Informatique théorique III
Série 9 solution

Prof. Nestmann, 2004

1. Formes normales

Soit $G = (\{S, A, B, C, D, E\}, \{a, b\}, P, S)$ avec P tel que :

$$S \to aAa \mid bBb \mid \epsilon \qquad A \to C \mid a \qquad B \to C \mid b$$
$$C \to CDE \mid \epsilon \qquad D \to A \mid B \mid ab$$

1. Éliminer les $\epsilon-$productions,
2. puis éliminer les productions unitaires,
3. puis éliminer les symboles inutiles,
4. et mettre la grammaire résultante sous forme normale de Chomsky.
5. Comment auriez-vous pu vous simplifier la tâche ? □

Solution.

1. L'ensemble des symboles potentiellement vides est $\{S, A, B, C, D\}$. On obtient les productions suivantes :

$$S \to aAa \mid bBb \mid aa \mid bb \qquad A \to C \mid a \qquad B \to C \mid b$$
$$C \to CDE \mid DE \mid CE \mid E \qquad D \to A \mid B \mid ab$$

2. L'ensemble des paires unitaires des production précédentes est le suivant :

$$\{(S, S), (A, A), (A, C), (B, B), (B, C), (C, C), (D, D), (D, A), (D, B), (D, C), (E, E)\}$$

On obtient donc les productions suivantes :

$$S \to aAa \mid bBb \mid aa \mid bb \qquad A \to CDE \mid DE \mid CE \mid a$$
$$B \to CDE \mid DE \mid CE \mid b \qquad C \to CDE \mid DE \mid CE$$
$$D \to CDE \mid DE \mid CE \mid a \mid b \mid ab$$

3. L'ensemble des symboles génératifs des productions précédentes est $\{a, b, A, B, D, S\}$. On obtient donc les productions suivantes :

$$S \to aAa \mid bBb \mid aa \mid bb \qquad A \to a \qquad B \to b \qquad D \to a \mid b \mid ab$$

L'ensemble des symboles accessibles pour ces productions est $\{S, A, B\}$. Et l'on obtient finalement les productions suivantes :

$$S \to aAa \mid bBb \mid aa \mid bb \qquad A \to a \qquad B \to b$$

4. En jouant avec ces dernières productions, une grammaire FNC est donnée par $G' = (\{S, A, A', B, B'\}, \{a, b\}, P', S)$ avec P' tel que :

$$S \to AA \mid AA' \mid BB \mid BB' \quad A' \to AA \quad B' \to BB \quad A \to a \quad B \to b$$

5. En remarquant que la production $C \to CDE$ ne produit rien. Lorsque vous voyez directement qu'une production ne sert à rien, il est parfois utile de faire une première élimination des symboles inutiles avant de procéder aux trois

phases succesives. Pour G les symboles génératifs sont $\{a, b, A, B, C, D, S\}$ et les productions après élimination des non-génératifs sont :

$$S \to aAa \mid bBb \mid \epsilon \qquad A \to C \mid a \qquad B \to C \mid b$$
$$C \to \epsilon \qquad D \to A \mid B \mid ab$$

Les symboles accessibles sont $\{S, A, B, C\}$ en ne gardant que ceux-ci, la grammaire dont on part est alors :

$$S \to aAa \mid bBb \mid \epsilon \qquad A \to C \mid a \qquad B \to C \mid b \qquad C \to \epsilon$$

Cela simplifie notamment l'élimination des productions unitaires.

■

2. Forme anormale ?

Confirmer ou infirmer la proposition suivante par une preuve.

Proposition 2.1 *Tout langage non-contextuel L démuni du mot vide admet une grammaire $G = (V, \Sigma, P, S)$ telle que $\mathrm{L}(G) = L$ et dont ses productions sont uniquement de la forme*

1. *$A \to BCD$, avec $A, B, C, D \in V$ ou*
2. *$A \to a$, avec $A \in V, a \in \Sigma$.* □

Preuve. Prenons le langage $L = \{aa\}$. Supposons$^{(*)}$ qu'il existe une grammaire $G = (V, \Sigma, P, S)$ telle que $\mathrm{L}(G) = L$ et dont la forme des production est telle que donnée ci-dessus.

Dans ce cas, nous savons qu'il existe une dérivation la plus à gauche $S \Rightarrow^* aa$. Étant donné la forme des productions nous savons que la dérivation se termine par les dérivations directes suivantes $AB \Rightarrow aB \Rightarrow aa$ avec $A, B \in V$. Cependant la forme des productions ne nous permet pas de trouver l'indispensable dérivation $S \Rightarrow^* AB$. En effet, les productions introduisent soit exactement trois non-terminaux, soit exactement un terminal. Le mot aa n'est donc pas dérivable dans G et l'hypothèse $(*)$ est infirmée.

Il existe un langage non-contextuel ($\{aa\}$) qui n'admet pas de grammaire dont les productions sont de la forme ci-dessus. La proposition est donc fausse. ■

3. Langages non algébriques

Montrer que les langages suivants ne sont pas algébriques.

1. $L_1 = \{a^i \mid i \text{ est un nombre premier}\}$
2. $L_2 = \{a^i b^j c^k \mid i < j < k\}$ □

Solution. On utilise le lemme de gonflement.

1. $L_1 = \{a^i \mid i \text{ est un nombre premier}\}$

   On raisonne par l'absurde et on suppose que L_1 est algébrique. Soit alors n du lemme de gonflement. Soit p un nombre premier tel que $p \geq n$ et considérons alors le mot $z = a^p \in L_1$.

   Par définition de n, il existe $u, v, w, x, y \in \Sigma^*$ tel que
   - $z = uvwxy$
   - $vx \neq \epsilon$

- $|vwx| \leq n$
- pour tout $i \in \mathbb{N}$ on a $uv^iwx^iy \in L_1$

On a donc $a^p = z = uvwxy = a^{|u|}a^{|v|}a^{|w|}a^{|x|}a^{|y|}$ et donc $p = |u| + |v| + |w| + |x| + |y|$. De plus, posons $k = |v| + |x|$. Alors comme $vx \neq \epsilon$, on a $k > 0$.
On considère maintenant le mot $z' = uv^{p+1}wx^{p+1}y$. Par hypothèse, on a $z' \in L_1$.
Or $z' = a^{|u|}(a^{|v|})^{p+1}a^{|w|}(a^{|x|})^{p+1}a^{|y|} = a^{|u|+|v|+|w|+|x|+|y|+p\times(|v|+|x|)} = a^{p+p\times k} = a^{p\times(1+k)}$. Comme $k > 0$, $1+k > 1$, donc $p\times(1+k)$ n'est pas premier : cela contredit $z' \in L_1$.
Donc L_1 n'est pas algébrique.

2. $L_2 = \{a^ib^jc^k \mid i < j < k\}$
On raisonne par l'absurde et on suppose que L_2 est algébrique. Soit alors n du lemme de gonflement. Considérons alors le mot $z = a^nb^{n+1}c^{n+2} \in L_2$.
Par définition de n, il existe $u, v, w, x, y \in \Sigma^*$ tel que
   - $z = uvwxy$
   - $vx \neq \epsilon$
   - $|vwx| \leq n$
   - pour tout $i \in \mathbb{N}$ on a $uv^iwx^iy \in L_1$

   Comme $vx \neq \epsilon$, on a $v \neq \epsilon \vee x \neq \epsilon$.
De plus, comme $|vwx| \leq n$, on sait que vwx ne contient pas à la fois des a et des c (par construction de z) et donc vx ne contient pas à la fois des a et des c.
On doit alors examiner plusieurs cas :
   (a) si v ou x contient deux types de lettre différents (a et b ou b et c), alors en prenant $i = 2$, on arrive à une contradiction car on obtient un mot qui alterne ces deux lettres.
   (b) si vx ne contient que des a, en prenant $i = 2$, on arrive à une contradiction car on aura plus de a que de b.
   (c) si vx ne contient que des b, on prend $i = 2$ et on aboutit à une contradiction car on aura plus de b que de c.
   (d) si vx ne contient que des c, on prend $i = 0$ et on arrive à une contradiction car on aura moins de c que de b.
   (e) si v ne contient que des a et x que des b, on prend $i = 2$ et on arrive à une contradiction. En effet,
      - si $x \neq \epsilon$, on aura plus de b que de c (on aura ajouté au moins un b).
      - si $v \neq \epsilon$ et $x = \epsilon$, on aura plus de a que de b (on aura ajouté au moins un a et aucun b).
   (f) si v ne contient que des b et x que des c, on prend $i = 0$ et on arrive à une contradiction. En effet,
      - si $v \neq \epsilon$, on aura plus de a que de b (on aura retiré au moins un b).
      - si $x \neq \epsilon$ et $v = \epsilon$, on aura plus de b que de c (on aura retiré au moins un c et aucun b).

   Dans tous les cas, on est arrivé à une contradiction donc L_2 n'est pas algébrique.

■

4. Forme normale de Chomsky (bis repetita)

Dans la série 8, on a obtenu la grammaire $G = (V, \Sigma, P, S)$ avec les symboles non-terminaux :

$$\begin{aligned} V = \{S, &[q_0aq_0], [q_0bq_0], [q_0 \bullet q_0], \\ &[q_0aq_1], [q_0bq_1], [q_0 \bullet q_1], \\ &[q_1aq_1], [q_1bq_1], [q_1 \bullet q_1]\} \end{aligned}$$

et avec les règles de productions :

$$
\begin{array}{lll}
S \to [q_0 \bullet q_0] & S \to [q_0 \bullet q_1] & \\
[q_0 \bullet q_0] \to a[q_0 a q_0][q_0 \bullet q_0] & [q_0 \bullet q_0] \to b[q_0 b q_0][q_0 \bullet q_0] & [q_0 a q_0] \to a[q_0 a q_0][q_0 a q_0] \\
[q_0 \bullet q_1] \to a[q_0 a q_0][q_0 \bullet q_1] & [q_0 \bullet q_1] \to b[q_0 b q_0][q_0 \bullet q_1] & [q_0 a q_1] \to a[q_0 a q_0][q_0 a q_1] \\
[q_0 \bullet q_1] \to a[q_0 a q_1][q_1 \bullet q_1] & [q_0 \bullet q_1] \to b[q_0 b q_1][q_1 \bullet q_1] & [q_0 a q_1] \to a[q_0 a q_1][q_1 a q_1] \\
[q_0 a q_0] \to b[q_0 b q_0][q_0 a q_0] & [q_0 b q_0] \to a[q_0 a q_0][q_0 b q_0] & [q_0 b q_0] \to b[q_0 b q_0][q_0 b q_0] \\
[q_0 a q_1] \to b[q_0 b q_0][q_0 a q_1] & [q_0 b q_1] \to a[q_0 a q_0][q_0 b q_1] & [q_0 b q_1] \to b[q_0 b q_0][q_0 b q_1] \\
[q_0 a q_1] \to b[q_0 b q_1][q_1 a q_1] & [q_0 b q_1] \to a[q_0 a q_1][q_1 b q_1] & [q_0 b q_1] \to b[q_0 b q_1][q_1 b q_1] \\
[q_0 \bullet q_1] \to [q_1 \bullet q_1] & [q_0 a q_1] \to [q_1 a q_1] & [q_0 b q_1] \to [q_1 b q_1] \\
[q_1 a q_1] \to a & [q_1 b q_1] \to b & [q_1 \bullet q_1] \to \epsilon
\end{array}
$$

Transformer cette grammaire afin de trouver une grammaire sous forme normale de Chomsky générant $\mathrm{L}(G) \setminus \{\epsilon\}$.

Solution.

1. On élimine les ϵ-productions.
   L'ensemble des symboles potentiellement vide est $\{S, [q_1 \bullet q_1], [q_0 \bullet q_1]\}$.
2. On élimine ensuite les productions unitaires.
   On a $U_G = \left\{ \begin{array}{l} (S, [q_0 \bullet q_0]), (S, [q_0 \bullet q_1]), (S, [q_1 \bullet q_1]), \\ ([q_0 \bullet q_1], [q_1 \bullet q_1]), ([q_0 b q_1], [q_1 b q_1]), ([q_0 a q_1], [q_1 a q_1]) \end{array} \right\} \cup \mathrm{Id}_V$
3. On élimine ensuite les symboles inutiles.
   – Les symboles génératifs sont $\{a, b, S, [q_0 \bullet q_1], [q_0 a q_1], [q_1 a q_1], [q_0 b q_1], [q_1 b q_1]\}$.
   – Les symboles accessibles sont $\{a, b, S, [q_1 a q_1], [q_1 b q_1], [q_0 b q_1], [q_0 a q_1]\}$.

Après avoir appliqué tous ces calculs dans cet ordre-ci,
on obtient $G' = (\{S, [q_1 a q_1], [q_1 b q_1], [q_0 b q_1], [q_0 a q_1]\}, \{a, b\}, P, S)$ avec P' défini par

$$
\begin{aligned}
S &\to a[q_0 a q_1] \mid b[q_0 b q_1] \\
[q_0 a q_1] &\to a \mid a[q_0 a q_1][q_1 a q_1] \mid b[q_0 b q_1][q_1 a q_1] \\
[q_0 b q_1] &\to b \mid a[q_0 a q_1][q_1 b q_1] \mid b[q_0 b q_1][q_1 b q_1] \\
[q_1 a q_1] &\to a \\
[q_1 b q_1] &\to b
\end{aligned}
$$

En renommant les non-terminaux, en éliminant $[q_1 a q_1]$ et $[q_1 b q_1]$, et en factorisant par S, on obtient $G'' = (\{S, C, D\}, \{a, b\}, P'', S)$ avec P'' défini par :

$$
\begin{aligned}
S &\to aC \mid bD \\
C &\to a \mid Sa \\
D &\to b \mid Sb
\end{aligned}
$$

En introduisant alors deux nouveaux non-terminaux A et B, on obtient finalement $G''' = (\{S, A, B, C, D\}, \{a, b\}, P''', S)$ avec P''' défini par :

$$
\begin{aligned}
S &\to AC \mid BD \\
A &\to a \\
B &\to b \\
C &\to a \mid SA \\
D &\to b \mid SB
\end{aligned}
$$

qui est sous forme normale de Chomsky. ∎