

Grammaires

Corrigé partiel de la feuille de travaux dirigés n°8

15 avril 2004

1. Numérotons les règles :

$$\begin{aligned} S &\rightarrow_1 aB & A &\rightarrow_5 bAA \\ S &\rightarrow_2 bA & B &\rightarrow_6 b \\ A &\rightarrow_3 a & B &\rightarrow_7 bS \\ A &\rightarrow_4 aS & B &\rightarrow_8 aBB \end{aligned}$$

a) On a trois dérivations gauches. Toutes trois commencent par $S \rightarrow_1 aB \rightarrow_8 aaBB \rightarrow_8 aaaBBB$ mais diffèrent ensuite

$$\begin{aligned} &aaaBBB \rightarrow_6 aaabBB \rightarrow_7 aaabbSB \rightarrow_1 aaabbaBB \\ &aaaBBB \rightarrow_7 aaabSBB \rightarrow_2 aaabbABB \rightarrow_3 aaabbaBB \\ &aaaBBB \rightarrow_6 aaabBB \rightarrow_6 aaabbB \rightarrow_8 aaabbaBB \end{aligned}$$

puis se terminent de la même manière : $aaabbaBB \rightarrow_6 aaabbabB \rightarrow_7 aaabbabbS \rightarrow_2 aaabbabbbaA \rightarrow_3 aaabbabbba$

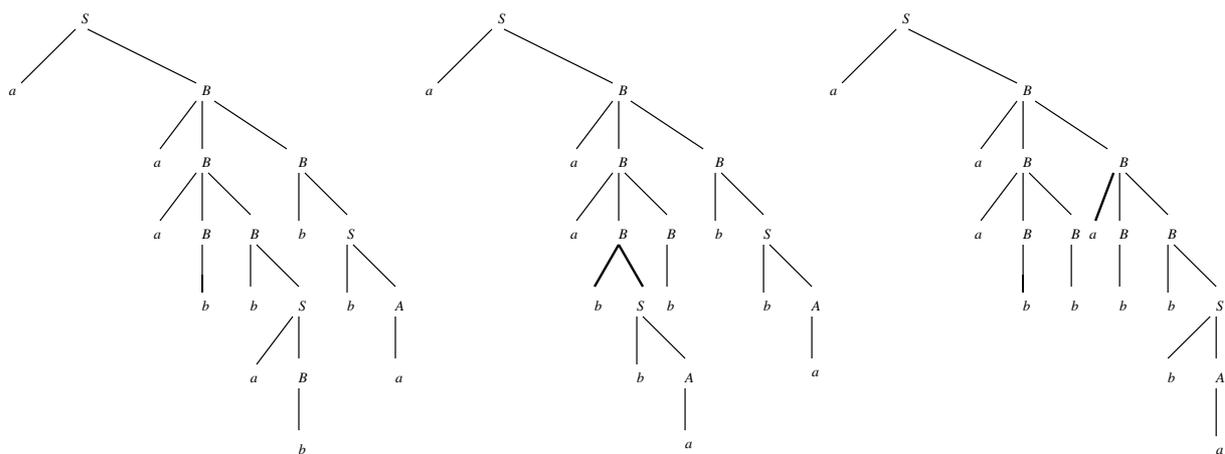
b) Comme pour la dérivation gauche, on a trois dérivations droites qui commencent de la même manière $S \rightarrow_1 aB \rightarrow_8 aaBB$ mais diffèrent ensuite

$$aaBB \rightarrow_7 aaBbS \rightarrow_2 aaBbbA \rightarrow_3 aaBbba \rightarrow_8 aaaBBbba \rightarrow_7 aaaBbSbba \rightarrow_1 aaaBbaBbba \rightarrow_6 aaaBbabbba \rightarrow_6 aaabbabbba$$

$$aaBB \rightarrow_7 aaBbS \rightarrow_2 aaBbbA \rightarrow_3 aaBbba \rightarrow_8 aaaBBbba \rightarrow_6 aaaBbbba \rightarrow_7 aaaBbSbba \rightarrow_2 aaabbAbbba \rightarrow_3 aaabbabbba$$

$$aaBB \rightarrow_8 aaBaBB \rightarrow_7 aaBaBbS \rightarrow_2 aaBaBbbA \rightarrow_3 aaBaBbba \rightarrow_6 aaBabbba \rightarrow_8 aaaBBabbba \rightarrow_6 aaaBbabbba \rightarrow_6 aaabbabbba$$

c) Un arbre syntaxique qui engendre le mot $aaabbabbba$:



Un parcours en profondeur gauche de cet arbre donne une dérivation gauche associée, idem pour la droite. Ce mot a en fait trois arbres syntaxiques, ce qui implique que la grammaire est ambiguë.

Remarque : On observe que :

- la grammaire donnée engendre l'ensemble des mots ayant autant de a que de b ,

- la variable A engendre l'ensemble des mots ayant exactement 1 a de plus que de b ,
- la variable B engendre l'ensemble des mots ayant exactement 1 b de plus que de a .

2.

a) On peut partir du fait que $L_1 = \{a^n b^n | n > 0\}c^+$. Il suffit alors de trouver une variable A qui engendre $\{a^n b^n | n > 0\}$, et une variable C qui engendre c^+ pour engendrer L_1 par une variable S avec la règle : $S \rightarrow AC$.

Or on peut engendrer le langage (rationnel) c^+ via les 2 règles suivantes : $C \rightarrow c | cC$ et le langage $\{a^n b^n | n > 0\}$ via les 2 règles suivantes : $A \rightarrow ab | aAb$. On obtient :

$N = \{S, A, C\}; T = \{a, b, c\}; S$ et $P = \{S \rightarrow AC, A \rightarrow ab | aAb, C \rightarrow c | cC\}$ qui engendre le langage algébrique L_1 .

b) L'idée est de dériver des mots dont le début est identique à la fin, puis à un moment donné d'engendrer une différence. A partir de ce moment, la seule contrainte à respecter est la parité de la longueur. On obtient donc : $N = \{S, T, U\}; T = \{a, b\}; S$ et $P = \{S \rightarrow aSa | bSb | aTb | bTa, T \rightarrow \varepsilon | aU | bU, U \rightarrow aT | bT\}$

c) Le langage L_3 est l'union de 2 langages, $L_{31} = \{a^i b^j | i \neq j\}c^+$ et $L_{32} = a^+ \{b^i c^j | i \neq j\}$. Soient les variables S_{31} et S_{32} qui engendrent respectivement L_{31} et L_{32} , et la variable S_3 qui permet d'engendrer le langage union L_3 via la règle : $S_3 \rightarrow S_{31} | S_{32}$.

$N = \{S_3, S_{31}, S_{32}, T_{31}, T_{32}, A, B, C\}, T = \{a, b, c\}, S_3$

$$P \begin{cases} S_3 \rightarrow S_{31} | S_{32} & T_{32} \rightarrow bT_{32}c | bB | cC \\ S_{31} \rightarrow T_{31}C & A \rightarrow aA | \varepsilon \\ T_{31} \rightarrow aT_{31}b | aA | bB & B \rightarrow bB | \varepsilon \\ S_{32} \rightarrow AT_{32} & C \rightarrow cC | \varepsilon \end{cases}$$

d) On peut utiliser le langage et la grammaire de l'exercice 1. Il suffit de l'adapter, pour engendrer à la place des mots contenant autant de a que de b les mots contenant soit plus de a que de b , soit plus de b que de a . Ainsi A (resp. B) engendre les mots ayants un excédent de a (resp. b) et E engendre les mots équilibrés. On

obtient : $N = \{S, A, B, E, C, D\}, T = \{a, b\}, S, P \begin{cases} S \rightarrow A | B & E \rightarrow aD | bC | \varepsilon \\ A \rightarrow aE | aA | EA & C \rightarrow a | aE | bCC \\ B \rightarrow bE | bB | EB & D \rightarrow b | bE | aDD \end{cases}$

e) Il suffit d'utiliser les règles qui permettent de créer des expressions rationnelles (comme dans la grammaire

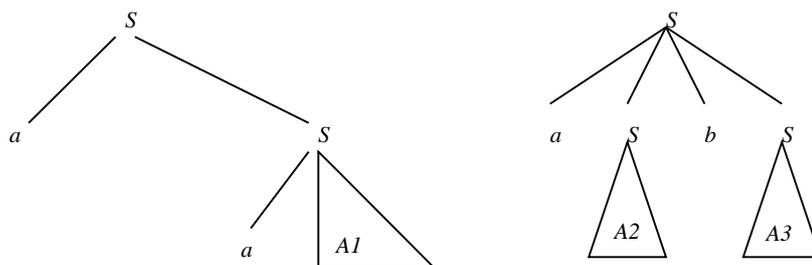
ETF). $N = \{E, T, F\}, T = \{a, b, c, *, +, (,), e\}$ où e représente ε , $E P \begin{cases} E \rightarrow TE | T * | T \\ T \rightarrow F + T | F \\ F \rightarrow (E) | a | b | c | e \end{cases}$

3. Notons L_S le langage engendré par S et L le langage des mots w vérifiant la propriété $P(w)$:

$P(w)$: tout préfixe de w a au moins autant de a que de b .

Il s'agit de montrer que $L_S = L$.

- On montre d'abord que tout mot w de L_S vérifie $P(w)$, c.a.d. $L_S \subseteq L$, par induction sur la hauteur des arbres syntaxiques. Pour les mots de L_S obtenus avec un arbre syntaxique de hauteur 1 (i.e. ε) ou de hauteur 2 (i.e. a ou ab) la propriété $P(w)$ est vraie. Soit n un entier, supposons $P(w)$ vraie pour tous les mots de L_S obtenus avec un arbre syntaxique de hauteur inférieure ou égale à n . Soit w un mot de L_S obtenu avec un arbre syntaxique de hauteur $n + 1$. Les arbres pouvant donner w sont de l'un des 2 types suivants :



où A_1 est un arbre syntaxique de hauteur n , et A_2, A_3 sont 2 arbres syntaxiques dont au moins 1 est de hauteur n . Par hypothèse d'induction, les mots obtenus à partir de ces arbres vérifient $P(w)$. Il est alors immédiat de vérifier que w vérifie aussi $P(w)$. Ainsi tous les mots de L_S vérifient $P(w)$ et donc $L_S \subseteq L$.

- Réciproquement, on montre que $L \subseteq L_S$. Les mots de L sont exactement les mots w tels que pour tout préfixe w' de w on a : $|w'|_a - |w'|_b \geq 0$. Remarquons que tout mot de L commence par a et qu'il y a 2 types de mots dans L :

1. les mots w tels que pour tout préfixe w' de w on a : $|w'|_a - |w'|_b > 0$ - ces mots s'écrivent am , où m est un mot de L
2. les mots w tels que pour au moins 1 préfixe w' on a : $|w'|_a - |w'|_b = 0$. Soit v le plus court des préfixes w' . $w = vv'$, où :
 - $v \in L$ (tout préfixe d'un mot de L est aussi un mot de L) qui se termine par b (car $|v|_a - |v|_b = 0$), et donc $v = aub$ où $u \in L$ (car v étant le plus petit mot du type 2., au est du type 1.)
 - $v' \in L$, car $|v'|_a - |v'|_b = 0$ et que vv' (i.e. w) est un mot de L .

On raisonne alors par induction sur la longueur des mots de L . Les mots de L de longueur 0 ou 1 sont bien engendrés par la grammaire. Soit n un entier, supposons que tous les mots de L de longueur inférieure ou égale à n sont obtenus par la grammaire. Soit w un mot de longueur $n + 1$:

1. soit il est du type 1., i.e. $w = am$ où m est un mot de L de longueur n et donc engendré par la grammaire, d'où en appliquant $S \rightarrow aS$, puis en dérivant S pour obtenir m , on obtient w à partir de S .
2. soit il est du type 2., i.e. $w = aubv$ où $u, v \in L$ sont de longueur inférieure à n et donc engendrés par la grammaire, en appliquant $S \rightarrow aSbS$, puis en dérivant le premier S pour obtenir u et le deuxième S pour obtenir v , on obtient w à partir de S .

Donc tous les mots de L sont engendrés par la grammaire et $L \subseteq L_S$.

D'où le résultat.