, alors $B, C \in V - \{S\}$ dans (i).

1. Mettre la grammaire suivante sous forme normale de Chomsky
$$S \to TbT$$
$$T \to TaT | ca$$
2. Donner des grammaires sous forme normale de Chomsky pour les langages algébriques suivants :
   – $a^n b^{2n} c^k, n \geq 1$
   – l'ensemble des palindromes sur $\{a, b\}$
3. Prouver le théorème suivant

   Théorème 1. *Pour toute grammaire hors-contexte $G = (V, \Sigma, P, S)$, il existe une grammaire $G' = (V', \Sigma, P', S)$ telle que $L(G') = L(G)$ et G' est sous forme normale de Chomsky.*

Exercice 3. *Dangling else*

Stmt $\to$ **if b then** *Stmt* | **if b then** *Stmt* **else** *Stmt* | **a**

1. Montrer que cette grammaire est ambiguë.
2. Proposer une grammaire non ambiguë pour le même langage.

Exercice 4. *Elle est toujours ambigue...*

1. Montrer qu'un langage rationnel ne peut pas être inhérement ambiguë.
2. Montrer que la grammaire suivante est ambiguë.

$$E \longrightarrow E + E \mid E * E \mid (E) \mid a \mid b$$

3. Trouver une grammaire non-ambiguë qui reconnaît le même langage que la grammaire précédente.
4. Trouver une grammaire hors-contexte qui reconnaît le langage

$$A = \left\{ a^i b^j c^k | i, j, k \geq 0 \text{ et } (i = j \text{ ou } j = k) \right\}$$

5. *(Bonus)* Montrer que toute grammaire pour le langage précédent est ambiguë.

   Pour cela, on admettra le lemme d'Ogden suivant (on peut trouver la preuve dans le livre *Langages Formels, Calculabilité et Complexité* d'Olivier Carton p.92) :

   Théorème 2. *Pour toute grammaire $G = (\Sigma, V, P)$ et toute variable $S \in V$, il existe un entier k tel que tout mot $f \in (\Sigma \cup V)^*$ engendré par une dérivation depuis S ayant au moins k lettres distinguées se factorise en $f = \alpha u \beta v \gamma$, où $\alpha, u, \beta, v, \gamma \in (\Sigma \cup V)^*$, avec :*

   1. *$S \longrightarrow^* \alpha T \gamma$ et $T \longrightarrow^* uTv + \beta$*
   2. *soit α, u, et β, soit β, v et γ contiennent tous les trois des lettres distinguées.*
   3. *$u\beta v$ contient moins de k lettres distinguées.*

   En particulier, tous les mots de la forme $\alpha u^n \beta v^n \gamma$ sont engendrés par une dérivation depuis S.

Exercice 5. *Fait divers*

Soit la grammaire suivante :
$S \longrightarrow GN\ GV$
$GN \longrightarrow Det\ N$
$GN \longrightarrow GN\ GP$
$GN \longrightarrow N$
$Det \longrightarrow$ un | une | des | le | la | les |
$N \longrightarrow$ chat | Jean | Mickey | télescope
$GV \longrightarrow V\ GN$
$GV \longrightarrow GV\ GP$
$V \longrightarrow$ regarde | mange
$GP \longrightarrow Prep\ GN$
$Prep \longrightarrow$ avec | chez

Montrer que cette grammaire est ambiguë.