LEMME D'OGDEN

Autebert (94) p105-106.

C'est un analogue du lemme de l'étoile pour les langages algébriques plutôt que rationnels.

lemme d'Ogden

Soit L le langage algébrique engendré par la grammaire algébrique $G=(\Sigma,\Gamma,R)$ ie $L=\widehat{L_G}(S)$ pour un certain $S\in\Gamma$.

Il existe alors $N\in\mathbb{N}$ tel que pour tout mot $f\in L$ contenant au moins N lettres distinguées, il existe une factorisa° de $f=\alpha u\beta v\gamma$ vérifiant :

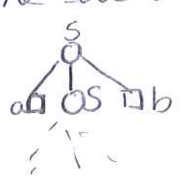
(1) il existe $T\in\Gamma$ tel que $\begin{cases} S \xrightarrow[G]{*} \alpha T\gamma \\ T \xrightarrow[G]{*} uTv \\ T \xrightarrow[G]{*} \beta \end{cases}$

(2) α, u et β contiennent des lettres distinguées ou β, v et γ ——

(3) $u\beta v$ contient moins de N lettres distinguées

Rq La condition (1) implique notamment que $\{\alpha u^n\beta v^n\gamma \mid n\in\mathbb{N}\}\subset L$.
En quelque sorte u et v sont des facteurs co-itérants.

Preuve Notons m la longueur maximale des membres droits des règles de la grammaire.
Un mot f de $\widehat{L_G(S)}$ est la frontière d'un arbre de dérivation, dont la racine est étiquetée par S et dont le degré est m. En effet chaque noeud étiqueté par un non terminal a autant de fils que le membre droit (de la règle utilisée) a de terme, tandis que les nd étiqueté par des terminaux sont des feuilles.

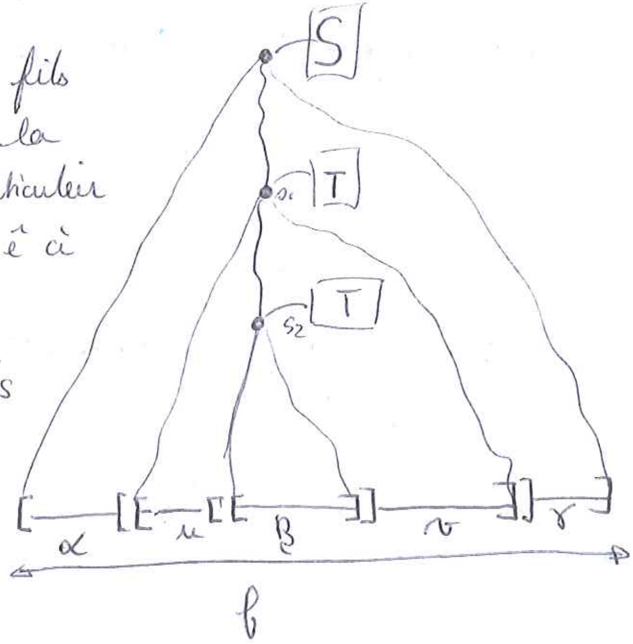


$S\to aSb$

Les feuilles de cet arbre correspondent chacune à des lettres de f, on peut donc considérer distinguées les feuilles de l'arbre correspondant à des lettres marquées de f. Dès lors on sait que si f a k feuilles marquées avec $k \geq m^r + 1$, nécessairem^t une branche de cet arbre a plus de r sommets particuliers, par contraposée du lemme : Si aucune branche d'un arbre de degré m n'a plus de r sommets particuliers, alors il y a moins de m^r feuilles dist.

En particulier pour $r = \# V$ le nombre de non terminaux, on sait qu'une branche de cet arbre présente deux sommets particuliers s_1 et s_2 étiquetés par le m̂ non terminal T.

Un sommet particulier a au moins deux fils distingués, dont l'un au plus est sur la branche considérée. Donc chaque noeud particulier a au moins un fils distingué qui peut ê à droite ou à gauche de la branche. En considérant l'aîné de ses fils distingués (ie le + à gauche) on dira qu'un nd part est de type droit si ce fils est à droite de la branche et qu'il est de type gauche sinon.



Parmi les $r+1$ sommets particuliers de la branche considérée, ou bien il y en a plus de type gauche, ou bien il y en a plus de type droit. Ainsi en posant $r = 2 \times \max(\# V, 2)$, on sait qu'il y a une branche où il y a soit au moins $\max(\# V, 2) + 1$ nd part. de type gauche
soit ——————————————————— droit.

On choisit alors pour s et t les deux plus bas.

En se rappelant que $W \xrightarrow[G]{*} \omega$ ssi il existe un arbre de dérivation de G de frontière ω et dont la racine est étiquetée par W et en considérant ici les arbres respectivement issus de s_2, s_1 et la racine (arbres de dérivation partielle tronqué sous s_2, resp. sous s_1 pour les 2 derniers) on est assuré que $S \xrightarrow[G]{*} \alpha T \gamma$; $T \xrightarrow[G]{*} u T v$ et $T \xrightarrow[G]{*} \beta$. D'où (1).

- S'il y a $\max(\# V, 2)+1$ nd. part de type gauche, alors α, β et u contiennent des lettres distinguées.
  - ↳ β car s_2 est distingué, donc a des descendants feuilles qui le sont
  - ↳ u car s_1 est de type gauche et a donc un fils distingué dont toutes les feuilles descendantes sont dans u
  - ↳ α car il y a néc un autre sommet part de type gauche sur la partie de la branche entre la racine et s_1.
    (NB: c'est pour assurer cela qu'on a pris $\max(\# V, 2)$)

- S'il y a $\max(\# V, 2)+1$ nd. part de type droit, alors β, v et γ contiennent des lettres distinguées, pour les m̂ raisons.

D'où (2).

Puisque soit α soit γ contiennent des lettres distinguées, $f' = u\beta v$ contient strictement moins de lettre distinguées que f et on peut lui réappliquer ce qu'on vient de faire.
En itérant on arrive à $u^{(n)}\beta^{(n)}v^{(n)}$ qui contient moins de N lettres distinguées, reste alors à poser $\tilde{\alpha} = \alpha\alpha'\alpha'' \text{---} \alpha^{(n)}$ $\tilde{\beta} = \beta^{(n)}$
$\tilde{\gamma} = \gamma^{(n)} \text{---} \gamma'\gamma$ $\tilde{u} = u^{(n)}$
$\tilde{v} = v^{(n)}$

ainsi la décomposition $f = \tilde{\alpha}\tilde{u}\tilde{\beta}\tilde{v}\tilde{\gamma}$ convient, pour $N = m^{\max(\# V, 2)}$

D'où (3)

□

Cor [Bar-Hillel, Perles, Shamir]
Si L est un langage algébrique
alors il existe un entier $N \in \mathbb{N}$ tel que pour tout mot $f \in L$ de longueur $\geq N$, f admet une décomposition en $f = \alpha u \beta v \gamma$ tel que $\forall n \in \mathbb{N}, \alpha u^n \beta v^n \gamma \in L$

ex $L_1 = \{a^n b^n c^n \mid n \in \mathbb{N}\}$ sur $\Sigma = \{a,b,c\}$ n'est pas algébrique.

Par l'absurde supposons que L_1 soit algébrique.
Il est alors engendré par une grammaire G, et d'après le lemme d'Ogden un certain entier N est associé à cette grammaire.

On considère alors $f = a^N \underline{b^N} c^N$. $f \in L'$. On distingue dans f toutes les occurences de b, soit N lettres.
D'après le lemme d'Ogden il existe une factorisa° de f en $\alpha u \beta v \gamma$ tq $\rightarrow$ α, u, β ou β, v, γ contiennent des lettres dist
($\rightarrow$ $u\beta v$ contient moins de N lettres dist.)
$\rightarrow$ $\forall n \in \mathbb{N}$ $\alpha u^n \beta v^n \gamma \in L_1$.

- Si α, u, et β contiennent des lettres dist. c-à-d des b.
  - $\rightarrow$ α s'arrête après tous les a, après le début des b m̂.
  - $\rightarrow$ β commence avant la fin des b.

  Donc néc u n'est constitué que de b, et v ne contient aucun a.
  De ce fait itérer simultaném^t u et v fait varier le nbre de b, sans faire varier celui de a $\rightarrow$ on sort de L_1. <u>IMP</u>
- Si β, v et γ contiennent des lettres dist. alors néc v est constitué que de b tandis que u ne contient aucun c, on a le m̂ pb IMP

D'où <u>L_1 n'est pas algébrique</u>