Informatique théorique III
(Automates, langages & calculabilité)

Formulaire du Cours 2004-2005
Partie 9
EPFL – I&C

Uwe Nestmann
Sébastien Briais
Daniel C. Bünzli

14 février 2005

Bibliographie

[HMU03] John Hopcroft, Rajeev Motwani, and Jeffrey Ullman. *Introduction to Automata Theory, Languages and Computation*. Pearson Education International, 2003. ISBN 0321210298.

Table des matières

$Id: notes-9.tex,v 1.13 2004/12/13 16:20:38 uwe Exp $

semaine 9

Voir [HMU03] §7.1 et §7.2.

3.5 Formes normales des grammaires non-contextuelles

3.5.1 Définition (Symbole potentiellement vide)
Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle.
Un symbole $A \in V$ est *potentiellement vide*, si $A \Rightarrow_G^* \epsilon$.

3.5.2 Lemme Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle.
Soit $V_\epsilon \subseteq V$ défini de manière inductive par :

$$\frac{(A \to \epsilon) \in P}{A \in V_\epsilon} \qquad \frac{(A \to B_1 B_2 \cdots B_k) \in P \qquad \forall i \in [1, k] \,.\, B_i \in V_\epsilon}{A \in V_\epsilon}$$

Alors A est potentiellement vide ssi $A \in V_\epsilon$.

3.5.3 Lemme (Elimination des ϵ-productions)
Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle, et $\Gamma = V \cup \Sigma$.
Soit V_ϵ l'ensemble des symboles potentiellement vides.
Soit $G' = (V, \Sigma, P', S)$ avec P' défini de manière inductive par :

$$\frac{(A \to \alpha) \in P \qquad \alpha \in \Gamma^+}{(A \to \alpha) \in P'} \qquad \frac{(A \to \gamma B \gamma') \in P' \qquad B \in V_\epsilon \qquad \gamma\gamma' \in \Gamma^+}{(A \to \gamma\gamma') \in P'}$$

Alors $\mathrm{L}(G') = \mathrm{L}(G) \setminus \{\epsilon\}$.

3.5.4 Définition (Production unitaire)
Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle.
Une production $(A \to \alpha) \in P$ est dite *unitaire* si $\alpha \in V$.
Deux variables $A, B \in V$ forment une *paire unitaire* (A, B) dans G, s'il existe une dérivation $A \Rightarrow_G^* B$ qui utilise uniquement des productions unitaires.

3.5.5 Lemme Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle.
Soit $\mathrm{U}_G \subseteq V \times V$ défini de manière inductive par :

$$\frac{A \in V}{(A, A) \in \mathrm{U}_G} \qquad \frac{(A, B) \in \mathrm{U}_G \qquad (B \to C) \in P}{(A, C) \in \mathrm{U}_G}$$

Alors, (A, B) est une paire unitaire ssi $(A, B) \in \mathrm{U}_G$.

3.5.6 Lemme (Élimination des productions unitaires)
Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle, et U_G l'ensemble des paires unitaires de G. Soit $G' = (V, \Sigma, P', S)$ la grammaire dont les productions P' sont telles que :

$$\frac{(A, B) \in \mathrm{U}_G \qquad (B \to \alpha) \in P \qquad \alpha \notin V}{(A \to \alpha) \in P'}$$

Alors $\mathrm{L}(G') = \mathrm{L}(G)$.

3.5.7 Définition (Symboles utiles)
Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle et $\Gamma = V \cup \Sigma$.

1. Un symbole $X \in \Gamma$ est *génératif*
   s'il existe $w \in \Sigma^*$ tel que $X \Rightarrow_G^* w$.
2. Un symbole $X \in \Gamma$ est *accessible*
   s'il existe $\alpha, \beta \in \Gamma^*$ tel que $S \Rightarrow_G^* \alpha X \beta$.
3. Un symbole $X \in \Gamma$ est *utile* s'il est génératif et accessible.
   Sinon il est *inutile*.

3.5.8 Lemme
Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle, et $\Gamma = V \cup \Sigma$.
Soit $\Gamma_{\text{gen}} \subseteq \Gamma$ défini de manière inductive par :

$$\frac{s \in \Sigma}{s \in \Gamma_{\text{gen}}} \qquad \frac{(A \rightarrow B_1 B_2 \cdots B_k) \in P \qquad \forall i \in [1, k] \,.\, B_i \in \Gamma_{\text{gen}}}{A \in \Gamma_{\text{gen}}}$$

Soit $\Gamma_{\text{acc}} \subseteq \Gamma$ défini de manière inductive par :

$$\frac{}{S \in \Gamma_{\text{acc}}} \qquad \frac{A \in \Gamma_{\text{acc}} \qquad (A \rightarrow B_1 B_2 \cdots B_k) \in P \qquad i \in [1, k]}{B_i \in \Gamma_{\text{acc}}}$$

Alors

1. $X \in \Gamma$ est *génératif* ssi $X \in \Gamma_{\text{gen}}$.
2. $X \in \Gamma$ est *accessible* ssi $X \in \Gamma_{\text{acc}}$.

Noter que, dans le calcul des symboles génératifs, le cas $k = 0$ traite les productions de la forme $A \rightarrow \epsilon$.

3.5.9 Lemme Soit $G = (V, \Sigma, P, S)$ une grammaire non-contextuelle et Γ_{gen} tel que défini au lemme 3.5.8. Si $\mathrm{L}(G) = \emptyset$, alors $S \notin \Gamma_{\text{gen}}$.

3.5.10 Lemme (Élimination des symboles inutiles)
Soit $G \triangleq (V, \Sigma, P, S)$ une grammaire non-contextuelle
telle que $\mathrm{L}(G) \neq \emptyset$ et Γ_{gen} tel que défini au lemme 3.5.8.
Soit $G' \triangleq (V', \Sigma, P', S)$ avec $V' \triangleq V \cap \Gamma_{\text{gen}}$, $\Gamma' \triangleq V' \cup \Sigma$, et P' tel que :

$$\frac{(A \rightarrow \alpha) \in P \qquad A \in V' \qquad \alpha \in \Gamma'^*}{(A \rightarrow \alpha) \in P'}$$

Soit Γ'_{acc} tel que défini (pour G') au lemme 3.5.8, et $G'' \triangleq (V'', \Sigma'', P'', S)$ avec $V'' \triangleq V' \cap \Gamma'_{\text{acc}}$, $\Sigma'' \triangleq \Sigma \cap \Gamma'_{\text{acc}}$, $\Gamma'' \triangleq V'' \cup \Sigma''$ et P'' tel que :

$$\frac{(A \rightarrow \alpha) \in P' \qquad A \in V'' \qquad \alpha \in \Gamma''^*}{(A \rightarrow \alpha) \in P''}$$

Alors :

1. G'' ne contient que de symboles utiles.
2. $\mathrm{L}(G) = \mathrm{L}(G') = \mathrm{L}(G'')$.

3.5.11 Proposition Soit G une grammaire non-contextuelle.
Soit G' la grammaire résultant de (dans cet ordre) :

1. l'élimination des ϵ-productions selon le lemme 3.5.3,
2. puis de l'élimination des productions unitaires selon le lemme 3.5.6,
3. et de l'élimination des symboles inutiles selon le lemme 3.5.10.

Alors, G' ne contient aucune ϵ-production, aucune production unitaire, et aucun symboles inutile.

Noter que si les trois transformations ne sont pas appliquées dans l'ordre donné alors la proposition n'est pas garantie.

3.5.12 Définition (Forme normale de Chomsky)
Soit $G = (V, \Sigma, P, S)$ une grammaire.
G est sous *forme normale de Chomsky* (FNC) si :

1. $V \cup \Sigma$ ne contient pas de symboles inutiles.
2. Les productions de P sont de l'une des formes suivantes :
   (a) $A \to BC$ avec $A, B, C \in V$
   (b) $A \to a$ avec $A \in V$ et $a \in \Sigma$.

3.5.13 Théorème (Chomsky)
Soit G une grammaire non-contextuelle telle que $\mathrm{L}(G) \setminus \{\epsilon\} \neq \emptyset$.
Alors, il existe une grammaire G' en FNC telle que $\mathrm{L}(G') = \mathrm{L}(G) \setminus \{\epsilon\}$.

3.6 Lemme de gonflement

3.6.1 Définition La *hauteur* $\mathrm{h}(t)$ d'un arbre $t \in \mathcal{T}_N$ est défini par :

$$\begin{aligned} \mathrm{h}\big((n)\big) &= 0 \\ \mathrm{h}\big((n\, t_1 \ldots t_k)\big) &= 1 + \max\{\, \mathrm{h}(t) \mid t \in \{\, t_1 \ldots t_k) \,\} \,\} \end{aligned}$$

3.6.2 Lemme Soit $G = (V, \Sigma, P, S)$ une grammaire en FNC.
Soit (t, ψ) un arbre de dérivation pour $w \in \Sigma$, et $n = \mathrm{h}(t)$.
Alors, $|w| \leq 2^{n-1}$.

3.6.3 Théorème Soit Σ un alphabet et L un langage sur Σ.
Si L est non-contextuel, alors :

$$\exists n \in \mathbb{N} : \forall z \in L : |z| \geq n \Rightarrow \left(\begin{array}{ll} \exists u, v, w, x, y \in \Sigma^* : & z = uvwxy \\ & \wedge\, vx \neq \epsilon \\ & \wedge\, |vwx| \leq n \\ & \wedge\, \forall i \in \mathbb{N} : uv^iwx^iy \in L \end{array} \right)$$