Informatique théorique III
Série 10 solution

Prof. Nestmann, 2004

1. Intersection de deux langages non-contextuels

Soit $\Sigma = \{a, b, c\}$. Soit L_1 et L_2 les langages sur Σ définis par :

$$L_1 = \left\{a^n b^{2n} c^m \mid n, m \in \mathbb{N}\right\}$$
$$L_2 = \left\{a^n b^m c^{2m} \mid n, m \in \mathbb{N}\right\}$$

1. Montrer que L_1 et L_2 sont non-contextuels en donnant deux grammaires non-contextuelles G_1 et G_2 telles que $\mathrm{L}(G_1) = L_1$ et $\mathrm{L}(G_2) = L_2$.
2. Montrer que $L_1 \cap L_2$ n'est pas non-contextuel en utilisant les propriétés de stabilité.

Solution.

1. La grammaire $G_1 = (\{S, S', C\}, \{a, b, c\}, P_1, S)$ avec P_1 défini par :

$$S \to S'C$$
$$S' \to aS'bb \mid \epsilon$$
$$C \to cC \mid \epsilon$$

   est une grammaire non-contextuelle telle que $\mathrm{L}(G_1) = L_1$ donc L_1 est non-contextuel.
   La grammaire $G_2 = (\{S, S', A\}, \{a, b, c\}, P_2, S)$ avec P_2 défini par :

$$S \to AS'$$
$$A \to aA \mid \epsilon$$
$$S' \to bS'cc \mid \epsilon$$

   est une grammaire non-contextuelle telle que $\mathrm{L}(G_2) = L_2$ donc L_2 est non-contextuel.
2. Clairement $L_1 \cap L_2 = \left\{a^n b^{2n} c^{4n} \mid n \in \mathbb{N}\right\}$.
   Soit $h : \Sigma^* \to \Sigma^*$ l'homomorphisme de mots défini par $h(a) = a$, $h(b) = bb$ et $h(c) = cccc$.
   On a alors $h^R(L_1 \cap L_2) = \{a^n b^n c^n \mid n \in \mathbb{N}\}$.
   Par l'absurde, si $L_1 \cap L_2$ était non-contextuel, on pourrait donc conclure que $\{a^n b^n c^n \mid n \in \mathbb{N}\}$ est non-contextuel : ce qui n'est pas. Donc $L_1 \cap L_2$ n'est pas non-contextuel.

■

2. Propriété de stabilité

Si L est un langage sur un alphabet Σ, on définit $\mathrm{init}(L)$ par :

$$\mathrm{init}(L) = \{x \in \Sigma^* \mid \exists y \in \Sigma^* . xy \in L\}$$

Montrer que si L est non-contextuel alors $\mathrm{init}(L)$ est non-contextuel.

On pourra partir d'une grammaire G sous forme normale de Chomsky telle que $\mathrm{L}(G) = L \backslash \{\epsilon\}$ et donner une grammaire non-contextuelle ayant pour langage $\mathrm{init}(L)$.

Énoncer ensuite les résultats intermédiaires qu'il faudrait prouver pour démontrer que la grammaire donnée convient et indiquer rapidement comment prouver ces résultats.

Solution. Si L est vide, c'est trivial. Supposons maintenant L non vide.

Comme L est un langage non-contextuel, il existe une grammaire G_L non-contextuelle ayant pour langage L.

Si $L = \{\epsilon\}$ alors $\mathrm{init}(L) = \{\epsilon\}$ est non-contextuel.

Sinon, il existe une grammaire $G = (V, \Sigma, P, S)$ sous forme normale de Chomsky ayant pour langage $L \setminus \{\epsilon\}$.

On va transformer G en G' de telle sorte que $\mathrm{L}(G') = \mathrm{init}(\mathrm{L}(G))$. L'idée est d'ajouter pour chaque non-terminal $A \in V$, un nouveau non-terminal A' qui va représenter le langage des préfixes du langage de A.

On définit donc $G' = (V', \Sigma, P', S')$ avec $V' = V \cup \{A' \mid A \in V\}$ et P' défini par :

$$
\begin{array}{ll}
A \to_{G'} BC & \text{si } A \to_G BC \in P \\
A' \to_{G'} B' \mid BC' & \text{si } A \to_G BC \in P \\
A \to_{G'} a & \text{si } A \to_G a \in P \\
A' \to_{G'} a \mid \epsilon & \text{si } A \to_G a \in P
\end{array}
$$

Autrement dit, P' contient toutes les productions de P plus des variantes avec des primes sur les non-terminaux. En effet, si $A \to_G BC$ est une production, alors les préfixes du langage de A sont soit des préfixes du langage de B ($A' \to_{G'} B'$), soit des mots du langage de B concaténés à des préfixes du langage de C ($A' \to_{G'} BC'$).

Nous démontrons maintenant que $\mathrm{L}(G') = \mathrm{init}(L)$. Pour cela, nous avons besoin de quatre résultats :

1. toute dérivation de G est aussi une dérivation de G' ;
2. toute dérivation de G' partant d'un mot de $(V \cup \Sigma)^*$ (c'est-à-dire où les non-terminaux primés que l'on a ajouté à G' n'apparaissent pas) est une dérivation de G ;
3. pour tout non terminal $A \in V$, si $A' \Rightarrow^*_{G'} w$ avec $w \in \Sigma^*$ alors il existe $\gamma \in (V \cup \Sigma)^*$ tel que $A \Rightarrow^*_G w\gamma$;
4. pour tout non terminal $A \in V$, si $A \Rightarrow^*_G w$ alors pour tout $u \in \mathrm{init}(\{w\})$, on a $A' \Rightarrow^*_{G'} u$.

Preuve. Les grammaires étant non-contextuelles, on utilise la notion de dérivation le plus à gauche.

1. c'est évident car les productions de G sont aussi des productions de G'.

2. c'est évident car les productions de G' réécrivent les non-terminaux $A \in V$ en un mot de $(V \cup \Sigma)^*$.

3. on prouve ce résultat par récurrence sur la longueur n de la dérivation $A' \Rightarrow_{G'}{}^n\ w$
   – Si $n = 0$, la dérivation $A \Rightarrow_{G'} w$ n'est pas possible donc la propriété est vraie car le faux implique n'importe quoi.
   – Soit $n \in \mathbb{N}$ et supposons le résultat vrai pour toutes les dérivations de longueur inférieure ou égale à n. Supposons alors que $A' \Rightarrow_{G'}{}^{n+1} w$. On regarde alors la première étape de cette dérivation le plus à gauche :
     – Si $A' \Rightarrow_{G'} a \Rightarrow_{G'}{}^n\ w$ alors nécessairement $n = 0$ et $w = a$. De plus, cela signifie que $A \to_G a \in P$. En prenant $\gamma = \epsilon$, on a donc $A \Rightarrow_G a = a\gamma$.
     – Si $A' \Rightarrow_{G'} \epsilon \Rightarrow_{G'}{}^n\ w$ alors comme ci-dessus on a $n = 0$ et $A \to_G a \in P$. En prenant $\gamma = a$, on a le résultat.

- Si $A' \Rightarrow_{G'} B' \Rightarrow_{G'}{}^{n} w$ alors on a $A \rightarrow_G BC \in P$.
  Par hypothèse d'induction (appliquée à $B' \Rightarrow_{G'}{}^{n} w$), il existe $\gamma \in (V \cup \Sigma)^*$ tel que $B \Rightarrow_G^* w\gamma$.
  En prenant $\gamma' = \gamma C$, on a $A \Rightarrow_G BC \Rightarrow_G^* w\gamma C = w\gamma'$, ce qui est le résultat voulu au rang $n+1$.
- Si $A' \Rightarrow_{G'} BC' \Rightarrow_{G'}{}^{n} w$ alors on a $A \rightarrow_G BC \in P$.
  Il est clair qu'il existe $u, v \in \Sigma^*$ tel que $w = uv$, $B \Rightarrow_{G'}{}^{k} u$ et $C' \Rightarrow_{G'}{}^{l} v$ avec $k + l = n$ (dans la dérivation le plus à gauche, on réduit d'abord B puis C').
  Par hypothèse d'induction (appliquée à $C' \Rightarrow_{G'}{}^{l} v$ où $l \leq n$), il existe $\gamma \in (V \cup \Sigma)^*$ tel que $C \Rightarrow_G^* v\gamma$.
  D'après le point (2), on sait aussi que $B \Rightarrow_{G'}{}^{k} u$ est aussi une réduction de G (on a donc $B \Rightarrow_G^* u$).
  En recollant les morceaux, on obtient $A \Rightarrow_G BC \Rightarrow_G^* uC \Rightarrow_G^* uv\gamma = w\gamma$. D'où le résultat au rang $n+1$.

4. ici encore, on prouve ce résultat par récurrence sur la longueur n de la dérivation $A \Rightarrow_G{}^{n} w$
   - Si $n = 0$, alors la dérivation $A \Rightarrow_G{}^{n} w$ n'est pas possible donc la propriété est vraie car le faux implique n'importe quoi.
   - Soit $n \in \mathbb{N}$ et supposons le résultat vrai pour toutes les dérivations de longueur inférieure ou égale à n. Supposons maintenant que $A \Rightarrow_G{}^{n+1} w$. On regarde alors la première étape de cette dérivation le plus à gauche :
     - Si $A \Rightarrow_G a \Rightarrow_G{}^{n} w$ alors nécessairement $n = 0$ et $w = a$ et aussi $A \rightarrow_G a \in P$. On a $\mathrm{init}(\{a\}) = \{\epsilon, a\}$. Comme $A \rightarrow_G a \in P$, par construction de G' on a $A' \rightarrow_{G'} a \mid \epsilon \in P'$ et donc pour tout mot u de $\mathrm{init}(\{a\})$ on a $A' \Rightarrow_{G'}^* u$.
     - Sinon, si $A \Rightarrow_G BC \Rightarrow_G{}^{n} w$ alors on a $A \rightarrow_G BC \in P$.
       En dérivant le plus à gauche, il existe $u, v \in \Sigma^*$ tels que $w = uv$, $B \Rightarrow_G{}^{k} u$ et $C \Rightarrow_G{}^{l} u$ avec $k + l = n$.
       Soit $w' \in \mathrm{init}(\{w\}) = \mathrm{init}(\{uv\})$.
       Alors on a soit $w' \in \mathrm{init}(\{u\})$, soit $w' \in \{u\}\,\mathrm{init}(\{v\})$.
       (a) premier cas : $w' \in \mathrm{init}(\{u\})$
       Comme on a $B \Rightarrow_G{}^{k} u$ avec $k \leq n$, l'hypothèse d'induction nous donne $B' \Rightarrow_{G'}^* w'$.
       De plus, comme $A \rightarrow_G BC \in P$, on a $A' \rightarrow_{G'} B' \in P'$.
       On a donc la dérivation $A' \Rightarrow_{G'} B' \Rightarrow_{G'}^* w'$, d'où le résultat au rang $n+1$.
       (b) second cas : $w' \in \{u\}\,\mathrm{init}(\{v\})$
       Dans ce cas, il existe $v' \in \mathrm{init}(\{v\})$ tel que $w' = uv'$.
       Comme on a $C \Rightarrow_{G'}{}^{l} v$ avec $l \leq n$, l'hypothèse d'induction nous donne $C' \Rightarrow_{G'}^* v'$.
       De plus, d'après (1), la dérivation $B \Rightarrow_G{}^{k} u$ est aussi une dérivation de G' : on a donc $B \Rightarrow_{G'}^* u$.
       Comme $A \rightarrow_G BC \in P$, on a par construction $A' \rightarrow_{G'} BC' \in P'$.
       On a donc la dérivation $A \Rightarrow_{G'} BC' \Rightarrow_{G'}^* uC \Rightarrow_{G'}^* uv' = w'$, d'où le résultat au rang $n+1$.

■

Il est alors facile de conclure que $\mathrm{L}(G') = \mathrm{init}(\mathrm{L}(G)) = \mathrm{init}(L)$.

En effet, le point (3) dit en particulier que pour tout mot $w \in \mathrm{L}(G')$, il existe γ tel que $S \Rightarrow_G^* w\gamma$. Comme G ne contient que des symboles utiles, on sait alors qu'il existe $w' \in \Sigma^*$ tel que $\gamma \Rightarrow_G^* w'$. Autrement dit $ww' \in \mathrm{L}(G)$ ce qui prouve que $w \in \mathrm{init}(\mathrm{L}(G))$. On a donc $\mathrm{L}(G') \subseteq \mathrm{init}(\mathrm{L}(G))$.

Pour la réciproque, le point (4) nous donne en particulier que si $w \in \mathrm{L}(G)$ alors tout préfixe de w est un mot de $\mathrm{L}(G')$ et donc $\mathrm{init}(\mathrm{L}(G)) \subseteq \mathrm{L}(G')$.

Il est de plus clair que quel que soit L non vide, non réduit à ϵ, on a $\mathrm{init}(L) = \mathrm{init}(L \setminus \{\epsilon\})$.

Donc $\mathrm{L}(G') = \mathrm{init}(L)$ est non-contextuel. ■