LA3 — **Le lemme de l'étoile en action**

Le « lemme de l'étoile » (ou lemme d'itération, ou de la pompe, ou de pompage, etc.) est une propriété des langages reconnaissables. Informellement, il montre que dans tout mot u assez grand appartenant à un langage reconnaissable L, il existe une partie du mot qu'on peut répéter un nombre arbitraire de fois tout en restant dans L.

$$\begin{array}{c} u \in L \\ \begin{array}{|c|c|c|}\hline \quad x \quad & \ y \ & \qquad\qquad z \qquad\qquad \\ \hline\end{array} \\ \downarrow \\ \begin{array}{|c|c|c|c|c|c|}\hline \quad x \quad & \ y \ & \ y \ & \ \cdots \ & \ y \ & \qquad\qquad z \qquad\qquad \\ \hline\end{array} \ \in L\end{array}$$

Il existe plusieurs versions de ce lemme selon ce qu'on peut choisir, ou non, concernant la zone à l'intérieur de laquelle on itère. Nous ne citerons que la version « usuelle » qui est souvent suffisante (si, lors de l'examen, un énoncé plus fort est utile, il sera rappelé).

> **Lemme** *(version habituelle)*
> Soit L un langage reconnaissable. Il existe un entier N tel que pour tout mot $u \in L$, si $|u| \geq N$ alors il existe des mots x, y, z tels que
>
> $$\begin{cases} u = xyz \\ \text{et } y \neq \varepsilon \\ \text{et } |xy| \leq N \end{cases} \qquad \text{et} \quad \forall k \in \mathbb{N}, xy^k z \in L.$$

Dans cette version, on garantit seulement que y n'est « pas trop loin » du début du mot, mais attention, à part cette contrainte, on ne choisit pas le découpage x, y, z : il nous est imposé.

Ce lemme sert principalement à montrer qu'un langage **n'est pas** reconnaissable : il suffit en effet de montrer qu'il ne satisfait pas la propriété du lemme. Voyons un **exemple d'utilisation**.

> Soit $L = \{u \in \{a,b\}^\star \mid |u|_a = |u|_b\}$ (l'ensemble des mots ayant autant de a que de b). Montrons que L n'est pas reconnaissable.
> On raisonne par l'absurde. Si L était reconnaissable, alors il existerait un entier N donné par le lemme de l'étoile. Soit $u = a^N b^N$: puisque $|u| \geq N$ et $u \in L$, il existe x, y, z comme dans l'énoncé du lemme. Puisque $|xy| \leq N$ et que les N premières lettres de u sont des a, on a $y = a^i$ pour un certain $i > 0$. Donc
>
> $$xy^k z = a^{(N-i)+ik} b^N.$$
>
> Ainsi, pour $k = 0$, $xy^k z = a^{N-i} b^N \notin L$ puisque ce mot n'a pas autant de a que de b. C'est une contradiction avec le lemme de l'étoile, donc L n'est pas reconnaissable.

$\longrightarrow$ *On n'écrit donc **pas** des choses comme « $\forall k, u = xy^k z$ » puisque ce n'est vrai que pour $k = 1$, ou encore « prenons $x = a$, $y = a^3 b$ et $z = b^3$ » puisque, d'une part, on ne choisit pas le découpage x, y, z, et d'autre part, il faut un mot $u \in L$ de taille supérieure à N qui peut être grand et qu'on ne maîtrise pas.*