Définitions :

Structure d'un langage formel : un langage formel possède :

- Un vocabulaire (exemple si (a > b) alors a=a+1 ;)
- Des règles de grammaire entièrement connues (exemple : la règle de la structure si ... alors et les relations entre les différentes expressions de l'instruction).

Définitions des notions de base :

1. **Symbole** : c'est une brique élémentaire, un atome.
2. **Alphabet** (vocabulaire) : c'est un ensemble fini et non vide de *symboles*. (exemple : en français on utilise l'alphabet latin).
   Les alphabets sont décrits comme des ensembles de symboles (entre accolades) composés de plusieurs caractères.
   Exemple :
   Soit l'alphabet A = {1,2,3, a,d,k} ; A = {a,b,c,….,z}

3. **Mot** (chaîne) : un mot sur un alphabet A est une suite finie de symboles $m_1m_2 \ldots m_n$ de cet alphabet.
   Exemple **: 12adk ; adzc ;**

   La longueur d'une chaine est le nombre de symboles la composant, on le note |m|.
   Exemple : |12adk| = 5 ; | adzc | = 4

4. Le **mot vide :** c'est le mot de longueur 0, c'est-a-dire ne contenant aucun symbole. Il est souvent note **ε**.
   Exemple : A={a,b,c} : un alphabet
   abbc, aaaa, bc, **ε** sont des mots sur l'alphabet A, respectivement de longueur : 4,4,2 et 0.

5. Les symboles représentés par des signes de ponctuation seront entre quottes.

Opérations sur les chaines :

- **1- La concaténation** : soit u et v deux chaines, alors la concaténation de u et v se fait sans l'utilisation d'un opérateur, juste on juxtapose les deux chaines uv.
  Exemple : soit u = abaa et v = bab alors :
  uv = et abaabab vu = bababaa
  La concaténation n'est pas commutative (uv est différent de vu).

  La chaine vide **ε** est neutre pour la concaténation. (u ε = ε u = u).

- 2- On note x^n, $n \in N$ la concaténation de n symboles identiques.
  Exemple : Le mot aaaaa est noté a^5.

- 3 – Soit U une chaine, On note u^n, $n \in N$ la concaténation de n chaines identiques.

- 4 - Le symbole U(i) désigne le i-ème symbole de la chaîne u.
  Exemple : soit le mot u = aabb , le u(3) = b.

- 5- Le symbole |U|x désigne le nombre de fois de répétition du symbole x dans la chaine U.

- 6 - Chaine miroir : La chaine miroir d'une chaine U est la chaine formée par la suite des symboles composant U mais pris dans l'ordre inverse. si u = 0101 alors miroir(u) = 1010.

- 7 - Un mot est dit palindrome s'il est égal à son miroir.

- **8 Préfixe** :

Soit $w = w_1 w_2 \ldots\ldots w_l$ un mot sur un alphabet A.

$$\varepsilon, w_1, w_1 w_2, \ldots\ldots, w_1 \ldots w_{l-1}, w_1 \ldots w_l = w$$

sont les préfixes de w. Un préfixe de w différent de ε et de w est dit propre.
Exemple :
Soit le mot w = abbcc sur l'alphabet A= {a,b,c}.
Les mots : ε, a, ab, abb, abbc, abbcc sont les prefixes du mot w.

Les mots : a, ab, abb, abbc sont les prefixes propres du mot w.

- **9 Suffixe** : on procède à définir les suffixes de la même manière que les préfixes sauf que cette fois ci on commence de droite à gauche.

Exemple :

Soit le mote w = abbcc sur l'alphabet A= {a,b,c}.

Les mots : ε , c, cc, bcc, bbcc, abbcc sont les suffixes du mot w.

Les mots : c, cc, bcc, bbcc sont les suffixes propres du mot w.

- **10 Facteur** :

Soit $1 \leq i \leq j \leq l$ avec i et j des entiers. Le mot $w_i \ldots w_j$ est un facteur du mot w. On note l'ensemble des facteurs de w par : Fact(w).

Exemple :

Soit le mot w = abbcc
sur l'alphabet A= {a,b,c}.
Les mots : abb, bbc, bcc sont des facteurs du mot w

Le mot : aa n'est pas un facteur du mot w.

- **11 Sous mots** : les sous-mots sont construits à partir de plusieurs symboles non nécessairement contigus, mais dans lesquels leur ordre d'apparition est respecté.

Exemple :

Soit le mot w = abbcc sur l'alphabet A= {a,b,c}.
Les mots : abc, ac, acc, abcc sont des sous-mots du mot w.

Notations :

- Soit l'alphabet V :
    - V^0 : désigne la chaine vide **ε**

- $V^1 = V$: désigne l'ensemble des chaines de longueur 1
- $V^2 = V.V$: désigne la concaténation de V avec V (Tous les mots de longueur 2)
- $V^n = V....V$: désigne la concaténation n fois de V
- $V^* = \cup V^i \; i \geq 0$: c'est l'ensemble de tous les mots sur V (c'est un ensemble infini)
- <u>Exemple</u> : soit l'alphabet A = {0,1}, donner A^1 A^2 et A^3

  $A^0=\{\varepsilon\}$;

  $A^1 = A =\{0, 1\}$;

  $A^2 = \{0,1\}^2 = \{00, 01, 10, 11\}$;

  $A^3 = \{0,1\}^3 = \{000, 001, 010, 011, 100, 101,110, 111\}$;