

Chapitre I : Introduction

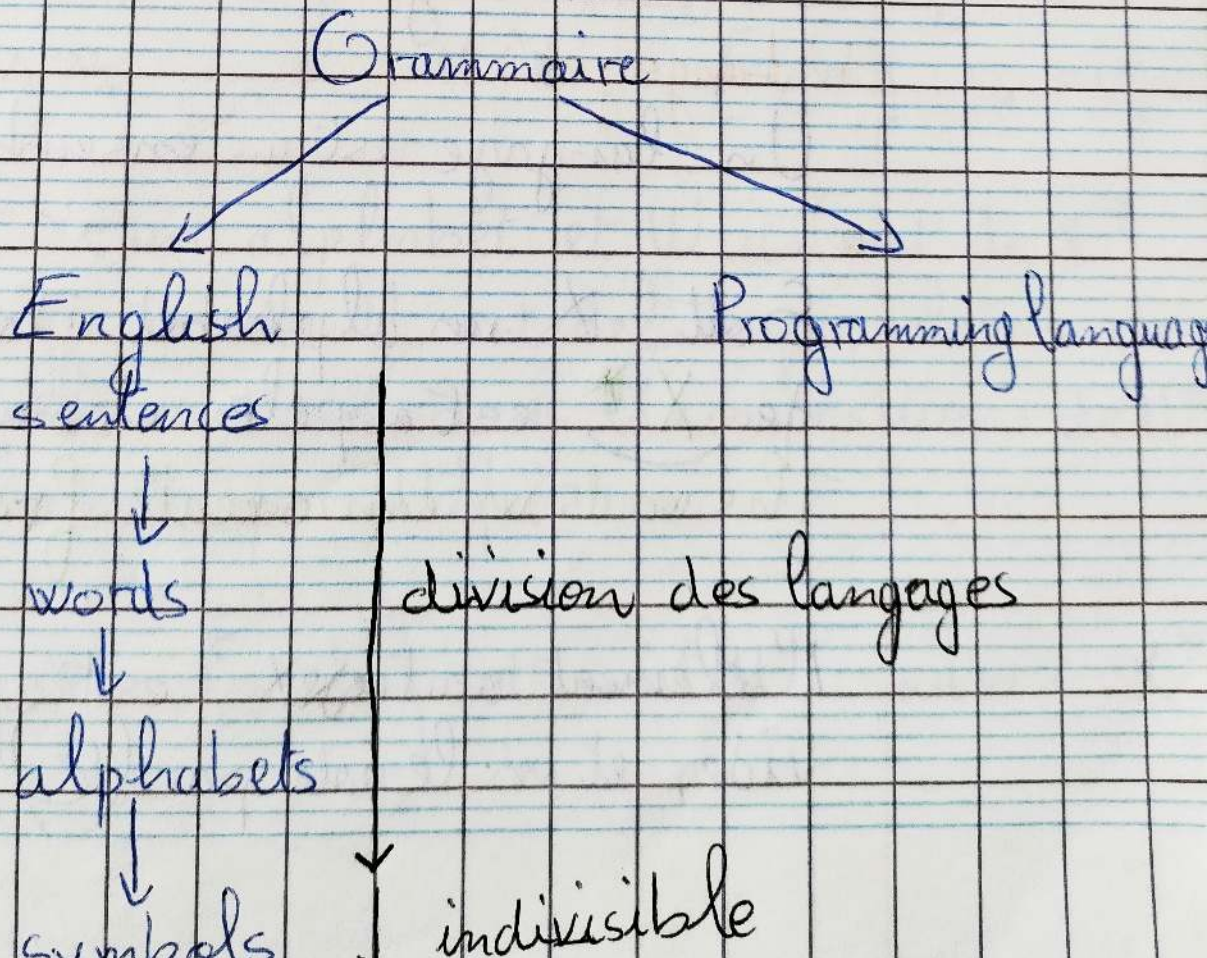
Objectif

Le but de ce chapitre est d'introduire les concepts de bases de la théorie des langages et compilations.

Il s'agit de définir les 3 notions suivantes :

- les langages
- les grammaires
- les analyses syntaxique

1) Concept de base de la théorie de langages



1.2) Définitions :

a) Alphabet :

Un alphabet est un ensemble ^{finie} non vide de symboles. Il est noté par l'ensemble $X(\Sigma)$.

b) Les symboles :

Les symboles sont des éléments indivisibles qui vont servir pour construire des mots.

Exp : 26 lettres de l'alphabet.

c) Les mots :

Un mot est un ensemble de symboles. Il est noté généralement W .

d) Langages :

Un langage est un ensemble des mots.

e)

Soit X un alphabet. Tout élément de X^* est appelé un mot.

The words we can create from X .

L'élément neutre X^* est appelé le mot vide, et on le note par (ϵ) .

La loi de composition sur X^* est appelée concatenation ou produit.

Exp: Soit $X = \{0, 1\}$

$W = 001$

est un mot construit sur l'alphabet X .

$W \in X^*$

f)

2 mots sont égaux ssi elles sont formées de la même suite de symboles.

Soit X un alphabet et soient $u \in X^*$ et $v \in X^*$, tel que $u = u_1 \cdot u_2 \cdot \dots \cdot u_n$ et $v = v_1 \cdot v_2 \cdot \dots \cdot v_n$ et $\forall i \in [1, n]; u_i = v_i$

g)

Soit X un alphabet et W un mot de X^* .
la longueur de W est définie par le nb des symboles entrant dans la composition de W .
On la note par $|W|$.

Exp:

Soit $X = \{a, b\}$ et soit $W = aab \in X^*$

$$\Rightarrow |W| = 3$$

$$* \text{ Soit } X = \{aa, bb\}$$

$$W = aab \notin X^*$$

$$W = bb aa bb \in X^*; |W| = 3$$

h) On note par $|W|_a$: le nb d'occurrences de symbole a dans W .

Exp:

$$\text{Soit } X = \{a, b\}$$

$$W = aab \in X^*$$

$$\Rightarrow |W|_a = 2$$

$$\text{Soit } X = \{aa, bb\}$$

$$W = aabbaa$$

$$|W|_{aa} = 2$$

i) Soit X un alphabet, $u \in X^*$ et $v \in X^*$

$$u = u_1 \cdot u_2 \cdots u_n \text{ avec } u_i \in X$$

$$\text{et } v = v_1 \cdot v_2 \cdots v_n \text{ avec } v_i \in X$$

La concaténation de u et v est notée:

$$u \cdot v = u_1 u_2 \dots u_n \cdot v_1 v_2 \dots v_m$$

~~NB:~~

NB:

* Soit $W = u \cdot v$ alors $|W| = |u| + |v|$
 * Soit $W \in X^*$
 alors $(u \cdot v) \cdot w = u \cdot (v \cdot w)$

* Soit ϵ l'élément neutre alors $u \epsilon = \epsilon u = u$

j) Soit X un alphabet
 et soit $W \in X^*$, $u \in X^*$ est un facteur de
 W (sous chaîne). ssi $W = u' \cdot u \cdot u''$, u' et $u'' \in X^*$
 $X = \{a, b, x, y\}$

Rqs:

* Si u' et $u'' \neq \epsilon$ alors dans ce cas nous
 désignons par u un facteur propre de W
 * Si u'' est égale ϵ alors $W = u \cdot u''$
 et dans ce cas nous désignons par u un facteur
 gauche.

* Si $u'' = \epsilon$ alors $W = u' \cdot u$ et dans ce cas nous désignons par u un facteur droit.

1.3) Notion de grammaire et d'analyse syntaxique

Considérons par exemple la phrase suivante :
Le vieux chat attrape le petit rat.

Le but est de construire une grammaire qui permet de produire cette phrase.

Il faut tout d'abord préciser l'alphabet de symbole dans notre nous pouvons prendre comme alphabet l'ensemble.

$X = \{ \text{le, chat, vieux, attrape, rat, petit} \}$

Voyons maintenant comment ces symboles seront assemblés. Dans la structure de la phrase. On peut distinguer un groupe sujet, un verbe est un grp complément d'objet.

Les grps sujet et complément d'objet sont eux même des groupes nominaux.

Un grp nominale est formé en article suivi par d'adjectif

qui peuvent être
ou suivie

Ainsi, on peut considérer par exemple la grammaire
suivante pouvant produire une phrase elle comporte

11 règles de grammaire :

- 1 - $\langle \text{phrase} \rangle \rightarrow \langle \text{groupe sujet} \rangle \langle \text{verbe} \rangle$
- 2 - $\langle \text{groupe sujet} \rangle \rightarrow \langle \text{groupe nominal} \rangle$
- 3 - $\langle \text{groupe co} \rangle \rightarrow \langle \text{groupe nominal} \rangle$
- 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 - $\langle \text{article} \rangle \rightarrow \text{Le}$
- 7 - $\langle \text{nom} \rangle \rightarrow \text{chat}$
- 8 - $\langle \text{nom} \rangle \rightarrow \text{rat}$
- 9 - $\langle \text{adjectif} \rangle \rightarrow \text{vieux}$
- 10 - $\langle \text{adjectif} \rangle \rightarrow \text{petit}$
- 11 - $\langle \text{verbe} \rangle \rightarrow \text{attrape}$

- Le pt de départ de la grammaire s'appelle l'axiome (start) [$\langle \text{phrase} \rangle$]
- Les variables de la grammaire sont [phrase, groupe sujet, verbe]
- Les règles 1 à 5 sont des règles syntaxique
- Le reste des règles sont des règles complètes terminales.

... Le langage engendré par la grammaire est l'ensemble de toutes les phrases qu'on peut produire à partir de l'action en utilisant les règles de grammaire.

On se basant sur les règles de la grammaire on peut produire par exemple les phrases suivantes :

~~1) < phrases > < groupe sujet > < verbe >~~

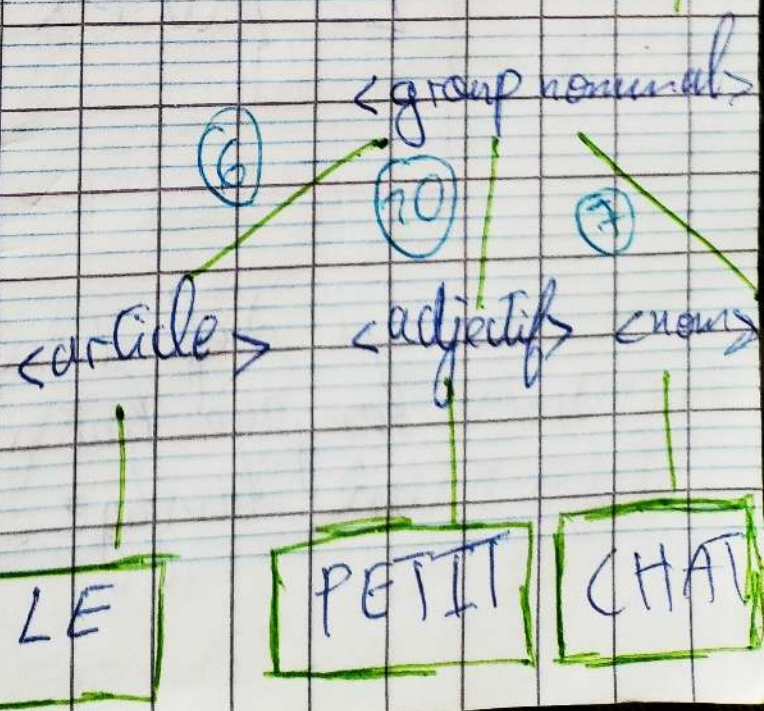
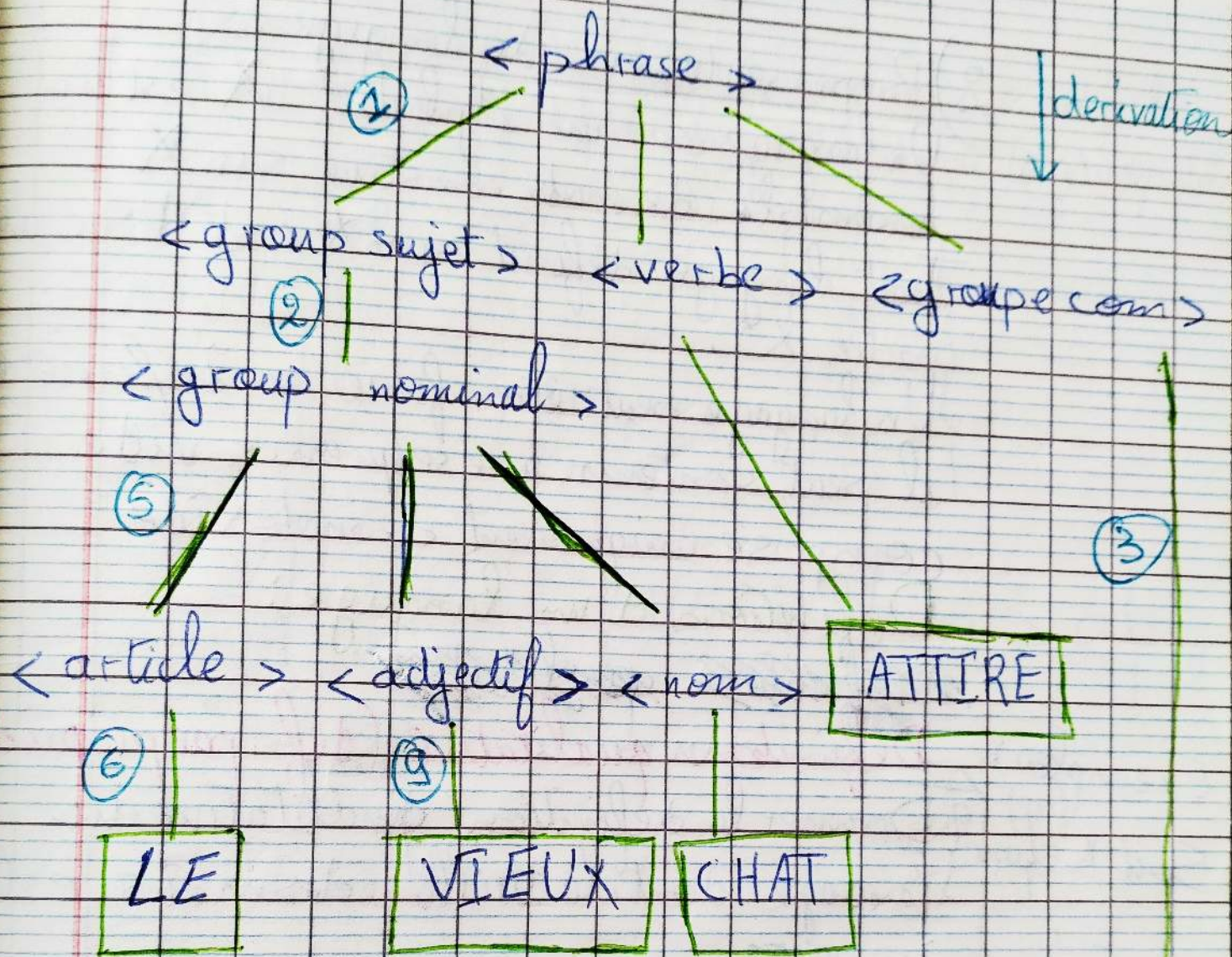
① LE RAT ATTRAPE LE PETIT CHAT

② LE VIEUX CHAT ATTRAPPE LE PETIT CHAT.

Exp:

On peut associer à chaque ainsi produite un arbre de dérivation où figurent les variables auxquelles on a appliqué les règles de la grammaire

L'arbre de dérivation est représenté comme suit



2) Représentation de langage :

Un langage sur un alphabet X est un ensemble de mots construits sur X .

Tout langage défini sur X est donc une partie X^* .

Un langage peut être fini ou infini et il peut contenir une séquence vide composée uniquement de mots vides.

Definition d'un langage :

~~Definition qualitative~~

exemple : $\leftarrow$ **Definition qualitative (def. par propriété)**
Soit $X = \{a, b\}$ Donner la définition qualitative du langage dont la taille des mots est impaire.

$$L = \{ w \in X^* \mid |w| = 2k + 1, k \geq 0 \}$$

propriété de langage

La définition des éléments d'un langage fait référence aux éléments de n° langage.

Definition recursive:

Donner une définition recursive de la langue a dont les mots sont constitués de suite des symboles a .

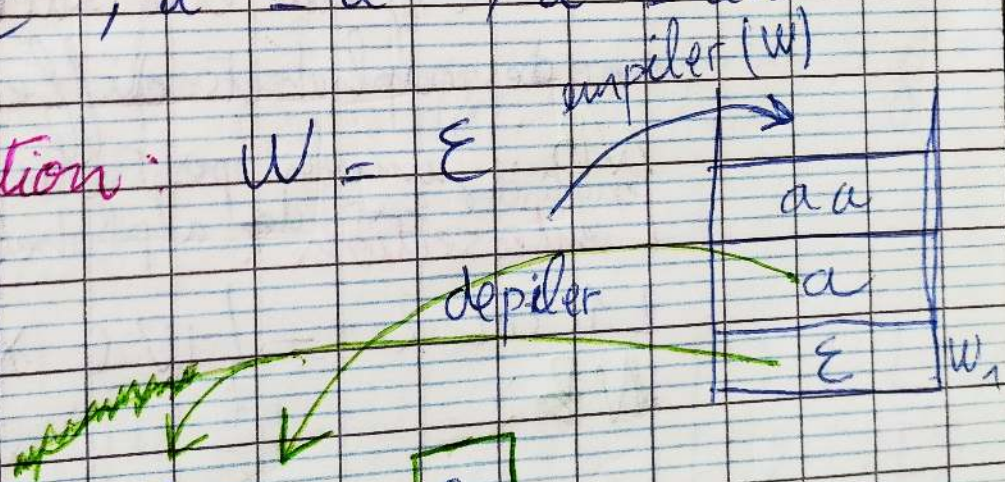
Exemple: Soit $X = \{a, b\}$
 $L = \{\epsilon, a, aa, aaa, \dots\}$

Definition propriétés:

$$L = \{w \in X^* \mid w = a^n, n \geq 0\}$$

$$a^0 = \epsilon, a^1 = a, a^2 = aa$$

Initiation: $w = \epsilon$



1 It: $w = \epsilon$ $a = \boxed{a}$

2 It: $w = a.a = aa$
 $L = \{w \in X^* \mid w = \epsilon \text{ ou } w = w_1.a, w_1 \in L\}$

Rq: w_1 est un ancien mot construit sur X et appartient à L

2.2) Operation sur les langages:

Soit X un alphabet et soient L_1 et L_2 deux langages engendré de X^*

L'union:

$$L_1 \cup L_2 = \{w \in X^* / w \in L_1 \text{ ou } w \in L_2\}$$

L'union peut être sur des langages engendrés sur des alphabets différents.

NB: l'union peut être appliquée par des langages engendrés sur des alphabets différents.

L'intersection:

$$L_1 \cap L_2 = \{w \in X^* / w \in L_1 \text{ et } w \in L_2\}$$

NB:

Le langage doit être de m alphabet (du présence)

La différence:

$$L_1 - L_2 = \{w \in X^* / w \in L_1 \text{ et } w \notin L_2\}$$

Produit ou concatenation:

$$L_1 \cdot L_2 = \{ w \in x^* \mid w = uv; u \in L_1 \text{ et } v \in L_2 \}$$

Exp: Soient $L_1 = \{ \epsilon, ab, ba \}$

$$L_2 = \{ xyz, xz \}$$

$$L_1 \cdot L_2 = \{ xyz, xz, abxz, baxz \}$$

$$L_2 \cdot L_1 = \{ xyz, xyzab, xyzba, xz, xzab, xzba \}$$

Autres propriétés:

- $L_1 \cdot (L_2 \cup L_3) = L_1 \cdot L_2 \cup L_1 \cdot L_3$
- $(L_1 \cup L_2) \cdot L_3 = L_1 \cdot L_3 \cup L_2 \cdot L_3$
- Si $L_1 \subseteq L_2$ alors $L_1 \cdot L_3 \subseteq L_2 \cdot L_3$
- $L^0 = \{ \epsilon \}$
- $L^1 = L$
- $L^2 = L \cdot L$

$$\begin{aligned}
 \bullet L^n &= L \cdot L \cdot \dots \cdot L \quad (n \text{ fois}) \\
 \bullet L^* &= \bigcup_{n \in \mathbb{N}} L^n = L^0 \cup L^1 \cup L^2 \cup \dots \cup L^n
 \end{aligned}$$

Remarque

L'étoile de Kleen d'un langage L^* donné est le langage obtenu en réunissant tout les langages L^n avec $(n \geq 0)$

$$L^+ = \bigcup_{n \geq 1} L^n = L^1 \cup L^2 \cup \dots$$

3) Les langages réguliers : régularisation

Un langage L sur X est obtenu ~~en peut~~ par récursivité :

~~* Le langage est ob~~

* A partir des langages de base

- Langage vide

- Langage $\{\epsilon\}$

- Langage de la forme $\{a\}; a \in X$

* et à l'aide des opérations :

- Union
- concaténation
- étoile de Kleen

Remarque :

Si P et Q sont 2 ensembles langages réguliers sur X alors les ensembles suivants sont réguliers :

- * $P \cup Q$
- * $P \cdot Q$
- * P^* ou Q^*
- * $P^* \cup Q^*$
- * $P^* \cdot Q^*$

Expression régulière :

Une expression régulière est une expression décrivant des langages réguliers engendrés sur un alphabet X

- $\emptyset$: l'ensemble vide
- a : ($a \in X$) : expression régulière décrivant l'ensemble $\{a\}$

- Si p et q sont 2 expressions régulières décrivant les 2 ensembles réguliers P et Q

Q alors l'expression suivante sont
régulières

$$P \cup Q \rightarrow P \cup Q$$

$$P \cdot Q \rightarrow P \cdot Q$$

$$P^* \rightarrow (P^*)$$

Exemple : $P^* \text{ ou } Q^* \rightarrow (P^* \text{ ou } Q^*)$

Soit $X = \{a, b, c\}$

Le langage L des mots commençant
par a est régulier, il peut être décrit
par une expression régulière de la forme

$$L = a \cdot (a \cup b \cup c)^*$$

$$(a \cup b \cup c)^0$$

$$(a \cup b \cup c)^1$$

$$(a \cup b \cup c)^2$$

,

.

.

.

$$(a \cup b \cup c)^n$$

Ré :

Soit $X = \{a, b\}$

- * L'expression régulière $a \cup b$: représente le langage formé par des mots a ou b
- * L'expression régulière $(a \cup b)^*$ représente le langage des mots qu'on peut former avec a et b (toutes les combinaisons possibles)

$$(a \cup b)^* = (a \cup b)^0 \cup (a \cup b)^1 \cup (a \cup b)^2 \cup \dots$$

$$L = \{ \epsilon, a, b, aa, ab, ba, bb, \dots \}$$

Manipulation des expressions régulières:
 Les opérations de l'union, de concaténation
 l'étoile de Kleene sont appelées des opérations
 régulières ayant la priorité suivantes.

priorité (*) > priorité (.) > priorité (U)

Loi de simplification de ER:

$$a) \alpha \cup (\beta \cup \theta) = (\alpha \cup \beta) \cup \theta$$

$$b) \alpha \cup \beta = \beta \cup \alpha$$

$$c) \alpha \cup \epsilon = \alpha$$

$$d) \alpha \cup \alpha = \alpha$$

$$e) \alpha (\beta \theta) = (\alpha \beta) \theta$$

$$f) \epsilon \alpha = \alpha \epsilon = \alpha$$

$$g) \alpha (\beta \cup \theta) = \alpha \beta \cup \alpha \theta$$

$$h) (\alpha \cup \beta) \theta = \alpha \theta \cup \beta \theta$$

$$i) \epsilon \cup \alpha \alpha^* = \epsilon \cup \alpha^+ = \alpha^+$$