

Lemme de pompage (hors programme)

MPI/MPI*, lycée Faidherbe

I Formes normales

I.1 Grammaires réduites

Pour toute grammaire $G = (\Sigma, V, P, S)$ et pour $X \in V$, on note $G(X) = (\Sigma, V, P, X)$; c'est la grammaire obtenue en changeant la variable de départ.

On note alors $L_G(X)$ le langage engendré par $G(X)$. En particulier $L(G) = L_G(S)$.

Une grammaire $G(\Sigma, V, P, S)$ est *réduite* si, pour toute variable $X \in V$, $L_G(X) \neq \emptyset$ et il existe une dérivation $S \xRightarrow{*} \alpha X \beta$ avec $\alpha, \beta \in (\Sigma \cup V)^*$. Une grammaire est donc réduite si toutes les variables sont utiles, elle interviennent dans au moins une dérivation possible d'un mot de $L(G)$.

Exercice 1

Solution page 3

Prouver que tout langage algébrique non vide est engendré par une grammaire réduite.

On peut formuler le résultat sous la forme

"Toute grammaire est faiblement équivalente à une grammaire réduite."

I.2 Grammaires propres

Dans une grammaire $G = (\Sigma, V, P, S)$, une variable X est annulable si $X \xRightarrow{*} \varepsilon$.

Exercice 2

Solution page 3

Prouver que, pour tout langage algébrique L , il existe une grammaire $G = (\Sigma, V, P, S)$ qui ne contient aucune règle de la forme $X \rightarrow \varepsilon$ qui engendre $L \setminus \{\varepsilon\}$.

Un grammaire $G = (\Sigma, V, P, S)$ est *propre* si elle ne contient aucune règle de la forme $X \rightarrow \varepsilon$ ou $X \rightarrow Y$ avec $Y \in V$.

Exercice 3

Solution page 3

Prouver que, si L est algébrique, il existe une grammaire propre qui engendre $L \setminus \{\varepsilon\}$.

Pour ajouter ε , il suffit d'ajouter la règle $S \rightarrow \varepsilon$.

I.3 Forme normale de Chomsky

Exercice 4

Solution page 3

Prouver que tout langage L ne contenant pas ε est engendré par une grammaire pour laquelle les règles de production sont de la forme

- $X \rightarrow X_1 X_2 \cdots X_p$ avec $X_i \in V$ pour tout i ,
- $X \rightarrow x$ avec $x \in \Sigma$,

Une grammaire en forme normale de Chomsky ou grammaire quadratique est une grammaire (Σ, V, P, S) dont les règles de production sont de la forme

- $X \rightarrow YZ$ avec $Y, Z \in V \setminus \{S\}$,
- $X \rightarrow x$ avec $x \in \Sigma$,

Exercice 5**Solution page 3**

Prouver que tout langage L ne contenant pas ε est engendré par une grammaire en forme normale de Chomsky.

Exercice 6**Solution page 4**

Montrer que l'intersection d'un langage algébrique et d'un langage rationnel est algébrique.

On pourra considérer une grammaire de Chomsky $G = (\Sigma, V, P, S)$ qui engendre le premier langage et un automate déterministe $Q = (\Sigma, \mathbb{S}, \delta, s_0, T)$ qui reconnaît le second et créer une grammaire de Chomsky dont les variables sont de la forme X_s^t avec $X \in V$ et $s, t \in \mathbb{S}$.

II Lemme de pompage

On va ici énoncer un résultat qui s'apparente au lemme de l'étoile. La principale différence est qu'il impose une itération simultanée en deux endroits d'un mot généré.

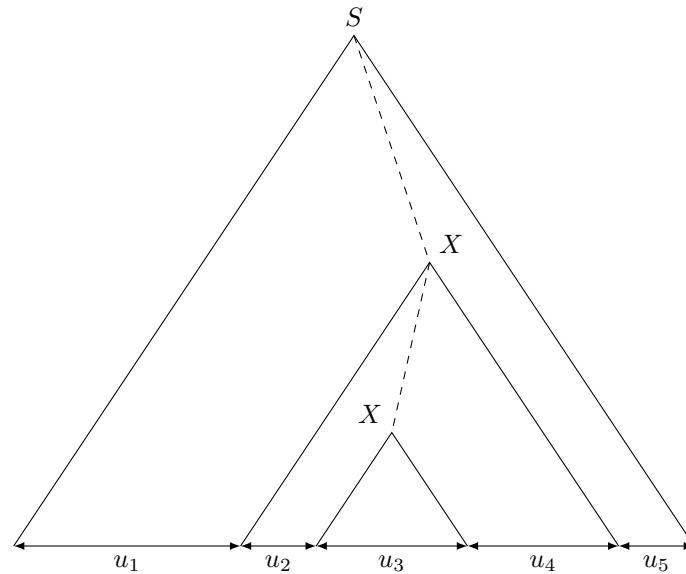
Exercice 7**Solution page 4**

Prouver que si $G = (\Sigma, V, P, S)$ est une grammaire en forme normale de Chomsky et si $u \in L(G)$ admet un arbre de dérivation de hauteur h alors $|u| \leq 2^{h-1}$.

Exercice 8**Solution page 4**

Prouver que si $G = (\Sigma, V, P, S)$ est une grammaire en forme normale de Chomsky et si $u \in L(G)$ vérifie $|u| \geq N = 2^{|V|}$ alors il existe une branche dans un arbre de dérivation de u contenant deux nœuds de même étiquette parmi les $|V| + 1$ derniers nœuds de la branche.

On a donc la situation suivante

**Exercice 9****Solution page 5**

Avec les notations ci-dessus, prouver qu'on a $|u_2u_3u_4| \leq N$ et $u_2u_4 \neq \varepsilon$.

Exercice 10 - Lemme de pompage**Solution page 5**

En déduire que si L est algébrique, il existe un entier N tel que tout mot $u \in L$ avec $|u| \geq N$ peut s'écrire $u = u_1u_2u_3u_4u_5$ avec $|u_2u_3u_4| \leq N$ et $u_2u_4 \neq \varepsilon$ et $u_1u_2^ku_3u_4^ku_5 \in L$ pour tout $k \in \mathbb{N}$.

Exercice 11 - Application**Solution page 6**

Prouver que $L_1 = \{a^n b^n c^n ; n \in \mathbb{N}\}$ et $L_2 = \{u \cdot v ; u, v \in \{a, b\}^*, |u| = |v|\}$ ne sont pas algébriques.

Solutions

Exercice 1

Soit L non vide engendré par $G = (\Sigma, V, P, S)$.

On note V' l'ensemble des variables X telles que $L_G(X) \neq \emptyset$ et P' l'ensemble des règles ne faisant intervenir que les variables de V' . $S \in V'$ car $L_G(S) = L(G) = L \neq \emptyset$.

$G' = (\Sigma, V', P', S)$ engendre aussi L .

En effet, toute dérivation dans G' est une dérivation dans G donc $L(G') \subset L(G)$.

Inversement, si $u \in L(G)$, il existe une dérivation qui génère u donc un arbre de dérivation dont u est la frontière. Pour tout nœud d'étiquette X de l'arbre, le sous-arbre dont X est la racine a une frontière qui est un facteur de u donc $L_G(X) \neq \emptyset$. Ainsi toutes les variables d'une dérivation appartiennent à V' donc $u \in L(G')$. On a bien $L(G) = L(G')$.

Si on note V'' l'ensemble des variables qui sont des étiquettes d'un nœud des arbres de dérivations des mots de $L(G')$ alors, par définition, (G'') engendre L aussi pour $G'' = (\Sigma, V'', P'', S)$ où P'' est l'ensemble des règles ne faisant intervenir que les variables de V'' .

De plus toute variable $X \in V''$ est dans un arbre de dérivation donc il existe une dérivation $S \xRightarrow{*} \alpha X \beta$ avec $\alpha, \beta \in (\Sigma \cup V'')^*$.

Exercice 2

On part d'une grammaire réduite $G = (\Sigma, V, P, S)$. De laquelle on retire les règles $X \rightarrow \varepsilon$.

Pour conserver la possibilité de dérivations qui utilisent une de ces règles on ajoute des règles qui éliminent les variables annulables.

Si $X \rightarrow \alpha$ est une règle qui contient les variables annulables $\{X_i ; i \in I\}$:

- si α contient d'autres lettres, on ajoute, pour chaque $J \subset I$, les règles $X \rightarrow \alpha_J$ où α_J est obtenu en retirant de α les variables X_i pour $i \in J$,
- si $\alpha = X_1 X_2 \cdots X_p$ avec X_i annulable pour tout i , on ajoute, pour chaque $J \subsetneq \{1, 2, 3, \dots, p\}$ avec $J \neq \emptyset$ la règle $X \rightarrow \prod_{i \in J} X_i$

On note $G' = (\Sigma, V, P', S)$ la grammaire obtenue.

Si $u \in L(G')$, on considère un arbre de dérivation pour u dans G' .

Pour chaque usage d'une règle de dérivation tronquée de G , on ajoute dans l'arbre les variables annulables qui ont été retirées que l'on prolonge en arbre de frontière ε . L'arbre obtenu est un arbre de dérivation de G dont la frontière est toujours u .

Inversement, si $u \neq \varepsilon$ est généré par g on considère un arbre de dérivation pour u dans G . Si l'arbre contient des feuilles ε , on peut retirer les branches qui les génèrent jusqu'à l'apparition d'une variable dans une règle qui n'est pas de la forme ε , et on obtient des règles tronquées de G' avec un arbre non vide donc $u \in L(G)$: les deux grammaires sont faiblement équivalentes.

Exercice 3

D'après l'exercice précédent $L \setminus \{\varepsilon\}$ est engendré par une grammaire réduite sans règle $X \rightarrow \varepsilon$.

Il reste à éliminer les règles $X \rightarrow Y$.

Toute dérivation $X \xRightarrow{*} X'$ est une suite de règles $X_i \rightarrow X_{i+1}$ car il n'y a pas de règle $X_i \rightarrow \varepsilon$; si, de plus il existe une règle $X' \rightarrow \alpha$ avec α non réduit à une variable, on retranche les règles $X_i \rightarrow X_{i+1}$ de P et on ajoute $X_i \rightarrow \alpha$. On obtient une grammaire propre.

Comme dans l'exercice précédent, on montre qu'on a conservé le langage engendré.

Exercice 4

On part d'une grammaire propre réduite qui engendre L . On ajoute une variable X_x la règle $X_x \rightarrow x$ pour chaque lettre $x \in \Sigma$ pour laquelle il n'y a pas de règle $T \rightarrow x$.

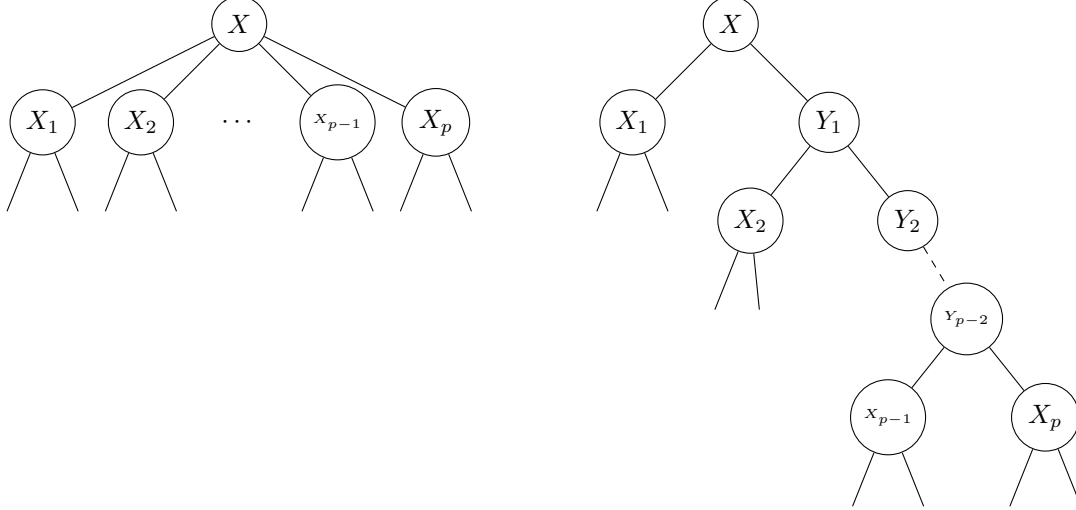
On transforme les règles de départ en remplaçant chaque lettre qui apparaît à droite par la variable correspondante.

La nouvelle grammaire est faiblement équivalente car les arbres de dérivations pour G' s'obtiennent en remplaçant les feuilles étiquetées par x en l'arbre de racine X_x et de fils unique la feuille d'étiquette x .

Exercice 5

On part d'une grammaire définie par l'exercice ci-dessus. Pour chaque $X \rightarrow X_1 X_2 \cdots X_p$ on ajoute $p - 2$ variables $Y_1, Y_2, \dots, Y_{p-2}$ et les règles $X \rightarrow X_1 Y_1, Y_1 \rightarrow X_2 Y_2, \dots, Y_{p-3} \rightarrow X_{p-2} Y_{p-2}$ et $Y_{p-2} \rightarrow X_{p-1} X_p$.

Les grammaires obtenues sont faiblement équivalentes car on transforme un arbre de dérivation de la façon suivante



Exercice 6

On note $V' = \{S_0\} \sqcup \{X_s^t ; X \in V \text{ et } s, t \in \mathbb{S}\}$.

On va définir une grammaire $G' = (\Sigma, V', P', S_0)$.

On crée les règles $S_0 \rightarrow S_{s_0}^t$ pour tout $t \in T$.

Si $X \rightarrow a \in P$ et $t = \delta(s, a)$, on crée la règle $X_s^t \rightarrow a$ pour tout $s \in \mathbb{S}$.

Si $X \rightarrow YZ \in P$, pour tous $s_1, s_2, s_3 \in \mathbb{S}$, on crée dans P' la règle $X_{s_1}^{s_3} \rightarrow Y_{s_1}^{s_2} Z_{s_2}^{s_3}$.

On montre par récurrence sur la longueur de la dérivation que

$X_s^t \xRightarrow{*} u$ pour G' si et seulement si $t = \delta^*(s, u)$ et $X \xRightarrow{*} u$ pour G .

Ainsi, si $u \in L(G) \cap L(Q)$, en notant $t = \delta^*(s_0, u)$, on a $S \Rightarrow S_{s_0}^t$ et $S \xRightarrow{*} u$ pour G donc $S_{s_0}^t \xRightarrow{*} u$ puis $S_0 \xRightarrow{*} u$ pour $G' : u \in L(G')$.

Inversement, si $u \in L(G')$, on a $S_0 \xRightarrow{*} u$ pour G' .

La première dérivation doit être $S_0 \Rightarrow S_{s_0}^t$ avec $t \in T$ et la fin est $S_{s_0}^t \xRightarrow{*} u$ dans G' donc $S_0 \xRightarrow{*} u$ dans G , $u \in L(G)$ et $t = \delta^*(s_0, u)$ donc $u \in L(Q)$. On en déduit que $u \in L(G) \cap L(Q)$.

G' engendre bien $L(G) \cap L(Q)$.

N.B. G' peut ne pas être réduite.

Exercice 7

Dans les arbres de dérivations des grammaires en forme normale de Chomsky chaque nœud admet au plus 2 fils.

De plus il n'existe pas d'arbre de dérivation de hauteur 0.

Soit $\mathcal{P}(h)$ la propriété un arbre de dérivation d'une grammaire en forme normale de Chomsky de hauteur h admet n feuilles avec $h \leq n \leq 2^{h-1}$.

Un arbre de hauteur 1 admet une feuille donc $\mathcal{P}(1)$ est vérifiée.

On suppose que $\mathcal{P}(k)$ est valide pour tout $k \leq h$ et $k > 0$.

Si un arbre de dérivation est de hauteur $h + 1$, la racine admet deux fils de hauteur h au plus ; de plus l'un est de hauteur h et l'autre de hauteur 1 au moins.

Le nombre de feuille vérifie donc $1 + h \leq n \leq 2^{h-1} + 2^{h-1} : \mathcal{P}(h + 1)$ est valide.

Ainsi, par récurrence, $\mathcal{P}(h)$ est valide pour tout h .

Comme chaque feuille contient une lettre, le résultat s'en déduit.

Exercice 8

Si $|u| \geq 2^{|V|}$, l'exercice précédent montre que la hauteur d'un arbre de dérivation est au moins $|V| + 1$.

On considère une branche de l'arbre de longueur maximale donc au moins $|V| + 1$.

Au dessus de la feuille il y a au moins $|V| + 1$ nœuds donc, parmi les $|V| + 1$ derniers il y a 2 nœuds de même étiquette.

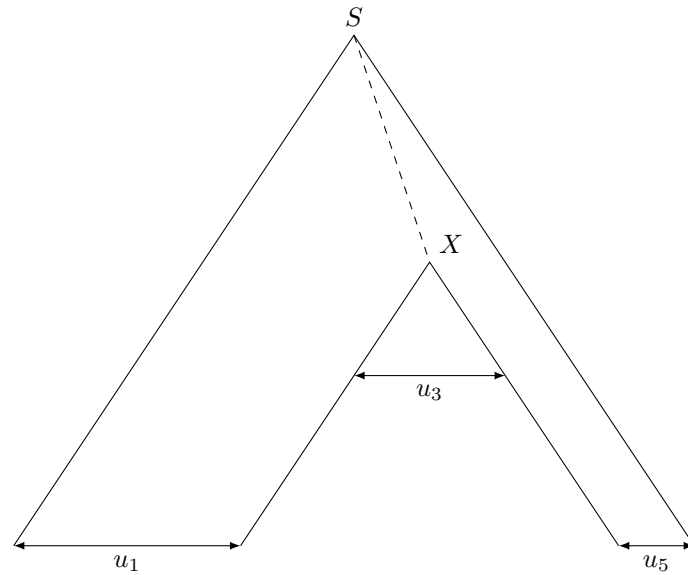
Exercice 9

La hauteur de l'arbre de racine le premier X est au plus $|V| + 1$ donc le mot généré, $u_2 u_3 u_4$ est de longueur au plus $2^{|V|+1-1} = N$.

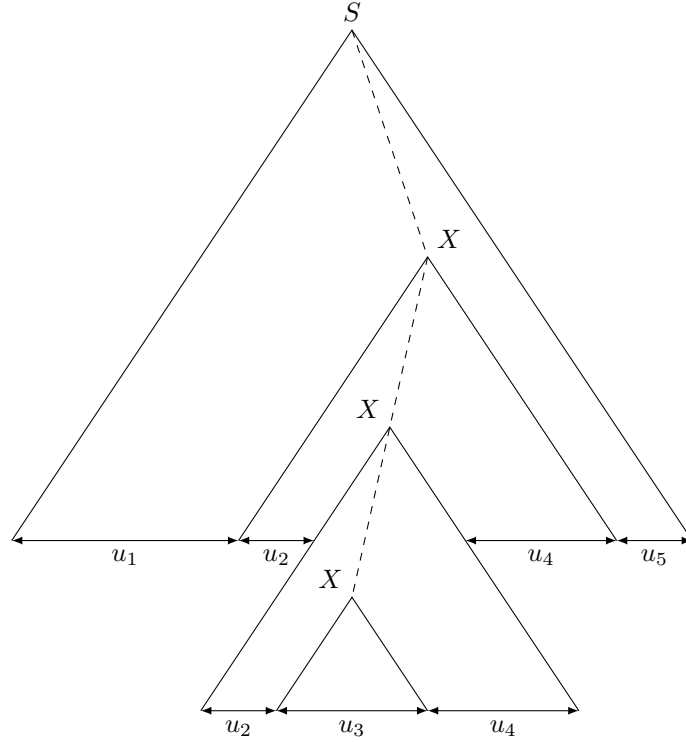
Cet arbre admet 2 fils car X est une variable donc le fils qui ne contient pas le second X génère un mot non vide v qui est un suffixe de u_4 ou un préfixe de u_2 selon le cas.

Exercice 10

La grammaire génère $u_1 u_3 u_5$



La grammaire génère $u_1 u_2^2 u_3 u_4^2 u_5$



On peut continuer, pas-à-pas, par récurrence sur k .

Exercice 11

Si L_1 est algébrique, on considère un entier N découlant du lemme de pompage.

$a^N b^N c^N = u_1 u_2 u_3 u_4 u_5$. Un mot non vide, u_2 ou u_4 , est de la forme a^p , b^p ou c^p car sinon $u_1 u_2^2 u_3 u_4^2 u_5$ ne serait pas de la forme $a^p b^q c^r$.

Alors $u_1 u_3 u_5$ admet au plus deux lettres avec un exposant strictement inférieur à N : ce n'est pas un mot de L .

La contradiction montre que L_1 n'est pas algébrique.

Si L_2 est algébrique, on considère un entier N découlant du lemme de pompage.

On choisit $u = a^N b^N a^N b^N \in L_2$, on peut écrire $u = u_1 u_2 u_3 u_4 u_5$ avec $|u_2 u_3 u_4| \leq N$ et $|u_2 u_4| \geq 1$.

On écrit aussi $u = v_1 v_2 v_3 v_4$ avec $v_1 = v_3 = a^N$ et $v_2 = v_4 = b^N$.

Comme $|u_2 u_3 u_4| \leq N$, $u_2 u_3 u_4$ est un facteur de v_1 , $v_1 v_2$, v_2 , $v_2 v_3$, v_3 , $v_3 v_4$ ou v_4 .

- Si $u_2 u_3 u_4$ est un facteur de v_1 alors $u_2 = a^p$ et $u_3 = a^q$ avec $p + q \geq 1$ et $u_1 u_3 u_5 = a^{N-(p+q)} b^N a^N b^N$ n'appartient pas à L_2 .
- De même si $u_2 u_3 u_4$ est un facteur de v_2 , v_3 ou v_4 .
- Si $u_2 u_3 u_4$ est un facteur de $v_1 v_2$ alors $u_2 = a^p$ et $u_4 = a^q b^r$ ou $u_2 = a^p b^q$ et $u_4 = b^r$ avec $p + r > 0$.
Dans le premier cas $u_1 u_3 u_5 = a^{N-(p+q)} b^{N-r} a^N b^N$ n'appartient pas à L_2 .
De même dans le second cas.
- De même si $u_2 u_3 u_4$ est un facteur de $v_2 v_3$ ou $v_3 v_4$.

Dans tous les cas $u_1 u_2^0 u_3 u_4^0 u_5 \notin L_2$.

La contradiction montre que L_2 n'est pas algébrique.