

## CHAPITRE 2 : ANALYSE LEXICALE

L'analyse lexicale est la 1<sup>ère</sup> étape de la compilation.

Le texte est lu caractère par caractère

C'est l'unique module qui lit le texte source

Elle reconnaît les unités lexicales ( = tokens) qui sont les **mots** du langage

Elle présente les mots à l'analyseur syntaxique

**Il s'agit d'un automate fini**

### Principales unités lexicales :

- Caractères simples : + - \* / ; ( )
- Caractères doubles : := >= ++
- Mots clés : if, then, for, class, while
- Unités lexicales génériques :
  - o Identificateurs (nom de variables / procédures)
  - o Constantes

L'analyseur lexical supprime tous les espaces, tabulations, retour chariots

Il supprime aussi tous les commentaires

Il affiche également les erreurs :

- o Identificateur mal formé
- o Constante trop grande
- o Caractère inconnu

Son autre tâche est de coder les unités lexicales et construire un lexique

*Exemple :*

Soit le codage :

Caractère	Code
+	1
-	2
*	3
/	4
:=	5
Identificateur	7
Constante	8

Soit le texte source :

```
/* Un exemple */
x := y * 2 + x
```

Analyseur lexical

Liste de mots :

(7, 'x'), 5, (7, 'y'), 3, (8,2), 1, (7, 'x')

Si on crée un lexique :

Code lexical	Caractère
1	'x'
2	'y'

La liste de mots devient alors : (7,1), 5, (7,2), 3, (8,2), 1, (7,1)