TD9 : mise sous forme normale des grammaires algébriques

Soit $\mathcal{G} = \langle\ \mathcal{A}, \mathcal{V}, \mathcal{P}\ \rangle$ une grammaire algébrique avec $\mathcal{A}$ l'alphabet des symboles terminaux, $\mathcal{V}$ l'ensemble des variables (*i.e.* les symboles non-terminaux) et $\mathcal{P}$ les règles de production.

La grammaire $\mathcal{G}$ est *pré-réduite* si :

(1) aucun non-terminal n'engendre le langage vide : $\quad \forall T \in \mathcal{V}, L_{\mathcal{G}}(T) \neq \emptyset$.

La grammaire $\mathcal{G}$ est *réduite vis-à-vis d'un non-terminal* $T \in \mathcal{V}$ si elle est pré-réduite et si :

(2) tous les non-terminaux peuvent apparaître par dérivation à partir de T :

$$\forall S \in \mathcal{V}, \exists u_1, u_2 \in (\mathcal{A} \cup \mathcal{V})^* \text{ vérifiant } u_1 S u_2 \in \widehat{L}_{\mathcal{G}}(T)$$

Pour réduire $\mathcal{G}$, on peut définir :

(1) $U_0 = \mathcal{A}$, $U_i = U_{i-1} \cup \{S \in \mathcal{V} : \exists m \in U_{i-1}^*, S \longrightarrow m \in \mathcal{P}\}$ et $U = (\bigcup_{i\geq 0} U_i) \setminus \mathcal{A}$ l'ensemble des variables *productives* ;

(2) $W_0 = \{T\}$, $W_i = W_{i-1} \cup \{S \in \mathcal{V} : \exists S' \in W_{i-1}, \exists m_1, m_2 \in (\mathcal{A} \cup \mathcal{V})^*, S' \longrightarrow m_1 S m_2 \in \mathcal{P}\}$ et $W = \bigcup_{i\geq 0} W_i$ l'ensemble des variables *accessibles* à partir de T.

On obtient une grammaire pré-réduite $\mathcal{G}'$ (*resp.* réduite $\mathcal{G}''$) en supprimant dans $\mathcal{G}$ (*resp.* dans $\mathcal{G}'$) toutes les règles où apparaît une variable non productive (*resp.* non accessible).

1 Vérifier que les grammaires suivantes sont réduites (vis-à-vis de S) :

1. $\left\{\begin{array}{lcl} S & \rightarrow & ASB \mid \varepsilon \\ A & \rightarrow & aAS \mid a \\ B & \rightarrow & SbS \mid A \mid bb \end{array}\right.$

2. $\left\{\begin{array}{lcl} S & \rightarrow & 0A0 \mid 1B1 \mid BB \\ A & \rightarrow & C \\ B & \rightarrow & S \mid A \\ C & \rightarrow & S \mid \varepsilon \end{array}\right.$

3. $\left\{\begin{array}{lcl} S & \rightarrow & AAA \mid B \\ A & \rightarrow & aA \mid B \\ B & \rightarrow & \varepsilon \end{array}\right.$

La grammaire $\mathcal{G} = \langle\ \mathcal{A}, \mathcal{V}, \mathcal{P}\ \rangle$ est *propre* si :

(3) $\mathcal{G}$ ne contient aucune règle de la forme $S \rightarrow \varepsilon$ (ε-règle) ;

(4) $\mathcal{G}$ ne contient aucune règle de la forme $S \rightarrow T$ (règle unitaire).

On peut définir :

(3) $N_0 = \{S \in \mathcal{V} : S \rightarrow \varepsilon \in \mathcal{P}\}$, $N_i = N_{i-1} \cup \{S \in \mathcal{V} : \exists m \in N_{i-1}^*, S \longrightarrow m \in \mathcal{P}\}$ et $N = \bigcup_{i\geq 0} N_i$ l'ensemble des variables *annulables* ;

(4) la relation $\leq$ par $S \leq T \iff S \overset{*}{\longrightarrow} T$ et l'équivalence $\sim$ par $S \sim T \iff S \geq T \ \&\ T \geq S$.

Pour obtenir une grammaire propre à partir d'une grammaire réduite :

(3) pour chaque variable annulable S, on remplace, dans les membres droits, chaque occurrence de S par $S + \varepsilon$, puis on élimine les règles $S \longrightarrow \varepsilon$;

(4) pour chaque variable S maximale dans l'ordre quotient, on remplace chaque règle $T \rightarrow S$ par les règles $T \rightarrow m$ où m décrit les membres droits des règles issues de S, ce processus étant itéré avec les variables devenues maximales.

2 Réduire vis-à-vis de S_0, puis rendre propre les grammaires suivantes :

1. $\left\{\begin{array}{lcl} S_0 & \to & S_1S_2 \mid S_3S_4 \mid S_5 \\ S_1 & \to & S_1S_1 \mid S_1S_4 \\ S_2 & \to & aS_2 \mid S_3 \\ S_3 & \to & S_2 \mid S_4 \mid S_5aS_3 \\ S_4 & \to & cS_4 \mid \varepsilon \\ S_5 & \to & S_4 \mid b \end{array}\right.$
2. $\left\{\begin{array}{lcl} S_0 & \to & S_0S_1 \mid S_1 \mid S_2 \\ S_1 & \to & bS_1 \mid aS_3S_5 \mid \varepsilon \\ S_2 & \to & bS_2 \mid abS_4 \mid \varepsilon \\ S_3 & \to & bS_1 \mid S_2S_3 \\ S_4 & \to & aS_5S_2 \mid bS_1S_6S_7 \\ S_5 & \to & aS_5 \mid bS_5s_6 \\ S_6 & \to & aS_6b \mid aS_4 \\ S_7 & \to & S_5S_2 \mid aS_7S_6 \end{array}\right.$
3. $\left\{\begin{array}{lcl} S_0 & \to & S_1 \mid S_0aS_2 \mid bS_2S_3 \mid \varepsilon \\ S_1 & \to & S_1bS_0 \mid S_3aS_1 \\ S_2 & \to & S_0bS_0 \mid S_2S_3a \mid aS_1S_3 \\ S_3 & \to & S_1S_2 \mid S_3bS_2 \mid aS_0bS_3 \end{array}\right.$

La grammaire $\mathcal{G} = \langle\ \mathcal{A}, \mathcal{V}, \mathcal{P}\ \rangle$ est *sous forme normale de Chomsky* (ou *sous forme normale quadratique*) si on a $\mathcal{P} \subseteq \mathcal{V} \times \mathcal{V}\mathcal{V} \cup \mathcal{V} \times \mathcal{A}$.
À partir d'une grammaire $\mathcal{G}$ propre, décomposer les règles de productions ayant strictement plus de deux symboles : remplacer les règles du type $A \to u_1u_2\cdots u_k$ avec $u_i \in \mathcal{V} \cup \mathcal{A}$ et $k \geq 3$ par $A \to u_1A_1, A_1 \to u_2A_2, \ldots, A_{k-2} \to u_{k-1}u_k$ où les A_i sont de nouvelles variables.
On obtient alors une grammaire sous forme normale de Chomsky en introduisant une nouvelle variable T_a avec la règle $T_a \longrightarrow a$ pour chaque lettre terminale a et en remplaçant par T_a chaque occurrence de a dans tout membre droit de longueur 2.

3 Mettre les grammaires suivantes sous forme normale de Chomsky :

1. $\left\{\begin{array}{lcl} S & \to & aS \mid Sb \mid c \end{array}\right.$
2. $\left\{\begin{array}{lcl} S & \to & A \\ A & \to & b \mid c \end{array}\right.$
3. $\left\{\begin{array}{lcl} S & \to & abcdef \end{array}\right.$
4. $\left\{\begin{array}{lcl} S & \to & aS \mid Ab \\ A & \to & cB \\ B & \to & dA \mid e \end{array}\right.$
5. $\left\{\begin{array}{lcl} S & \to & aSAb \mid bSS \mid d \\ A & \to & cASaA \mid bcd \end{array}\right.$
6. $\left\{\begin{array}{lcl} S & \to & x \mid lSr \mid SpS \mid SmS \end{array}\right.$

La grammaire $\mathcal{G} = \langle\ \mathcal{A}, \mathcal{V}, \mathcal{P}\ \rangle$ est *sous forme normale de Greibach* (*resp. presque-Greibach*) si on a $\mathcal{P} \subseteq \mathcal{V} \times \mathcal{A}\mathcal{V}^*$ (*resp.* $\mathcal{P} \subseteq \mathcal{V} \times \mathcal{A}(\mathcal{A} \cup \mathcal{V})^*$).
Pour $\mathcal{V} = \{S_1, \ldots, S_r\}$, on pose $\mathcal{V}' = \{S'_1, \ldots, S'_r\}$. Pour $\mathcal{G}$ propre, on construit une suite $\mathcal{G} = \mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_{2r+1}$ de grammaires équivalentes (avec $\mathcal{G}_{2r+1}$ presque-Greibach) :

- $\mathcal{G}_{2i}$ s'obtient en remplaçant dans $\mathcal{G}_{2i-1}$ les règles $S_i \to S_im_1 \mid \cdots \mid S_im_k \mid w_1 \mid \cdots \mid w_p$ par

$$\left\{\begin{array}{lcl} S_i & \to & w_1S'_i \mid \cdots \mid w_pS'_i \mid w_1 \mid \cdots \mid w_p \\ S'_i & \to & m_1S'_i \mid \cdots \mid m_kS'_i \mid m_1 \mid \cdots \mid m_k \end{array}\right.$$

- $\mathcal{G}_{2i+1}$ s'obtient en remplaçant dans toute règle de $\mathcal{G}_{2i}$ de la forme $S \to S_jm$ avec $j \leq i$, cette occurrence de S_j par les membres droits issus de S_j dans $\mathcal{G}_{2i}$.

On obtient alors une grammaire sous forme normale de Greibach en introduisant une nouvelle variable T_a avec la règle $T_a \longrightarrow a$ pour chaque lettre terminale a et en remplaçant par T_a chaque occurrence de a non initiale d'un membre droit.

4 Mettre les grammaires suivantes sous forme normale de Greibach :

1. $\left\{\begin{array}{lcl} S_1 & \to & S_2S_3 \\ S_2 & \to & S_1S_2 \mid a \\ S_3 & \to & b \end{array}\right.$
2. $\left\{\begin{array}{lcl} S_1 & \to & S_2S_3 \\ S_2 & \to & S_1S_2 \mid a \\ S_3 & \to & S_3S_1 \mid b \end{array}\right.$
3. $\left\{\begin{array}{lcl} S & \to & ST \mid a \\ T & \to & TS \mid b \end{array}\right.$
4. $\left\{\begin{array}{lcl} S & \to & SaT \mid TT \mid b \\ T & \to & Td \mid TSa \mid aS \mid c \end{array}\right.$
5. $\left\{\begin{array}{lcl} S_1 & \to & S_1S_1 \mid S_1S_2 \mid S_3S_2 \mid S_2b \\ S_2 & \to & S_2S_1 \mid S_1S_2 \mid S_3S_3 \mid S_1a \\ S_3 & \to & S_1S_3 \mid S_2S_3 \mid S_3S_1 \mid a \end{array}\right.$