Langages Formels
Série 9 - Lemme de l'étoile
Hors-Contexte

24 Novembre 2025

Pensez à justifier vos réponses.

Rappel - Lemme de l'étoile Hors-Contexte :

Le lemme de l'étoile (ou lemme de pompage) pour les langages Hors-Contexte s'utilise similairement au premier lemme de l'étoile (pour les langages réguliers) : on ne peut pas s'en servir pour montrer que des langages sont Hors-Contextes (car il s'agit d'une condition nécessaire mais pas suffisante), mais il est en revanche extrêmement utile pour montrer que des langages ne sont pas Hors-Contextes.

On va donc se servir de la contraposée du lemme de l'étoile Hors-Contexte pour montrer qu'un langage n'est pas Hors-Contexte. On rappelle donc l'énoncé de la contraposée :

Soit L un langage infini. Si

$$\forall N \in \mathbb{N}$$

$$\exists z \in L \text{ avec } |z| \geq N$$

$$\forall u, v, w, x, y \text{ tels que } z = uvwxy,\ |vwx| \leq N,\ |vx| \geq 1$$

$$\exists i \in \mathbb{N},\ z' = uv^iwx^iy \notin L$$

Alors L n'est pas Hors-Contexte.

Remarque - On a donc la même structure (∀ N, ∃ mot, ∀ decoupage, ∃ i) que pour le premier lemme : la différence c'est que la forme et les conditions des découpages est bien plus complexe, et traiter tous les cas de découpages sera donc plus difficile (voir exemple ci-après).

Exemple

Soit le langage $L_1 = \{w \in \{a, b, c\} \mid |w|_a = |w|_b = |w|_c\}$.
On va montrer qu'il n'est pas Hors-Contexte en appliquant le lemme (sa contraposée), on doit donc montrer que :
$\forall N$: Pour tout N,
$\exists z = a^N b^N c^N \in L_1$, $|w| \geq N$: Il existe un mot z appartenant au langage, avec une longueur supérieure ou égale à N. Ici, on choisit $z = a^N b^N c^N$. Il appartient bien au langage, et sa longueur est de $3N$.
$\forall u, v, w, x, y$, tels que $z = uvwxy$, $|vwx| \leq N, |vx| \geq 1$: On doit construire tous les découpages possibles de z en cinq facteurs u, v, w, x, y,
tels que v, w et x contiennent au maximum N caractères au total,
et v et x contiennent à eux deux au moins un caractère (au moins l'un des deux n'est pas vide).
$\exists i \in N \mid z' = uv^i wx^i y \notin L_1$: Pour chaque découpage, on doit montrer qu'il existe une valeur i, telle que si on pompe i-fois les parties v et x, alors le mot généré, z', n'appartient pas au langage.
Il est impossible d'exprimer tous les découpages possibles respectant ces conditions en une seule formule facilement, on va donc séparer en plusieurs cas ces découpages possibles, et à chaque fois, on donne un i permettant de construire un z' sortant du langage :

1. Le premier cas correspond à celui où v et x contiennent uniquement des "a" :
   $u = a^j$
   $v = a^k$
   $w = a^l$
   $x = a^m$
   $y = a^{N-(j+k+l+m)} b^N c^N$.
   Avec $j, k, l, m \geq 0$, et $k + l + m \leq N$, et $k + m \geq 1$.
   Soit $i = 2$, on obtient alors un mot $z' = a^{N+k+m} b^N c^N \notin L_1$
   Pour cette partie des découpages, on a bien un i tel que le mot z' sort du langage. Mais il faut le montrer pour tous les découpages possibles d'après les conditions du lemme. On continue donc avec les autres cas possibles.

2. Le second cas pour ce langage est celui où v et x contiennent uniquement des "b" :
   $u = a^N b^j$
   $v = b^k$
   $w = b^l$
   $x = b^m$
   $y = b^{N-(j+k+l+m)} c^N$
   Avec $j, k, l, m \geq 0$, et $k + l + m \leq N$, et $k + m \geq 1$.
   Pour $i = 2$, on obtient un mot de la forme $z' = a^N b^{N+k+m} c^N \notin L_1$.

3. Troisième cas avec v et x qui contiennent uniquement des "c" :
   $u = a^N b^N c^j$
   $v = c^k$

$w = c^l$
$x = c^m$
$y = c^{N-(j+k+l+m)}$
Avec $j, k, l, m \geq 0$, et $k + l + m \leq N$, et $k + m \geq 1$.
Pour $i = 2$, on obtient un mot de la forme $z' = a^N b^N c^{N+k+m} \notin L_1$.

4. Ensuite, il reste les cas où v et y contiennent plusieurs types de caractères. On commence par les cas où v et y contiennent à la fois des "a" et des "b". On a trois sous-cas possibles (on peut se contenter de décrire seulement v et x : cela suffit à définir l'entièreté du découpage, car les blocs u, w et y sont implicitement définis par v et x) :
   a) $v = a^j$ et $x = b^k$
   (implicitement, u contient ce qui est avant v, w ce qui est entre v et x, et y ce qui est après x)
   Pour $i = 2$, on obtient alors un mot de la forme $w' = a^{N+j} b^{N+k} c^N \notin L_1$.
   b) $v = a^j b^k$ et $x = b^m$
   Pour $i = 2$, on obtient alors un mot de la forme $w' = a^N b^k a^j b^{N+m} c^N \notin L_1$.
   c) $v = a^j$ et $x = a^k b^m$
   Pour $i = 2$, on obtient alors un mot de la forme $w' = a^{N+j} b^m a^k b^N c^N \notin L_1$.
   Pour tous les cas où v et x contiennent à la fois des "a" et des "b", on a donc bien un i tel que le mot z' sort du langage.

5. Enfin, il reste les cas où v et x contiennent à la fois des "b" et des "c". De même que pour le cas "a et b à la fois", on a trois sous-cas :
   a) $v = b^j$ et $x = c^k$
   Pour $i = 2$, on obtient alors un mot de la forme $w' = a^N b^{N+j} c^{N+k} \notin L_1$.
   b) $v = b^j c^k$ et $x = c^m$
   Pour $i = 2$, on obtient alors un mot de la forme $w' = a^N b^N c^k b^j c^{N+m} \notin L_1$.
   c) $v = b^j$ et $x = b^k c^m$
   Pour $i = 2$, on obtient alors un mot de la forme $w' = a^N b^{N+j} c^m b^k c^N \notin L_1$.
   De nouveau, pour tous ces cas, on a bien un i qui crée un mot z' sortant du langage.

On a donc bien montré que pour tout découpage possible d'après le lemme, il existe un i tel que z' sort du langage. Donc L_1 n'est pas Hors-Contexte.

Exercices

Pour chaque application du lemme, pour chaque forme de découpage possible, indiquez <u>au minimum</u> v et x, puis le i choisi et le mot généré z' !

1. En utilisant le second lemme de l'étoile, pour langages Hors-Contexte (c.f. ci-dessus), montrez que les langages suivants ne sont pas hors-contexte :
   - $L_2 = \{M^n E^{2n} W^{3n} \mid n \geq 0\}$

- $L_3 = \{(0^j 1^k)^n \mid j, k, n \geq 0\}$

2. Pour chacun des langages suivants, montrez s'ils sont réguliers, hors-contextes, ou aucun des deux :

   - $L_4 = \{a^n b^i c^n d^j \mid i, j \geq 0,\ n \geq 1\}$
   - $L_5 = \{w \in \{a, b, c, d\}^* \mid |w|_b = |w|_d\}$
   - $L_6 = L_4 \cap L_5$