

TD langages algébriques

David A. Madore

23 janvier 2017

INF105

Git:d851d50 Mon Jan 23 17:10:23 2017 +0100

Exercice 1.

Considérons le fragment simplifié suivant de la grammaire d'un langage de programmation hypothétique :

Instruction → foo | bar | qux | *Conditional*
 | begin *InstrList* end
Conditional → if *Expression* then *Instruction* else *Instruction*
 | if *Expression* then *Instruction*
InstrList → *Instruction* | *Instruction InstrList*
Expression → true | false | happy | trippy

(Ici, les « lettres » ou tokens ont été écrits comme des mots, par exemple foo est une « lettre » : les terminaux sont écrits en police à espacement fixe tandis que les nonterminaux sont en italique et commencent par une majuscule. On prendra *Instruction* pour axiome.)

(1) Donner l'arbre d'analyse de : if happy then if trippy then foo else bar else qux ; expliquer brièvement pourquoi il n'y en a qu'un.

(2) Donner deux arbres d'analyse distincts de : if happy then if trippy then foo else bar. Que peut-on dire de la grammaire présentée ?

(3) En supposant que, dans ce langage, begin *I* end (où *I* est une instruction) a le même effet que *I* seul, comment un programmeur peut-il réécrire l'instruction considérée en (2) pour obtenir un comportant équivalent à l'une ou l'autre des deux interprétations ?

(4) Modifier légèrement la grammaire proposée de manière à obtenir une grammaire faiblement équivalente dans laquelle seul l'un des arbres d'analyse

obtenus en (2) est possible (i.e., une grammaire qui force cette interprétation-là par défaut); on pourra être amené à introduire des nouveaux nonterminaux pour des variantes de *Instruction* et *Conditional* qui interdisent récursivement les conditionnelles sans else.

Exercice 2.

Soit $\Sigma = \{a, b\}$. On considère le langage M des mots qui *ne s'écrivent pas* sous la forme ww avec $w \in \Sigma^*$ (c'est-à-dire sous la forme d'un carré; autrement dit, le langage M est le *complémentaire* du langage Q des carrés considéré dans l'exercice 3): par exemple, M contient les mots a, b, ab, aab et $aabb$ mais pas $\varepsilon, aa, abab$ ni $abaaba$.

(0) Expliquer pourquoi tout mot sur Σ de longueur impaire est dans M , et pourquoi un mot $x_1 \cdots x_{2n}$ de longueur paire $2n$ est dans M si et seulement si il existe i tel que $x_i \neq x_{n+i}$.

On considère par ailleurs la grammaire hors contexte G (d'axiome S)

$$\begin{aligned} S &\rightarrow A \mid B \mid AB \mid BA \\ A &\rightarrow a \mid aAa \mid aAb \mid bAa \mid bAb \\ B &\rightarrow b \mid aBa \mid aBb \mid bBa \mid bBb \end{aligned}$$

(1) Décrire le langage $L(G, A)$ des mots dérivant de A dans la grammaire G (autrement dit, le langage engendré par la grammaire identique à G mais ayant pour axiome A). Décrire de même $L(G, B)$.

(2) Montrer que tout mot de longueur impaire est dans le langage $L(G)$ engendré par G .

(3) Montrer que tout mot $t \in M$ de longueur paire est dans $L(G)$. (Indication: si $t = x_1 \cdots x_{2n}$ est de longueur paire $2n$ et que $x_i \neq x_{n+i}$, on peut considérer la factorisation de t en $x_1 \cdots x_{2i-1}$ et $x_{2i} \cdots x_{2n}$.)

(4) Montrer que tout mot de $L(G)$ de longueur paire est dans M .

(5) En déduire que M est algébrique.

Exercice 3.

Soit $\Sigma = \{a, b\}$. Montrer que le langage $Q := \{ww : w \in \Sigma^*\}$ constitué des mots de la forme ww (autrement dit, des carrés; par exemple, $\varepsilon, aa, abab, abaaba$ ou encore $aabbaabb$ sont dans Q) n'est pas algébrique. On pourra pour cela considérer son intersection avec le langage L_0 dénoté par l'expression rationnelle $a*b*a*b*$ et appliquer le lemme de pompage.

Remarque : Les exercices 2 et 3 mis ensemble donnent un exemple explicite d'un langage M algébrique dont le complémentaire Q n'est pas algébrique.