

Théorie des Langages

Samia Mazouz

FEI-USTHB

Département Informatique

2018-2019

Campusvirtuel.usthb.dz



LANGUAGE

Un langage est un ensemble de phrases ou mots satisfaisant certaines règles.

Exemple simple de langage serait qu'une phrase est de la forme :

- Phrase = Sujet Verbe
- Sujet = Pronom
- Pronom = il ou elle
- Verbe = écoute ou regarde
- L'ensemble des phrases formées est :
il écoute, elle écoute, il regarde, elle regarde.



LANGUAGE

De très nombreux langages existent, par exemple :

- Les langages naturels (anglais, arabe, ...).
- Les langages utilisés en mathématiques et en logique mathématique (langage de la logique propositionnel, langage de la logique des prédicats).
- Les langages informatiques qui sont la principale motivation de ce cours.



LANGAGE

Intérêt des langages permettent :

- De communiquer soit
 - Entre Nous
 - Avec la machine
 - Entre machines
- De décrire des systèmes, des documents,
- De formaliser des problèmes afin de les résoudre



THEORIE DES LANGAGES

La théorie des langages trouve son origine dans la tentative de formalisation du langage naturel par le linguistique Noam Chomsky 1956.



Par la suite, il y a eu le besoin des informaticiens de l'époque :

- de décrire de **manière finie certains langages infinis**
- de mettre en place les premiers langages de programmation (Fortan, Algol,...).



THÉORIE DES LANGAGES

- Besoin de décrire formellement un langage \Rightarrow Grammaire
- Reconnaître formellement un langage \Rightarrow Automate

Grammaire formelle : permet d'**engendrer** les mots du langage, **généralement infini**, en utilisant un **ensemble fini de règles**.

Chomsky a décomposé les grammaires formelles en catégories de pouvoir d'expression croissant, c'est-à-dire en groupes successifs pouvant chacun générer une variété de langages plus large que le groupe précédent.

On parle de **hiérarchie de Chomsky**.

On verra particulièrement :

- les grammaires régulières
- et les grammaires à contexte libre.



THÉORIE DES LANGAGES

○ **Automate** : est une machine abstraite qui permet de **reconnaître** les mots d'un langage.

Etant donné un mot fourni en entrée, l'automate **lit les symboles du mot** un par un et va d'état en état selon les transitions. Le mot lu est soit accepté par l'automate soit rejeté.



THÉORIE DES LANGAGES

La théorie des langages établit des correspondances entre :

- descriptions **génératives (grammaires)**
- **et descriptions analytiques (automates).**

Chaque niveau de grammaire est strictement isomorphe à un type particulier d'automate:

- Les grammaires régulières correspondent aux automates d'états finis.
- Les grammaires à contexte libre correspondent aux automates à pile.



CONVERTIR UN « PROBLÈME » EN UN LANGAGE

La résolution du problème revient à l'analyse d'un élément de ce langage.

Un automate qui résout le problème prend en entrée un mot et décide s'il est accepté ou non.

Par exemple, le problème de savoir si un entier N est pair (test de parité) peut se traduire comme suit :

- On représente tous les entiers naturels par des chaînes binaires (écriture en base 2).
- Dans ce langage, les mots représentant des nombres pairs forment un sous-ensemble.
- Le problème du test de parité consiste alors à savoir si la chaîne binaire représentant un nombre N appartient à ce sous-ensemble ou non.
- Un automate approprié prend en entrée une chaîne binaire et l'accepte précisément lorsqu'elle représente un nombre pair.



APPLICATION

- **En compilation** : reconnaître qu'un programme d'un langage donné est syntaxiquement correcte.
- **Recherche des motifs** :
 - Recherche d'une chaîne dans un fichier (Recherche sous word)
 - Recherche dans un répertoire (Recherche sous Dos/Linux)
 - Recherche dans le web (Moteur de recherche)
- **Réaliser des contrôles** : Vérifier qu'une donnée entrée par un utilisateur a bien le format spécifié par exemple une adresse IP.
- **Industries de la langue** :
 - Traitement automatique des langues naturelles
 - Correction orthographique automatique
 - Génération automatique de textes
 - Réalisation de dictionnaire électronique (gain de mémoires)
- **Vérification des circuits électronique** .
- **Biologie (génomique)** : Recherche de gènes particuliers dans une chaîne. En génomique, on dispose d'un alphabet de 4 lettres correspondant aux bases de l'ADN) ou de protéomes (textes sur un alphabet de 20 lettres correspondant aux acides aminés qui constituent les protéines).

