Td n° 14

Grammaires non contextuelles (1)

Hugo Salou MPI*

Dernière mise à jour le 13 mars 2023

1 Exemples de dérivations

1. Les dérivations (b), et (d) sont vraies.
2. Les dérivations (a), (c), et (d).
3. On a $\{ab, ba, aab\} \subseteq \mathscr{L}(\mathscr{G})$. En effet,
   — $\mathrm{S} \Rightarrow \mathrm{D} \Rightarrow a\mathrm{T}b \Rightarrow ab$,
   — $\mathrm{S} \Rightarrow \mathrm{D} \Rightarrow b\mathrm{T}a \Rightarrow ba$,
   — $\mathrm{S} \Rightarrow \mathrm{D} \Rightarrow a\mathrm{T}b \Rightarrow a\mathrm{X}b \Rightarrow aab$.
4. On a $\varepsilon \notin \mathscr{L}(\mathscr{G})$, $aa \notin \mathscr{L}(\mathscr{G})$ et $bb \notin \mathscr{L}(\mathscr{G})$.
5. Le langage $\mathscr{L}(\mathscr{G})$ est l'ensemble des mots non palindromes.

2 Arbres de dérivations

1.

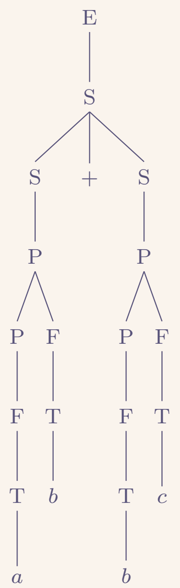

Figure 1 – Arbre de dérivation de $ab+bc$ dans la grammaire $\mathscr{G}$

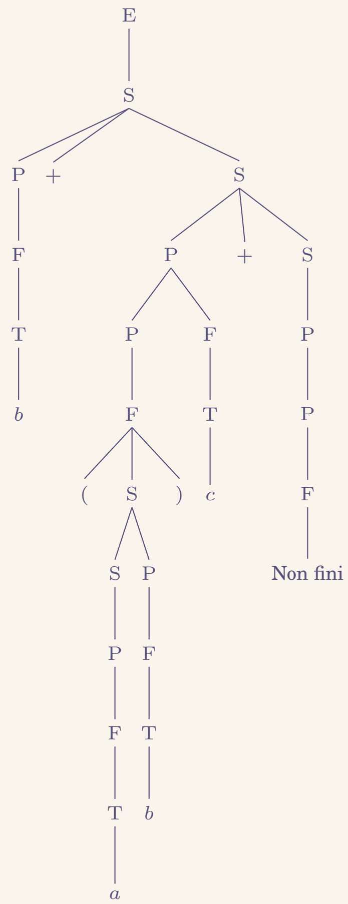

Figure 2

2.

3 Construction de grammaires

1. — $\mathscr{G}_0 = (\Sigma, \{\mathrm{L}, \mathrm{A}, \mathrm{S}\}, \{\mathrm{S} \to \mathrm{LLLA}, \mathrm{L} \to a \mid b, \mathrm{A} \to \mathrm{LA} \mid \varepsilon\}, \mathrm{S})$,
   — $\mathscr{G}_1 = (\Sigma, \{\mathrm{L}, \mathrm{A}, \mathrm{S}\}, \{\mathrm{S} \to \mathrm{A}a\mathrm{A}a\mathrm{A}a\mathrm{A}, \mathrm{A} \to \mathrm{LA} \mid \varepsilon, \mathrm{L} \to a \mid b\}, \mathrm{S})$,
   — $\mathscr{G}_2 = (\Sigma, \{\mathrm{S}, \mathrm{V}\}, \{\mathrm{S} \to a\mathrm{X}a \mid b\mathrm{X}b, \mathrm{X} \to a\mathrm{X} \mid b\mathrm{X} \mid \varepsilon\}, \mathrm{S})$,

— $\mathcal{G}_3(\Sigma, \{S, X\}, \{S \to X, X \to S\}, S)$.

2. — $\mathcal{G}_4 = (\Sigma, \{L, S, X\}, \{S \to LX, L \to a \mid b, X \to LLX \mid \varepsilon\}, S)$,
   — $\mathcal{G}_5 = (\Sigma, \{S, X\}, \{S \to XaX, X \to aX \mid bX \mid \varepsilon\}, S)$,
   — $\mathcal{G}_6 = (\Sigma, \{S, X\}, \{S \to aSa \mid bSb \mid X, X \to \varepsilon \mid a \mid b\}, S)$,
   — *c.f.* exercice 1.
3. — $\mathcal{G}_8 = (\Sigma, \{S\}, \{S \to \varepsilon \mid aSb\}, S)$,
   — $\mathcal{G}_9 = (\Sigma, \{S, X\}, \{S \to aSb \mid SX, X \to bX \mid \varepsilon\}, S)$,
   — $\mathcal{G}_{10} = (\Sigma, \{S\}, \{S \to aSbb \mid Sb \mid \varepsilon\}, S)$
4. — $\mathcal{G}_{11} = (\Sigma, \{S \to aSb \mid bSa \mid aS \mid SS \mid \varepsilon\}, S)$,

4 Raisonner par induction sur une grammaire

1. Montrons le par induction.
   — **Cas S $\to$ aS.** Soit $w = aw'$ un mot, où ba n'est pas un sous-mot de w'. Alors, ba n'est pas un sous-mot de $w = aw'$.
   — **Cas S $\to$ Sb.** Soit $w = w'b$ un mot, où ba n'est pas un sous-mot de w'. Alors, ba n'est pas un sous-mot de $w = w'b$.
   — **Cas $S \to \varepsilon \mid a \mid b$.** Le mot ba n'est pas un sous-mot de a, ni de b, ni de ε.
2. Montrons, par double-inclusion, que $\mathcal{L}(\mathcal{G}) = \mathcal{L}(a^\star b^\star)$.
   "$\supseteq$" Soit $w \in \mathcal{L}(a^\star b^\star)$. Il existe m et n deux entiers tels que $w = a^n b^m$. On applique la dérivation

$$\underbrace{S \Rightarrow aS \Rightarrow aaS \Rightarrow \cdots \Rightarrow a^n S}_{n \text{ fois}} \Rightarrow \underbrace{a^n Sb \Rightarrow a^n Sbb \Rightarrow \cdots \Rightarrow a^n Sb^m}_{m \text{ fois}} \Rightarrow a^n b^m.$$

   "$\subseteq$" D'après la question 1, on a ce sens de l'inclusion.

5 Ambigüité

6 Langage de Dyck

1. On suppose ce langage reconnaissable par un automate à n états. On considère le mot $w = (^n \cdot)^n$, donc $|w| \geqslant n$. Ainsi, il existe x, y et z trois mots tels que $w = xyz$, $|xy| \leqslant n$, $y \neq \varepsilon$ et $\forall p \in \mathbb{N}$, $xy^pz \in \mathcal{L}(\mathcal{G})$. Soit alors $p \in [\![1, n-1]\!]$ et $q \in [\![1, n-p]\!]$ tels que $x = (^p$, $y = (^q$ et $z = (^{n-q-p} \cdot)^n$. Ainsi, $xy \in \mathcal{L}(\mathcal{G})$, ce qui est absurde. On en déduit que $\mathcal{L}(\mathcal{G})$ n'est pas reconnaissable, il n'est donc pas régulier.
2. On pose $\mathcal{G} = (\Sigma, \{S\}, \{S \to (S) \mid SS \mid \varepsilon\}, S)$.
3. — On le montre par induction.
      — **Cas S $\to$ ε.** On a $|\varepsilon|_(= 0 = |\varepsilon|_)$.
      — **Cas S $\to$ (S).** Soit $u \in \mathcal{L}(\mathcal{G})$ avec $|u|_(= |u|_) = n$. Ainsi, $|(u)|_(= |(u)|_) = n + 1$.
      — **Cas S $\to$ SS.** Soient u et v deux mots de $\mathcal{L}(\mathcal{G})$ tels que $|u|_(= |u|_) = n$ et $|v|_(= |v|_) = m$. Alors, $|u \cdot v|_(= |v \cdot u|_) = n + m$.
   — Montrons par induction $\mathcal{P}_u$: « pour tout v préfixe de u, $|v|_(\geqslant |v|_)$. »
      — **Cas S $\to$ ε.** Le seul préfixe de ε est ε, et on a bien $|\varepsilon|_(= 0 \geqslant 0 = |\varepsilon|_)$.
      — **Cas S $\to$ (S).** Soit u un mot de $\mathcal{L}(\mathcal{G})$ vérifiant $\mathcal{P}_u$. Soit v un préfixe de (u). On procède par induction sur v.
         — **Cas $v = \varepsilon$ ou (.** ok.
         — **Cas ($\tilde{u}$,** où $\tilde{u}$ est un préfixe de u. Par hypothèse d'induction, $|\tilde{u}|_(\geqslant |\tilde{u}|_)$ donc $|(\tilde{u}|_(= |(\tilde{u}|_)$.
         — **Cas (u).** Par hypothèse d'induction, $|u|_(\geqslant |u|_)$ donc $|(u)|_(\geqslant |(u)|_)$.
4. On note $\overline{w}^j = |w_{[\![0,j]\!]}|_(- |w_{[\![0,j]\!]}|_)$. Alors les deux conditions se traduisent par $\overline{w}^{|w|} = 0$ et $\forall i \in [\![0, |w|-1]\!]$, $\overline{w}^i \geqslant 0$.

7 Listes OCaml

On pose $\mathcal{G} = (\Sigma, \{\mathrm{S}, \mathrm{L}, \mathrm{B}\}, \{\mathrm{S} \to \texttt{[L]} \mid \texttt{[]}, \mathrm{L} \to \mathrm{B}\texttt{;}\mathrm{L} \mid \mathrm{B}, \mathrm{B} \to \texttt{true} \mid \texttt{false}\}, \mathrm{S})$.

8 Mots de Łukasiewicz

1. On a $\Sigma \cap \mathcal{L} = \{\square\}$, $\Sigma^2 \cap \mathcal{L} = \varnothing$ et $\Sigma^3 \cap \mathcal{L} = \{\bigcirc\square\square\}$.
2. Montrons que $\Sigma^{2N} \cap \mathcal{L} = \varnothing$, pour tout $N \in \mathbb{N}^\star$. (Dans le cas $N = 0$, le seul mot est ε, et il n'est pas dans $\mathcal{L}$.) Soit $w \in \Sigma^{2N} \cap \mathcal{L}$. On sait que $-1 = \overline{w}^{|w|} = \overline{w}^{|w|-1} + \mathsf{va}(w_{|w|}) \geqslant \mathsf{va}(w_{|w|})$, d'où $w_{|w|} = \square$, et $\overline{w}^{|w|-1} = 0$. On pose $w = u \cdot \square$, le mot u est donc de longueur impaire. Or, la somme $\sum_{k=1}^{|u|} \mathsf{va}(w_i)$ est une somme d'un nombre impaire de termes valant -1 ou 1, elle ne peut pas être nulle. Mais, comme $\overline{w}^{|w|-1} = 0$, donc elle est nulle, une contradiction. On en déduit que $\mathcal{L} \cap \Sigma^{2N} = \varnothing$. Par suite, on conclut que $\mathcal{L}$ ne contient pas de mots pairs.
3. On suppose $\mathcal{L}$ reconnaissable par un automate $\mathcal{A}$ à n états. On considère le mot $w = \bigcirc^n \cdot \square^n \cdot \square \in \mathcal{L}$. Alors, par application du lemme de l'étoile à l'automate $\mathcal{A}$ avec ce mot w, il existe donc x, y et z trois mots de $\Sigma^\star$ tels que $w = xyz$, $|xy| \leqslant n$, $y \neq \varepsilon$ et, pour tout $p \in \mathbb{N}$, $xy^pz \in \mathcal{L}$. D'où, $x = \bigcirc^k$, $y = \bigcirc^j$ et $z = \bigcirc^{n-k-j} \cdot \square^{n+1}$, où k et j sont des entiers. De plus, $j \neq 0$ car, sinon, $y = \varepsilon$. Alors, $xy^pz \in \mathcal{L}$, pour tout $p \in \mathbb{N}$. En particulier, pour $p = 0$, $xz \in \mathcal{L}$. Or, $xz = \bigcirc^{n-j} \cdot \square^{n+1} \notin \mathcal{L}$ car $\overline{xz}^{|xz|} = (n-j) - (n+1) = -j-1 \neq -1$ car $j \neq 0$. On en déduit donc que que $\mathcal{L}$ n'est pas reconnaissable par un automate à n états. Ceci étant vrai pour tout n, alors $\mathcal{L}$ n'est pas un langage régulier.
4.

```ocaml
let est_luka (w: word): bool =
  let l = List.map va w in
  let rec aux (l: int list) (s: int): bool =
    match l with
    | []      -> s = -1
    | x :: q -> s >= 0 && aux q (s + x)
  in aux l 0
```

Code 1 – Fonction `est_luka` testant si w est un mot de Łukasiewicz

5. Soit $w = u \cdot v \in \mathcal{L}$, où $u \neq \varepsilon$ est un préfixe strict de w. Alors, $0 \leqslant \overline{w}^{|u|} = \overline{u}^{|u|} \neq -1$ donc $u \notin \mathcal{L}$.
6. Soient u et v deux mots de $\mathcal{L}$. On pose $w = \bigcirc \cdot u \cdot v$. Montrons que $w \in \mathcal{L}$. On pose $n = |u|$, et $m = |v|$.
   - On a $\overline{w}^0 = 0 \geqslant 0$,
   - et $\overline{w}^1 = \mathsf{va}(\bigcirc) = 1 \geqslant 0$,
   - et, pour tout $i \in [\![1, n[\![$, $\overline{w}^{i+1} = \overline{u}^i + \mathsf{va}(\bigcirc) \geqslant \mathsf{va}(\bigcirc) \geqslant 0$,
   - et, pour tout $i \in [\![1, m[\![$, $\overline{w}^{i+n+1} = \overline{u}^n + \overline{v}^i + \mathsf{va}(\bigcirc) = \overline{v}^i \geqslant 0$,
   - et finalement, $\overline{w}^{n+m+1} = \overline{u}^n + \overline{v}^m + \mathsf{va}(\bigcirc) = -1 - 1 + 1 = -1$.

   On en conclut que $w \in \mathcal{L}$.
7. Soit $w \in \mathcal{L}$, un mot de taille n. Comme montré précédemment, $\overline{w}^{n-1} = 0$. Ainsi, l'ensemble $\{k \in [\![1, n]\!] \mid \overline{w}^i = 0\}$ est une partie de $\mathbb{N}$ non vide, elle admet donc un minimum que l'on notera k. De plus, on a montré que $w_0 = \bigcirc$. On en déduit que w peut être décomposé en
$$w = \bigcirc \cdot \underbrace{w_2 w_3 \dots w_k}_{u} \cdot \underbrace{w_{k+1} w_{k+2} \dots w_n}_{v}.$$
   Montrons que u et v sont des mots de $\mathcal{L}$. D'une part, pour $i \in [\![1, k-2]\!]$, $0 < \overline{w}^{i+1} = \overline{u}^i + \mathsf{va}(\bigcirc)$, d'où $\overline{u}^i \geqslant 0$. De plus, $0 = \overline{w}^k = \overline{u}^{k-1} + \mathsf{va}(\bigcirc)$, d'où $\overline{u}^{k-1} = \overline{u}^{|u|} = -1$. Également, pour $i \in [\![1, n-k-1]\!]$, $0 \leqslant \overline{w}^{k+i} = \overline{w}^k + \overline{v}^i = \overline{v}^i$. Finalement, $-1 = \overline{w}^n = \overline{w}^k + \overline{v}^{n-k} = \overline{v}^{|v|}$. On en conclut que u et v sont bien des mots de $\mathcal{L}$.
   Montrons, à présent, l'unicité de la décomposition. On suppose qu'il existe u' et v' deux mots de $\mathcal{L}$ tels que $w = \bigcirc \cdot u' \cdot v'$. Nous savons, en particulier, que $\overline{w}^{1+|u'|} = 0$, afin de

maintenir la condition $u \in \mathscr{L}$. Alors, $1 + |u'| \leqslant p$, et donc $|u'| \geqslant |u|$. Le mot u' est donc un préfixe de u. Au vu de la question 5., afin que u' soit un mot de $\mathscr{L}$, il est nécessaire que u' ne soit pas un préfixe strict de u. On en déduit que $u' = u$.

8. On utilise habilement la question précédente, et on pose

$$\mathscr{G} = (\{\mathrm{L}\}, \Sigma, \{\mathrm{L} \to \bigcirc\mathrm{LL} \mid \bigcirc\}, \mathrm{L}).$$

   La question précédente montre la non-ambigüité de la grammaire, et que $\mathscr{L}(\mathscr{G}) = \mathscr{L}$.

9.