

Théorie des Langages et Automates

2025-2026

Plan du cours

- ▶ Automates finis et langages réguliers
 - ▶ Notion de langage
 - ▶ Automates finis déterministes
 - ▶ Automates finis non déterministes + Déterminisation
 - ▶ Lemme de Pumpage ◦ Grammaires régulières ◦ Expressions régulières
 - ▶ Equivalence entre automates finis, grammaires régulières et expressions régulières
 - ▶ Limites des langages réguliers
- ▶ Automates à pile, langages non contextuels
 - ▶ Automates à pile
 - ▶ Grammaires non contextuelles
 - ▶ Equivalence automates à pile et grammaires non contextuelles
 - ▶ Lemme de pompage
- ▶ Machines de Turing
 - ▶ Définitions
 - ▶ Langages Turing acceptables
 - ▶ Problème de l'arrêt

Introduction générale

- ▶ En linguistique et en informatique, on parle de théorie des langages pour décrire les langages formels.
- ▶ Plus généralement, la théorie des langages concerne tout langage fini ou infini qui peut être spécifié par une méthode ou un mécanisme fini, et explicite, qui permet de le **produire** ou de **l'analyser**.
- ▶ La théorie des langages ne se préoccupe pas du côté sémantique (le sens), mais plutôt **syntactique** (la forme) d'un langage.

Introduction générale

- ▶ On définit un **langage** notamment grâce aux **grammaires** et on analyse les **mots** d'un langage grâce aux **automates**.
- ▶ Ce cours va s'articuler autour de ces trois concepts que sont les **langages**, les **grammaires** et les **automates**.
- ▶ La théorie des langages est une base pour un certain nombre de disciplines informatique dont principalement la compilation.

Chapitre 1: Automates finis et langages réguliers

Plan

- ▶ Rappel sur la théorie des ensembles
- ▶ Notion de langage
- ▶ Automates finis déterministes
- ▶ Automates finis non déterministes + Déterminisation
- ▶ Lemme de Pumpage
 - Grammaires régulières
 - Expressions régulières
- ▶ Equivalence entre automates finis, grammaires régulières et expressions régulières
- ▶ Limites des langages réguliers

Rappel sur la théorie des ensembles

Définitions :

- ▶ Un ensemble est une collection d'objets sans répétition.
- ▶ Si un objet appartient à un ensemble A , on dit qu'il est élément de cet ensemble et l'on note $x \in A$.
- ▶ On distingue un ensemble particulier noté \emptyset qui ne contient aucun élément.
- ▶ Un ensemble est noté par $\{...\}$ où les pointillés indiquent les éléments de l'ensemble ou une caractéristique de ses éléments

Rappel sur la théorie des ensembles

- ▶ Il existe principalement trois moyens pour définir un ensemble :
 - ▶ **Définition par extension:** consiste à donner tous les éléments d'un ensemble.
Exemple: $\{0, 1, 2, 3, 4\}$.
 - ▶ **Définition par compréhension:** définit les éléments d'un ensemble par les propriétés qui les définissent. on écrira $\{x / x \text{ vérifie une propriété } P(x)\}$.
Exemple : $\{n \in \mathbb{N} / n \% 2 = 0\}$ définit les nombres entiers pairs.
 - ▶ **Définition par induction :** définit un ensemble par certains éléments triviaux et des règles d'induction permettant de retrouver d'autres éléments en fonction de ceux déjà connus. On écrit: $x \in A \Rightarrow f(x) \in A$ tel que f est un moyen permettant de construire d'autres éléments en fonction de l'argument.
Exemple : l'ensemble des entiers peut être représenté comme suit :
 $\mathbb{N} = \{0; x \in \mathbb{N} \Rightarrow (x + 1) \in \mathbb{N}\}$.

Opérations sur les ensembles

► Comparaison des ensembles:

► **Inclusion:** $A \subseteq B$, si $\forall x \in A, x \in B$. On dira alors que A est un sous ensemble de B .

► Exemple : $\{a, b, c\} \subseteq \{a, b, c, d\}$.

► **Egalité:** Deux ensembles A et B sont égaux si $A \subseteq B$ et $B \subseteq A$.

Opérations sur les ensembles

► Construction d'autres ensembles:

Soit A et B deux sous ensembles de Ω :

- L'union $A \cup B$: comporte tout élément appartenant à A ou B
- L'intersection $A \cap B$: comporte tout élément appartenant à A et B
- La différence $A - B$: comporte tout élément appartenant à A et qui n'appartient pas à B
- Le complément, noté $\bar{A} = \Omega - A$
- Le produit cartésien $A \times B$: est l'ensemble des paires (a, b) telles que $a \in A$ et $b \in B$.
- L'ensemble des parties de A , noté 2^A : est l'ensemble de tous les sous-ensembles de A

Opérations sur les ensembles

► **Exemple : $A = \{a, b\}$, $B = \{a, c\}$, $\Omega = \{a, b, c\}$:**

- $A \cap B = \{a\}$;
- $A \cup B = \{a, b, c\}$;
- $A - B = \{b\}$;
- $\overline{A} = \{c\}$;
- $A \times B = \{(a, a), (a, c), (b, a), (b, c)\}$;
- $2^A = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$

Opérations sur les ensembles

► Mesures sur les ensembles:

- La mesure la plus utilisée et celle de la cardinalité. Elle est notée $\text{card}(A)$ et représente le nombre d'éléments de A .
- Si l'ensemble est infini alors sa cardinalité est ∞ .
- $\text{card}(A \times B) = \text{card}(A) \cdot \text{card}(B)$
- $\text{card}(2A) = 2\text{card}(A)$.

Notion de langage

- ▶ La théorie des langages utilise un certain nombre de concepts ainsi qu'une certaine terminologie.
- ▶ Nous allons d'abord définir certaines notions capitales qui sont les bases de cette théorie:
 - ▶ Alphabet
 - ▶ Mot
 - ▶ Langage

Alphabet

► Définitions:

- Un **alphabet** est un ensemble fini de symboles.
- Un **symbole** est une entité abstraite qui ne peut pas être définie formellement comme les lettres, les chiffres, ..ect

► Exemples :

- $A = \{0, 1\}$, l'alphabet binaire
- $B = \{a, b, c\}$
- $C = \{\text{if}, \text{then}, \text{else}, a, b\}$, un alphabet du langage C
- $F = \{\rightarrow, \leftarrow, \uparrow, \downarrow\}$ un alphabet de fleches

Mot

► Définition:

- Un mot sur l'alphabet X est une séquence finie et ordonnée d'éléments de l'alphabet.
- C'est une concaténation de lettres.

► Exemples:

- abbac et ba sont deux mots de l'alphabet $\{a,b,c\}$.
- 00,01,10010 sont des mots de l'alphabet $\{0,1\}$

► Notations:

- Le mot vide, noté ε , correspond à la suite de symboles vides.
- L'ensemble des mots construits sur un alphabet X est un ensemble infini noté X^*
- L'ensemble des mots construits sur un alphabet X et qui ne contient pas le mot vide est un ensemble infini noté X^+

Mot

► **Longueur d'un mot:**

Soient X est un alphabet, $w \in X^*$, et x est un des symboles constituant le mot w :

- $|w|$ est la longueur du mot w
- $|w|_x$ est le nombre de l'occurrence de x dans le mot w

► **Exemples:**

- $|abbac| = 5$
- $|ba| = 2$
- $|\epsilon| = 0$.
- $|ababac|_a = 3$

Mot: Opérations sur les mots

► Concaténation

Soient deux mots $w, w' \in X^*$, la concaténation de w et w' est définie comme la juxtaposition de w et w' . Elle est notée par $w.w'$ ou bien ww' .

► Exemple:

- $X=\{0,1\}, w=10, w'=00$
 $w.w'=1000$
 $w'.w=0010$

Mot: Opérations sur les mots

► Puissance d'un mot

Soit un alphabet X , $w \in X^*$ et $n \in \mathbb{N}$, la puissance de w est donnée comme suit:

$$w^n = \begin{cases} \varepsilon & \text{si } n = 0 \\ w & \text{si } n = 1 \\ ww^{n-1} = w^{n-1}w & \text{si } n > 1 \end{cases}$$

► Exemple:

Soit $X = \{a, b\}$ et $w = aba$

- $w^0 = \varepsilon$

- $w^1 = w.\varepsilon = w = aba$

- $w^2 = w.w = ww = abaaba$

Mot: Opérations sur les mots

► Factorisation d'un mot

Soit un alphabet X , $w \in X^*$

- u est un facteur gauche (préfixe) de $w \Leftrightarrow \exists v \in X^*$ tel que $w = uv$
- u est un facteur droit (suffixe) de $w \Leftrightarrow \exists v \in X^*$ tel que $w = vu$
- u est un préfixe propre de $w \Leftrightarrow \exists v \in X^+$ tel que $w = uv$
- u est un suffixe propre de $w \Leftrightarrow \exists v \in X^+$ tel que $w = vu$

► Exemple:

Soit $X = \{a, b\}$ et $w = babb$:

- les préfixes de w sont : $b, ba, bab, babb$
- les suffixes de w sont : $b, bb, abb, babb$
- les préfixes propres de w sont : b, ba, bab
- les suffixes propres de w sont : b, bb, abb

Mot: Opérations sur les mots

► Inverse d'un mot ou Miroir

le miroir d'un mot $w = a_1 a_2 \dots a_n$ est le mot noté $w^R = a_n \dots a_2 a_1$ obtenu en inversant les symboles de w .

► Exemple:

► Soit $w = abbc$ un mot de l'alphabet $x = \{a, b, c\}$:
 $w^R = cbba$

► Remarques:

- Le miroir d'un mot composé d'un seul symbole est le mot lui-même
- $\varepsilon^R = \varepsilon$
- Un mot est un palindrome si $w^R = w$, exple: $(aba)^R = aba$

Langage

- ▶ **Définition:** Un langage est un ensemble (fini ou infini) de mots définis sur un alphabet donné. Il peut être défini par extension, par compréhension ou par induction

- ▶ **Exemples:**

Soit l'alphabet $X=\{a,b\}$

- ▶ \emptyset est le langage vide. Il ne contient aucun mot.
- ▶ $\{\epsilon\}$ est un langage.
- ▶ $\{a, b, aa, bb, aba\}$ est un langage
- ▶ $\{aa, ab, ba, bb\}$ est le Langage des mots de longueur 2
- ▶ $\{w \in X^* / w=a^n \text{ tel que } n > 0\}$ est un langage

- ▶ **Remarques:**

- ▶ Un langage sur un alphabet X peut être fini ou infini
- ▶ \emptyset est un langage défini sur n'importe quel alphabet

Opérations sur les langages

- ▶ Les langages sont des ensembles. On peut alors leur appliquer n'importe quelle opération ensembliste telles que l'union, l'intersection, et la complémentarité.
- ▶ Soient L , $L1$ et $L2$ trois langages définis sur l'alphabet X , on définit les opérations suivantes :
 - L'union : $L1 + L2 = \{w / w \in L1 \vee w \in L2\}$
 - L'intersection : $L1 \cap L2 = \{w / w \in L1 \wedge w \in L2\}$
 - Le complément : $L = \{\text{tous les mots } w \text{ sur } x / w \notin L1\}$
 - La concaténation : $L1.L2 = \{w / \exists u \in L1, \exists v \in L2 : w = uv\}$
 - Exposant : $L^n = \{w / \exists u_1, u_2, \dots, u_n \in L : w = u_1 u_2 \dots u_n\}$

Opérations sur les langages

► **Egalité:**

Soit $L1$ et $L2$ deux langages engendré sur X^* .

On dit que les deux langages sont égaux ($L1 = L2$) si et seulement si $\forall w \in L1$ alors $w \in L2$ et $\forall w \in L2$ alors $w \in L1$.

Opérations sur les langages

► **Union:**

Soit X un alphabet et soit deux langages $L1$ et $L2$ engendrés sur X^* .

L'union de $L1$ et $L2$ est l'ensemble des mots de $L1$ et $L2$:

$$L1 \cup L2 = \{w \in X^* / w \in L1 \text{ ou } w \in L2\}$$

► L'union est:

- Associative
- Commutative
- Élément neutre, le langage vide: $L \cup \emptyset = L$
- Élément absorbant, le vocabulaire X^* : $L \cup X^* = X^*$
- Notée $+$ dans la théorie des langages: $L1 \cup L2 = L1 + L2$

Opérations sur les langages

- ▶ Intersection:

L'intersection de L_1 et L_2 est l'ensemble des mots qui appartiennent à la fois à L_1 et L_2 : $L_1 \cap L_2 = \{w \in X^* / w \in L_1 \text{ et } w \in L_2\}$

- ▶ Elle est:

- ▶ Associative
- ▶ Commutative
- ▶ Element neutre est X^* : $L \cap X^* = L$
- ▶ Element absorbant, le langage vide: $L \cap \emptyset = \emptyset$

Opérations sur les langages

- ▶ Produit ou concaténation:

C'est le résultat de la concaténation de chaque mot de L_1 avec chaque mot de L_2 :

$$L_1.L_2 = \{w = u.v \mid u \in L_1, v \in L_2\}$$

- ▶ Le produit est:

- ▶ Associatif
- ▶ Il n'est pas commutatif
- ▶ Element neutre est $\{\epsilon\}$
- ▶ Element absorbant, le langage vide: \emptyset
- ▶ Le produit est distributif par rapport à l'union: $L_1.(L_2 \cup L_3) = (L_1.L_2) \cup (L_1.L_3)$

Opérations sur les langages

► **Puissance:**

La puissance d'un langage est défini comme suit:

$$\text{► } L^n = \begin{cases} \{\varepsilon\} & \text{si } n=0 \\ L & \text{si } n=1 \\ L.L^{n-1}=L^{n-1} & \text{si } n > 1 \end{cases}$$

Opérations sur les langages

► Complémentarité:

Soit un langage L défini sur l'alphabet X .

La complémentarité de L noté \overline{L} , est définie par:

$$\overline{L} = X^* \setminus L = \{w \in X^* / w \notin L\}$$

Opérations sur les langages

► Fermeture positive:

La fermeture positive de L est notée L^+ .

$$L^+ = \bigcup_{i=1}^{\infty} L^i$$

► Fermeture de Kleene:

La fermeture étoile de L est notée L^* .

$$L^* = \bigcup_{i=0}^{\infty} L^i = L^0 + \bigcup_{i=1}^{\infty} L^i = \{\varepsilon\} + L^+$$

Opérations sur les langages

► **Exemples:**

Soit $X=\{0,1\}$, $L1=\{00,11,01\}$ et $L2=\{01,10\}$

► $L1+L2=\{00,11,01,10\}$

► $L1 \cap L2=\{01\}$

► $L1.L2=\{0001,0010,1101,1110,0101,0110\}$

► $L2^*=\{\epsilon\}+\{L2\}+\{L2.L2\}+\dots+\{L2.L2.L2\dots L2.L2.L2\}$

Les grammaires

Introduction:

- ▶ Une grammaire est l'ensemble des règles à suivre pour parler et écrire correctement une langue.
- ▶ Ce même principe s'applique également à la théorie des langages, où en suivant les **règles de production** d'une grammaire, il est possible de créer un langage spécifique.

Grammaire

Définitions:

- ▶ Une grammaire est un ensemble de règles de production qui sont utilisées pour engendrer un langage
- ▶ Ensemble de règles pour générer les mots du langage sont sous la forme de règles de réécriture
 - Remplacer une séquence de symboles par une autre séquence
- ▶ Mots générés = mots obtenus à partir d'un symbole spécial appelé symbole de départ ou axiome

Grammaire

Exemple: Considérons la phrase suivante :

La vieille dame regarde la petite fille

- ▶ Peut-on construire une grammaire qui permet de générer cette phrase?
- ▶ Alphabet : $A = \{ \text{la, vieille, petite, dame, fille, regarde} \}$
- ▶ Structure de la phrase :
 - ▶ Un groupe sujet (article, adjectif, nom)
 - ▶ Un verbe
 - ▶ Un groupe complément d'objet (article, adjectif, nom)

Grammaire

Exemple:

► Règles de production:

1. $\langle \text{Phrase} \rangle \rightarrow \langle \text{Sujet} \rangle \langle \text{Verbe} \rangle \langle \text{Complément} \rangle$
2. $\langle \text{Sujet} \rangle \rightarrow \langle \text{Groupe Nominal} \rangle$
3. $\langle \text{Complément} \rangle \rightarrow \langle \text{Groupe Nominal} \rangle$
4. $\langle \text{Groupe Nominal} \rangle \rightarrow \langle \text{Article} \rangle \langle \text{Nom} \rangle$
5. $\langle \text{Groupe Nominal} \rangle \rightarrow \langle \text{Article} \rangle \langle \text{Adjectif} \rangle \langle \text{Nom} \rangle$
6. $\langle \text{Article} \rangle \rightarrow \text{la}$
7. $\langle \text{Nom} \rangle \rightarrow \text{dame} \mid \text{fille}$
8. $\langle \text{Adjectif} \rangle \rightarrow \text{vieille} \mid \text{petite}$
9. $\langle \text{Verbe} \rangle \rightarrow \text{regarde}$

Grammaire

► Définition formelle:

Une grammaire **G** est un quadruplet (N, T, P, S) tels que:

- **N**: ensemble fini de symboles non terminaux
- **T**: ensemble fini de symboles terminaux
- $N \cap T = \emptyset$
- **S**: symbole non terminal appelé axiome (point de départ de la dérivation) ;
- **P**: ensemble fini de règles de production de la forme:
 $\alpha \rightarrow \beta$ tel que $\alpha \in (N \cup T)^+$ et $\beta \in (N \cup T)^*$

La notation $\alpha \rightarrow \beta$ est appelée une **dérivation** et signifie que α peut être remplacé par β .

Grammaire

► Dérivation directe:

Un mot w' dérive directement d'un mot w qu'on note $w \Rightarrow w'$ si une règle G est appliquée une fois pour passer de w à w' .

Càd: il existe une règle $\alpha \rightarrow \beta$ dans P telle que:

$w = u\alpha v$ et $w' = u\beta v$ avec $u, v \in (N \cup T)^*$

► Dérivation au sens général:

Un mot w' dérive d'un mot w qu'on note $w \rightarrow^* w'$, si on applique n fois les règles de G pour passer de w vers w' tel que $n \geq 0$.

Grammaire

► Exemple:

soit la grammaire: $G = (\{S\}, \{a, b\}, \{S \rightarrow aSb, S \rightarrow \epsilon\}, S)$, nous avons:

- **aSb** dérive directement de S et on écrit: $S \Rightarrow aSb$ car il existe une règle $S \rightarrow aSb \in P$
- **ab** dérive directement de aSb et on écrit: $aSb \Rightarrow ab$ car il existe une règle $S \rightarrow \epsilon \in P$
- **$S \rightarrow aSb \rightarrow ab$** est une dérivation de longueur 2
- **$S \rightarrow aSb \rightarrow aaSbb \rightarrow aabb$** est une dérivation de longueur 3
- **$S \rightarrow aSb \rightarrow \dots \rightarrow a^n b^n$** est une dérivation de longueur $n+1$

Langage engendré par une grammaire

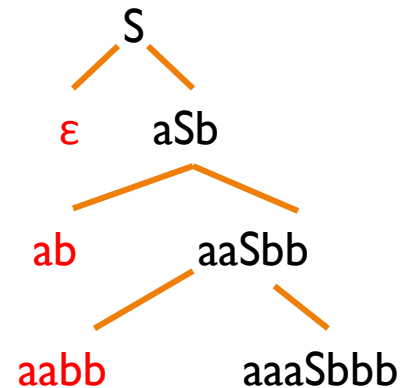
- ▶ **Définition:** Un langage engendré par une grammaire $G=(N,T,P,S)$ est l'ensemble des mots obtenus en appliquant des séquences de dérivations à partir de l'axiome S .

On note: $L(G)=\{w \in T^*/S \Rightarrow^*_G w\}$

- ▶ **Exemples:**

Pour la grammaire: $G=(\{S\}, \{a, b\}, \{S \rightarrow aSb, S \rightarrow \epsilon\}, S)$:

- ▶ Mot minimal: ϵ
- ▶ La forme générale: $L(G)=\{a^n b^n / n \geq 0\}$



- ▶ Remarque: Une grammaire définit un seul langage par contre un même langage peut être engendré par plusieurs grammaires différentes.

Grammaires équivalentes

- ▶ Deux grammaires sont équivalentes si elles engendrent le même langage.

G équivalente à $G' \Leftrightarrow L(G)=L(G')$

- ▶ **Exemples:**

- ▶ $G=(\{S,A,B\}, \{a,b\}, \{S \rightarrow aS \mid ABb, A \rightarrow Aa \mid a, B \rightarrow b\}, S)$

- ▶ $G'=(\{S,A,B\}, \{a,b\}, \{S \rightarrow aS \mid Aa, A \rightarrow Aa \mid bB, B \rightarrow b\}, S)$

- ▶ \Rightarrow On trouve que $L(G)=L(G')=\{a^k b^2 \mid k \geq 1\}$

Classification des grammaires

▶ **Hiérarchie de Chomsky**

Selon la classification de Chomsky, les grammaires sont regroupées en quatre types en fonction de la forme de leurs règles de production:

- ▶ Grammaire Syntagmatique -Type 0
- ▶ Grammaire Monotone – Type 1
- ▶ Grammaire Algébrique (hors contexte) – Type 2
- ▶ Grammaire Régulière – Type 3

Classification des grammaires

► **Grammaire Syntagmatique -Type 0**

Dite aussi grammaire sans restriction (grammaire générale): si toutes ses règles sont de la forme générale suivante:

$\alpha \rightarrow \beta$ tel que $\alpha \in (N \cup T)^+$ et $\beta \in (N \cup T)^*$

► **Exemple:**

$G = (\{S\}, \{a,b,c\}, \{S \rightarrow aS \mid Sb \mid c, aSb \rightarrow Sa \mid bS\}, S)$

Classification des grammaires

▶ **Grammaire monotone-Type I:**

si toutes ses règles sont de la forme:

$\alpha \rightarrow \beta$ avec $|\alpha| \leq |\beta|$ tel que $\alpha \in (N \cup T)^+$ et $\beta \in (N \cup T)^*$

Exception: axiome $\rightarrow \varepsilon$ peut appartenir à P

▶ **Exemple:**

$G = (\{S, R, T\}, \{a, b, c\}, \{S \rightarrow \varepsilon \mid abc, R \rightarrow aRTb \mid aTb, Tb \rightarrow bT, Tc \rightarrow cc\}, S)$

Classification des grammaires

▶ **Grammaire algébrique –type2:**

si toutes ses règles sont de la forme:

$A \rightarrow \beta$ avec $A \in N$ et $\beta \in (N \cup T)^*$

▶ **Exemple:**

▶ $G = (\{S\}, \{a, b\}, \{S \rightarrow aSb \mid \varepsilon\}, S)$

Mots: $\varepsilon, ab, aabb, aaabbb, \dots, a^n b^n$

▶ $G = (\{S\}, \{0, 1\}, \{S \rightarrow 0S0 \mid 1S1 \mid \varepsilon \mid 0 \mid 1\}, S)$

Mots: $00, 11, 010, 101, 000, 111, 101101, \dots$

C'est la grammaire des Palindromes

Classification des grammaires

▶ **Grammaire régulière- Type 3**

Dite aussi grammaire régulière si elle est régulière à gauche ou bien à droite

- ▶ Une grammaire G dite régulière à gauche si toutes ses règles sont de la forme:

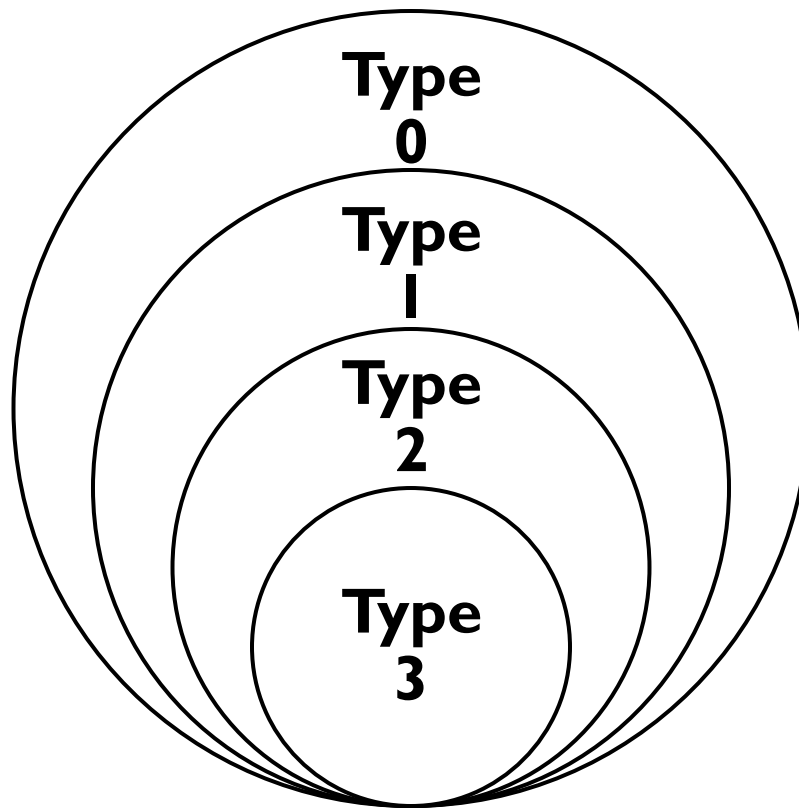
$A \rightarrow Bw$ ou $A \rightarrow w$ avec $A, B \in N$ et $w \in T^*$

- ▶ Une grammaire G dite régulière à droite si toutes ses règles sont de la forme:

$A \rightarrow wB$ ou $A \rightarrow w$ avec $A, B \in N$ et $w \in T^*$

Classification des grammaires

- ▶ Il existe une relation d'inclusion entre les types de grammaires selon la figure suivante :



Classification des grammaires

- ▶ Le type retenu pour une grammaire est le plus petit qui satisfait les conditions.
- ▶ Pour trouver la classe d'un langage on procède cependant comme suit :
 - ▶ Chercher une grammaire de type 3 qui le génère, si elle existe, le langage est de type 3 (ou **régulier**)
 - ▶ Sinon, chercher une grammaire de type 2 qui le génère, si elle existe, le langage est de type 2 (ou **algébrique**)
 - ▶ Sinon, chercher une grammaire de type 1 qui le génère, si elle existe, le langage est de type 1 (ou **contextuel**)
 - ▶ Sinon, le langage est de type 0.