TD 05 – Forme normale supérieure

Lemme (TD04 - 5.1). *Pour toute grammaire $G = (V, T, S, P)$ il existe une grammaire équivalente $G' = (V', T, S', P')$ telle que si $\alpha \to \beta \in P'$ et $|\beta| < |\alpha|$, alors $\alpha = X \in V'$ et $\beta = \epsilon$.*

Idée : $V' = V \cup \{Y\}$. P' est constitué des trois types de règles suivantes : $\alpha \to \beta$ si $\alpha \to \beta \in P$ et $|\alpha| \leq |\beta|$, $\alpha \to Y^{|\alpha|-|\beta|}\beta$ si $\alpha \to \beta \in P$ et $|\alpha| > |\beta|$, et $Y \to \epsilon$.

Lemme (TD04 - 5.2). *Pour toute grammaire $G = (V, T, S, P)$, il existe une grammaire équivalente $G'' = (V'', T, S'', P'')$ telle que*
– *si $\alpha \to \beta \in P''$ et $|\beta| < |\alpha|$, alors $\alpha = X \in V''$ et $\beta = \epsilon$;*
– *pour toute production $\alpha \to \beta \in P''$, $|\beta| \leq 2$.*

Idée : On utilise le lemme précédent pour avoir la première condition satisfaite. Ensuite, si $\alpha \to \beta \in P'$ avec $|\beta| \geq 3$, nécessairement $|\alpha| \leq |\beta|$. Si $|\alpha| < |\beta|$, on remplace $\alpha \to \beta$ par $\alpha \to \beta[0, |\alpha| - 1]X$ et $X \to \beta[|\alpha|, |\beta| - 1]$ où X est un nouveau symbole non terminal. Si $|\alpha| = |\beta|$, avec $\alpha = A_0 \dots A_{m-1}$ et $\beta = B_0 \dots B_{m-1}$, alors $\alpha \to \beta$ est remplacé par les m productions $A_0A_1 \to B_0X_0$, $X_iA_{i+2} \to B_{i+1}X_{i+1}$ pour $0 \leq i \leq m-2$ et $X_{m-2}A_m \to B_{m-2}B_{m-1}$, où les X_i sont de nouveaux symboles non terminaux. On répète ces deux opérations jusqu'à obtenir une grammaire de la bonne forme.

Définition. Une grammaire hors-contexte $G = (V, \Sigma, P, S)$ [1] 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).

Exercice 1. *Forme normale de Chomsky*

1. Mettre la grammaire suivante sous forme normale de Chomsky

$$S \to TbT$$
$$T \to TaT \mid ca$$

2. Prouver le théorème suivant

Théorème. *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.*

Définition. Une grammaire hors-contexte $G = (V, \Sigma, P, S)$ est sous *forme normale de Greibach* si toutes ses règles sont de la forme
i) $A \to aB_1 \dots B_n$
ii) $A \to a$
iii) $S \to \epsilon$
Où $B_1, \dots, B_n \in V - \{S\}$ et $a \in \Sigma$.
On l'appelle également *forme m-standard* lorsque $n \leq m$ pour toutes les règles.

1. rappel : $V \cap \Sigma = \emptyset$

Exercice 2. *forme normale de Greibach*

Nous allons prouver le théorème suivant

Théorème. *Tout langage hors-contexte peut être généré par une grammaire hors-contexte sous forme normale de Greibach.*

1. Montrer le lemme suivant

   Lemme (substitution). *Soit $G = (V, \Sigma, P, S)$ une GHC, $\pi = A \to \alpha_1 B \alpha_2$ une production de P avec $B \to \beta_1 | \dots | \beta_r$ la règle*[2] *de P avec B à gauche. Soit $G_1 = (V, \Sigma, P_1, S)$ avec*

   $$P_1 = (P - \{\pi\}) \cup \{A \to \alpha_1\beta_1\alpha_2 | \alpha_1\beta_2\alpha_2 | \ \dots \ | \alpha_1\beta_r\alpha_2\}$$

   alors $L(G_1) = L(G)$.

2. Montrer le lemme suivant

   Lemme (inversion). *Soit $G = (V, \Sigma, P, S)$ une GHC ne produisant pas le mot vide et*

   $$A \to A\alpha_1 | \ \dots | \ A\alpha_r$$

   avec $\alpha_i \neq \epsilon$ pour tout i, l'ensemble des règles avec A en tête telles que le symbole le plus à gauche du corps soit A. Soit

   $$A \to \beta_1 | \dots | \beta_s$$

   l'ensemble des règles restantes avec A en tête. Soit $G_1 = (V \cup \{Z\}, \Sigma, P_1, S)$, où Z est un nouveau symbole non-terminal et P_1 l'ensemble des productions de P avec toutes les production avec A en tête remplacées par

   $$\begin{aligned} A &\to \beta_1 Z | \ \dots \ | \ \beta_s Z | \ \beta_1 | \ \dots | \ \beta_s \\ Z &\to \alpha_1 Z | \ \dots \ | \alpha_r Z | \ \alpha_1 | \ \dots | \ \alpha_r \end{aligned}$$

   alors $L(G_1) = L(G)$.

Sans perte de généralité, on partira d'une grammaire $G = (V, \Sigma, P, S)$ sous forme normale de Chomsky. **Supposons pour le moment que G ne produit pas le mot vide**.

3. Soit $V = \{A_1, \dots, A_m\}$ avec $S = A_1$. L'ensemble des règles de production peut être classé en deux catégories

   $A_i \to BC$ avec $B, C \in V$
   $A_i \to a$

   Par induction sur les membres gauches des règles à partir de A_1, construisez une grammaire équivalente dont les règles de production peuvent être classées en trois catégories

   (a) $A_i \to A_j\alpha$ avec $i < j, \alpha \in ((V - \{S\}) \cup \{Z_1, \dots, Z_n\})^+$
   (b) $A_i \to a\alpha$ avec $a \ \in \Sigma, \alpha \in ((V - \{S\}) \cup \{Z_1, \dots, Z_n\})^*$
   (c) $Z_i \to \alpha$ avec $\alpha \in ((N - \{S\}) \cup \{Z_1, \dots, Z_n\}^+$

4. Toujours sous l'hypothèse que le mot vide n'appartient pas au langage, conclure.

5. Conclure si de plus ϵ est dans le langage.

6. En quoi la forme normale de Greibach est-elle intéressante ?

2. sous forme de Backus–Naur (BNF)

7. Mettre la grammaire suivante sous forme normale de Greibach

$$\begin{aligned} A_1 &\rightarrow A_2A_3 \\ A_2 &\rightarrow A_1A_2 |\, 1 \\ A_3 &\rightarrow A_1A_3|0 \end{aligned}$$

Définition. Soit $G = (V, \Sigma, P, S)$ une grammaire hors-contexte. La grammaire est dite *autoenchâssante*[3] s'il existe $X \in V$ tel que $X \overset{+}{\Rightarrow} \alpha X \beta$ avec $\alpha, \beta \in (V \cup \Sigma)^+$ $(\alpha, \beta \neq \epsilon)$.
Un langage hors-contexte L est dit *autoenchâssant* si toute grammaire hors-contexte le générant est autoenchâssante.

Exercice 3. *Autoenchâssements*

On souhaite montrer le théorème suivant :

Théorème. *Un langage hors-contexte est régulier si et seulement s'il n'est pas autoenchâssant.*

1. Montrer l'implication directe.

On veut montrer l'implication inverse. Soit donc un langage L hors-contexte non autoenchâssant. Soit $G = (V, \Sigma, P, S)$ une grammaire hors-contexte le générant telle que tout symbole non terminal est *accessible*[4] et sans règle de la forme $X \rightarrow Y$, $X, Y \in V$.

2. Supposons que pour tout X, il existe une dérivation $X \overset{\star}{\Rightarrow} \gamma S \gamma'$. Montrer qu'alors, G est nécessairement de type 3.

3. On suppose maintenant qu'il existe $X \in V$ tel que $X \overset{\star}{\not\Rightarrow} \gamma S \gamma'$. Montrer par induction sur $|V|$ que G est de type 3. On admettra[5] le résultat suivant : si L et L' sont deux langages régulier, avec a un terminal de L, alors le langage L'' constitué des mots de L dans lesquels on remplace occurence de a par un mot de L' est lui aussi régulier.

Définition. Une grammaire contextuelle $G = (V, \Sigma, P, S)$ est sous *forme normale de Kuroda* si toutes ses règles sont de la forme (i) $S \rightarrow SB$, (ii) $CD \rightarrow EF$, (iii) $G \rightarrow H$, (iv) $A \rightarrow a$ où $A, B, C, D, E, F, H \in V \setminus \{S\}$, $G \in V$, $A \in \Sigma$.

Exercice 4. *Forme normale de Kuroda*

✎ Montrer que toute grammaire contextuelle est équivalente à une grammaire sous forme normale de Kuroda. **Indication.** Partant d'une grammaire $G = (V, \Sigma, P, S)$ dont les productions sont soit de la forme $\alpha \rightarrow \beta$ avec $2 \geq |\beta| \geq |\alpha| \geq 0$ et $\alpha, \beta \in V^\star$ soit de la forme $A \rightarrow a$ avec $A \in V$ et $a \in \Sigma$, construire une nouvelle grammaire $G' = (V \cup \{S', Q\}, \Sigma, P', S')$ sous forme normale de Kuroda, où Q sert à compléter les règles $\alpha \rightarrow \beta$ qui ne sont d'aucune des formes (i) à (iii).

Exercice 5. *Langage miroir*

✎ Soit L un langage hors-contexte, et L^T son langage miroir : $L^T = \{x^T : x \in L\}$ où $(a_1 \ldots a_n)^T = a_n \ldots a_1$. Montrer que L^T est également hors-contexte.

---

3. Ou *self-embedding* en anglais.
4. Pour tout $X \in V$, il existe une dérivation $S \overset{\star}{\Rightarrow} \alpha X \beta$.
5. Ce résultat se prouve aisément à l'aide des automates finis.