Université de Rouen	Année en cours
Faculté des sciences	Théorie des langages
Licence Informatique $3^{\text{ème}}$ année	Résumé de cours pour le TD n° 1

Les langages

1 Définitions de base

Un *alphabet* est un ensemble fini de symboles appelés *lettres*. L'alphabet est souvent noté Σ.

Exemple :
$\Sigma = \{a, b, c\}$
$\Sigma = \{0, 1\}$ alphabet binaire
$\Sigma = \{+, -, *, /\}$ alphabet des opérateurs arithmétiques
$\Sigma = \{a, c, g, t\}$ alphabet des nucléotides

On travaillera généralement avec l'alphabet $\{a, b, c, \ldots\}$.
Un *mot* sur un alphabet donné est une suite de lettres de cet alphabet. On n'utilisera que des mots finis (suite finie de lettres). Les mots sont souvent notés u, v, w, x, y, z.

Exemple :
$u = aba$, $v = 00$, $w = acacgat$, $\ldots$

La *longueur* d'un mot est le nombre de lettres qui le composent. On la note $|u|$. On note ε ou 1 le mot vide.

Exemple :

$$\begin{array}{lcl} |\varepsilon| & = & 0 \\ |a| & = & 1 \\ |aaba| & = & 4 \end{array}$$

On note $|u|_a$ le *nombre d'occurrences* de a dans le mot u.

Exemple :

$$\begin{array}{lcl} |\varepsilon|_a & = & 0 \\ |a|_a & = & 1 \\ |aaba|_a & = & 3 \end{array}$$

Soit Σ un alphabet. On note Σ^* *l'ensemble de tous les mots de longueur finie* que l'on peut construire sur cet alphabet, y compris le mot vide.

On définit sur cet ensemble l'opération appelée *concaténation* et notée $\cdot$ de la manière suivante : si $u = aba$ et $v = bbab$ alors $u \cdot v = aba \cdot bbab = ababbab$.

On omet souvent le $\cdot$ de la concaténation et on note $u \cdot v$ simplement par uv. D'autre part, $\cdot$ a les propriétés suivantes :

$\varepsilon \cdot u = u \cdot \varepsilon = u$ pour tout u dans Σ^* (ε est élément neutre pour $\cdot$)
$(u \cdot v) \cdot w = u \cdot (v \cdot w)$ ($\cdot$ est associative)

La loi $\cdot$ est donc une loi de composition interne associative possédant un élément neutre ε et $(\Sigma^*, \cdot, \varepsilon)$ a une structure de monoïde. On l'appelle le *monoïde libre.*

On appelle *préfixe* ou *facteur gauche* (resp. *suffixe* ou *facteur droit*) d'un mot un début (resp. une fin) du mot.

Exemple : si $u = ababb$, l'ensemble de ses préfixes est $\{\varepsilon, a, ab, aba, abab, ababb\}$

le mot w est un *préfixe propre* de u si c'est un préfixe de u différent de u.

Exemple : le mot *aba* est un préfixe propre de *ababb*

le mot w est un *facteur* de u si il existe deux mots x et y tels que $u = x \cdot w \cdot y$ et c'est un *facteur propre* de u si x ou y est différent de ε.

Exemple :
si $u = ababb$, l'ensemble des facteurs de u est $\{\varepsilon{,}a{,}ab{,}aba{,}abab{,}ababb{,}abb{,}b{,}ba{,}bab{,}babb{,}bb\}$

On appelle *langage sur l'alphabet* Σ toute partie de Σ^*.

2 Opérations sur les langages

Nous ne donnerons ici que les opérations de base sur les langages.

Soient X et Y deux langages sur Σ^*

On note	$X \cup Y$ ou $X + Y$	l'union de X et Y
	$X \cap Y$	l'intersection de X et Y
	$\Sigma^* \setminus X$	le complémentaire de X dans Σ^*
	$X \cdot Y$	le produit de concaténation de X par Y

Le produit de concaténation est l'ensemble des mots obtenus en concaténant un mot de X et un mot de Y. On a les règles suivantes :
$X \cdot \emptyset = \emptyset \cdot X = \emptyset$ ($\emptyset$ est absorbant pour $\cdot$)
$\{\varepsilon\} \cdot X = X \cdot \{\varepsilon\} = X$ ($\{\varepsilon\}$ est élément neutre pour $\cdot$)

On peut alors définir X^n pour tout $n \geq 0$:
$X^0 = \{\varepsilon\}$
$X^n = X^{n-1} \cdot X$ pour tout $n > 0$. C'est la concaténation de n mots de X.
$X^* = X^0 + X^1 + X^2 + \ldots + X^n + \ldots$
$X^+ = X^1 + X^2 + \ldots + X^n + \ldots$
On en déduit que $\emptyset^* = \{\varepsilon\}$

Comme on définit les préfixes (resp.suffixes, resp. facteurs) d'un mot, on peut définir les préfixes (resp. suffixes, resp. facteurs) d'un langage. C'est l'ensemble des préfixes (resp.

suffixes, resp. facteurs) des mots du langage.

Exemple :
$X = \{ab, bab\}$ et $Y = \{bb, ca\}$ sur $\Sigma = \{a, b, c\}$

- $X \cup Y = \{ab, bab, bb, ca\}$.
- $X \cup \Sigma^* = \Sigma^*$.
- $X \cap Y = \emptyset$.
- $X \cap \Sigma^* = X$.
- $X \cdot Y = \{abbb, abca, babbb, babca\}$.
- $X^2 = \{abab, abbab, babab, babbab\}$.
- $\Sigma^2 = \Sigma \cdot \Sigma = \{aa, ab, ba, bb, ac, ca, bc, cb, cc\}$ est l'ensemble des mots de longueur 2 sur Σ.
- $\Sigma^n = \Sigma \cdot \Sigma \ldots \Sigma$ est l'ensemble des mots de longueur n sur Σ.
- $\Sigma^* = \Sigma^0 + \Sigma^1 + \ldots \Sigma^n + \ldots$ est l'ensemble des mots de longueur quelconque sur Σ.
- $\Sigma^+ = \Sigma^1 + \ldots \Sigma^n + \ldots$ est l'ensemble des mots de longueur strictement positive sur Σ.
- Préfixe$(X) = \{\varepsilon, a, ab, b, ba, bab\}$.

Un langage peut être donné explicitement par la liste de ses mots quand son cardinal le permet, sinon on donne une caractérisation des mots du langage.

Exemple :

- $X_1 = \{ab, bab\}$
- $X_2 = \{w \in \Sigma^*$ tel que w commence par $a\} = \{a\} \cdot \Sigma^*$
- $X_3 = \{w \in \Sigma^*$ tel que w contient $ab\} = \Sigma^* \cdot \{ab\} \cdot \Sigma^*$
- $X_4 = \{w \in \Sigma^*$ tel que w commence par un a suivi d'un nombre quelconque de $b\} = \{a\} \cdot \{b\}^*$
- $X_5 = \{w \in \Sigma^*$ tel que w commence par aba et $|w| <= 5\} = \{aba$, $abaa$, $abab$, $abaaa$, $abaab$, $ababa$, $ababb\}$
- $X_6 = \{w \in \Sigma^*$ tel que w ne contient que des $a\}$
  $= \{\varepsilon, a, aa, aaa, aaaa, \ldots\} = \{\varepsilon\} \cup \{a\} \cup \{aa\} \cup \{aaa\} \cup \{aaaa\} \ldots$
  $= \{a\}^0 + \{a\}^1 + \{a\}^2 + \{a\}^3 + \{a\}^4 + \ldots = \{a\}^*$
- $X_7 = \{w \in \Sigma^*$ tel que w contient une suite quelconque de a puis le même nombre de $b\} = \{\varepsilon, ab, \ldots, a^n b^n, \ldots\} = \{a^n b^n, n \geq 0\}$

2.1 Quelques propriétés

La concaténation est distributive par rapport à l'union

$$(X + Y) \cdot Z = (X \cdot Z) + (Y \cdot Z)$$

Les opérations unaire $*$ et $+$ conservent l'inclusion et sont idempotentes

$$X \subseteq Y \Rightarrow X^* \subseteq Y^* \text{ et } X^+ \subseteq Y^+$$
$$(X^*)^* = X^* \text{ et } (X^+)^+ = X^+$$

De plus

$$(X \cdot Y)^* \cdot X = X \cdot (Y \cdot X)^*$$