Université Toulouse III Année 2018-2019
TD Mathématiques discrètes **Feuille n° 2**

Notions sur les langages

Exercice 1 - Langages préfixes. On dit que $\mathcal{L}$ est un langage préfixe si

$$\text{pour tous } u, v \in \mathcal{L} \text{ on a } u \neq v \Longrightarrow u \text{ n'est pas un préfixe de } v.$$

Autrement dit si aucun des mots de $\mathcal{L}$ n'apparaît dans le début d'un autre mot de $\mathcal{L}$.

1. Donner un exemple de langage préfixe.
2. Montrer que si $\varepsilon \in \mathcal{L}$ et $\mathcal{L}$ est préfixe alors $\mathcal{L} = \{\varepsilon\}$;
3. Soient $\mathcal{L}_1$ et $\mathcal{L}_2$ deux langages préfixes, montrer que $\mathcal{L}_1 \cap \mathcal{L}_2$ est préfixe ;
4. Soient $\mathcal{L}_1$ et $\mathcal{L}_2$ deux langages préfixes, montrer que $\mathcal{L}_1.\mathcal{L}_2$ est préfixe ;
5. Soient $\mathcal{L}_1$ et $\mathcal{L}_2$ deux langages préfixes, montrer que $\mathcal{L}_1 \cup \mathcal{L}_2$ n'est pas forcément préfixe.

Exercice 2 - Quelques exemples. Définissez les langages suivants.

1. Le langage des identifiants du langage de programmation Python (qui commencent avec une lettre ou un *underscore* "_" et se poursuivent avec un nombre arbitraire de lettres, de chiffres ou underscores). Par exemple : `a12`, `x`, `y_3`, `_ab` mais pas `5_z`.
2. Le langage des nombres décimaux selon l'écriture habituelle (c'est à dire sans 0 à gauche, sauf si le nombre est 0). Par exemple : `123` mais pas `0123`.
3. Le langage des nombres hexadécimaux selon la notation utilisée dans le langage C, à savoir précédés d'un `0x` et avec des majuscules `A ... F` pour marquer les nombres de 10 à 15. Par exemple : `0x0`, `0x123`, `0xFA7`.

Exercice 3 - Monotonie. Montrez les propriétés de monotonie suivantes :

1. Si $L_1 \subseteq M_1$ et $L_2 \subseteq M_2$, alors $L_1.L_2 \subseteq M_1.M_2$.
2. Si $L \subseteq M$, alors pour tout n, $L^n \subseteq M^n$.
3. Si $L \subseteq M$, alors $L^* \subseteq M^*$.

Exercice 4 - Distributivité. Montrer ou invalider les propriétés de distributivité suivantes :

1. $L.(M_1 \cup M_2) = (L.M_1) \cup (L.M_2)$ et $(L_1 \cup L_2).M = (L_1.M) \cup (L_2.M)$
2. $(L.M)^* = L^*.M^*$
3. $(L \cup M)^* = L^* \cup M^*$

Exercice 5 - Opérateur de Kleene. Quelles égalités sont correctes ? Éventuellement, pensez à corriger l'équation pour qu'elle devienne correcte.

$$L^* = L.L^* \quad , \quad L^+ = L.L^* \quad , \quad L^* = L^*L^* \quad , \quad L^+ = L^+.L^+$$

Exercice 6 - Code. Un ensemble de mots X est un code si et seulement si pour tous entiers $n \geq 0$ et $p \geq 0$, pour tous mots, $x_1, \ldots, x_n, y_1, \ldots, y_p$ de X, l'égalité $x_1 \ldots x_n = y_1 \ldots y_p$ implique $n = p$ et $x_i = y_i$ pour tout i entre 1 et p.

1. Les ensembles suivants sont-ils des codes ?

$$\{a, b, ab\}, \ \{a, ab, bb\}, \ \{a, ab, ba\}, \ \{aa, ba, baa, bb, bba\}$$

2. Soit $\{u_1, \ldots, u_n\}$ un ensemble de n mots distincts de même longueur k. Montrer que c'est un code.
3. On dit que $X \subset \mathcal{A}^*$ est un ensemble préfixe si pour tous $x, y \in X$, si x est un préfixe de y alors $x = y$. Montrer que tout ensemble préfixe est un code.
4. Énoncer le théorème équivalent sur les suffixes.

Exercice 7 - Le lemme d'Arden. On considère $\mathcal{M}$ et $\mathcal{N}$ deux langages sur $\mathcal{A}$. On souhaite résoudre l'équation sur les langages $X = \mathcal{M}.X \cup \mathcal{N}$ où X est le langage inconnu.

1. Montrer que le langage $\mathcal{M}^*.\mathcal{N}$ est solution de l'équation.
2. Soit $\mathcal{L}$ une solution, montrer que pour tout $n \in \mathbb{N}$ on a

$$\mathcal{L} = \mathcal{M}^{n+1}.\mathcal{L} \cup \mathcal{M}^n.\mathcal{N} \cup \mathcal{M}^{n-1}.\mathcal{N} \cup \cdots \cup \mathcal{M}\mathcal{N} \cup \mathcal{N}.$$

3. En déduire que $\mathcal{M}^*.\mathcal{N} \subset \mathcal{L}$.
4. Montrer que si $\varepsilon \notin \mathcal{M}$ alors $\mathcal{M}^*.\mathcal{N}$ est la seule solution de l'équation.

Exercice 8 - Puissance d'un mot. Soient $u, v \in \mathcal{A}^*$. Montrer que les propositions suivantes sont équivalentes :

— $u.v = v.u$;
— il existe $w \in \mathcal{A}^*$ et $p, q \in \mathbb{N}$ tels que $u = w^p$ et $v = w^q$;
— il existe $n, m \in \mathbb{N}^*$ tels que $u^n = v^m$.

Notions sur les langages (Solutions)

Correction 1 1. $\mathcal{L} = \{a, ba, bba, bbba\}$ est un langage préfixe.

2. ε est préfixe de tout mot. Ainsi si $\varepsilon \in \mathcal{L}$ et $\mathcal{L}$ est préfixe alors ε ne peut être que préfixe de lui même dans $\mathcal{L}$. On en déduit que c'est le seul mot de $\mathcal{L}$.
3. Soient $\mathcal{L}_1$ et $\mathcal{L}_2$ deux langages préfixes. Soient $u, v \in \mathcal{L}_1 \cap \mathcal{L}_2$, comme $\mathcal{L}_1$ est préfixe, on en déduit que u n'est pas un préfixe de v. Ainsi $\mathcal{L}_1 \cap \mathcal{L}_2$ est préfixe.
4. Soient $\mathcal{L}_1$ et $\mathcal{L}_2$ deux langages préfixes et soient $u, v \in \mathcal{L}_1.\mathcal{L}_2$ tels que $u \neq v$. On peut donc écrire $u = u_1u_2$ et $v = v_1v_2$ avec $u_1, u_2 \in \mathcal{L}_1$ et $v_1, v_2 \in \mathcal{L}_2$.
   Supposons que u est préfixe de v alors $v = uv'$. On a donc les deux alternatives suivantes :
   - Si $|v_1| \geq |u_1|$, cela signifie que u_1 est préfixe de v_1 donc $u_1 = v_1$ car $\mathcal{L}_1$ est préfixe. On a donc u_2 qui est préfixe de v_2 et $u_2 \neq v_2$ ce qui est impossible car $\mathcal{L}_2$ est préfixe.
   - Si $|v_1| < |u_1|$, cela signifie que v_1 est préfixe de u_1 et $v_1 \neq u_1$ ce qui est impossible car $\mathcal{L}_1$ est préfixe.
5. Soit $\mathcal{L}_1 = \{a, ba, bba, bbba\}$ et $\mathcal{L}_2 = \{b, ab, abb\}$. $\mathcal{L}_1 \cup \mathcal{L}_2$ n'est pas préfixe car b est préfixe de ba.

Correction 2 On définit les ensembles $L = \{a \dots z, A \dots Z\}$, $C = \{0 \dots 9\}$, $U = \{\_\}$ et on définit :
— $\mathcal{L}_1 = (L \cup U).(L \cup C \cup U)^*$
— $\mathcal{L}_2 = \{0\} \cup ((C \backslash \{0\}).C^*)$
— $\mathcal{L}_3 = \{0\}.\{x\}.(C \cup (\{A, B, C, D, E, F\}))^*$

Correction 3 1. Soit $L_1 \subseteq M_1$ et $L_2 \subseteq M_2$. Alors $L_1.L_2 = \{u_1.u_2 : u_1 \in L_1, u_2 \in L_2\} \subseteq \{u_1.u_2 : u_1 \in M_1, u_2 \in M_2\} = M_1.M_2$

2. Soit $L \subseteq M$. Preuve par induction sur n pour prouver $L^n \subseteq M^n$:
   — $n = 0$: $L^0 = \{\varepsilon\} = M^0$
   — hérédité : soit $L^n \subseteq M^n$.
   Alors, par la monotonie de l'opérateur de concaténation : $L^{n+1} = L.L^n \subseteq M.M^n = M^{n+1}$
3. Soit $L \subseteq M$. Alors $L^* = \bigcup_{n\geq 0} L^n \subseteq \bigcup_{n\geq 0} M^n = M^*$, utilisant le fait que pour tout n, $L^n \subseteq M^n$ comme prouvé sous le point précédent.

Correction 4 1. correct

2. incorrect, prendre $L = \{l\}$ et $M = \{m\}$
3. On a $L^* \subseteq (L \cup M)^*$, voir l'exercice sur la monotonie, donc aussi $L^* \cup M^* \subseteq (L \cup M)^*$. Mais pas $(L \cup M)^* \subseteq L^* \cup M^*$, prendre $L = \{l\}$ et $M = \{m\}$, avec $lm \in (L \cup M)^*$ et $lm \notin L^* \cup M^*$

Correction 5 1. $L^* = L.L^*$ est incorrect, mais $L^* = \{\varepsilon\} \cup L.L^*$, parce que $L^* = \bigcup_{n\geq 0} L^n = L^0 \cup \bigcup_{n>0} L^n = \{\varepsilon\} \cup \bigcup_{n\geq 0} L.L^n = \{\varepsilon\} \cup L.\bigcup_{n\geq 0} L^n = \{\varepsilon\} \cup L.L^*$ Utilise la distributivité de la concaténation de l'exercice précédent (mais sur une union infinie).

2. $L^+ = L.L^*$ correct, preuve similaire.
3. $L^* = L^*.L^*$ On montre $L^* \subseteq L^*.L^*$, parce que $\{\varepsilon\} \subseteq L^*$ et $L^* = \{\varepsilon\}.L^* \subseteq L^*.L^*$
   Inversement, pour montrer $L^*.L^* \subseteq L^*$, soit $w \in L^*.L^*$. Alors, $w = u.v$ avec $|u| = n$ et $|v| = m$, donc $u \in L^n$ et $v \in L^m$, donc $w \in L^{n+m}$, donc $w \in L^*$
4. $L^+ = L^+.L^+$ incorrect : L^+ contient des mots de taille ≥ 1 et $L^+.L^+$ uniquement des mots de taille ≥ 2. On pourrait écrire $L + L^+ = L^+.L^+$ (ce qui n'a pas beaucoup d'intérêt).

Correction 6 1. Le premier n'est pas un code car $a.b = ab$.

Le deuxième, $X = \{a, ab, bb\}$, est un code. Soient $x_1, \dots, x_n, y_1, \dots, y_p \in X$ tel que $u = x_1 \dots x_n = y_1 \dots y_p$.
— Si u commence par b, le mot est formé que par cette lettre et donc les éléments $x_1, \dots, x_n, y_1, \dots, y_p$ ne peuvent être que l'élément bb de X. On en déduit l'égalité voulu.
— Si u commence par a, il est forcément suivi par des b.

- Si le nombre de b est pair alors $x_1 = y_1 = a$ et les autres mots sont bb.
- Si le nombre de b est impair alors $x_1 = y_1 = ab$ et les autres mots sont bb.

Le troisième n'est pas un code car $a.ba = ab.a$.

Le quatrième, $X = \{aa, ba, baa, bb, bba\}$ est un code. On le montre par récurrence sur le nombre de lettre du mot à décomposer puis en faisant des distinctions de cas suivant la parité du block de a qui sit le premier b.

2. Soit $X = \{u_1, \ldots, u_n\}$ un ensemble de n mots distincts de même longueur k. Soit $x_1, \ldots, x_n, y_1, \ldots, y_p$ des mots de X vérifiant $x_1 \ldots x_n = y_1 \ldots y_p$. Comme tout mot de X ont la même longueur k, on a $|x_1 \ldots x_n| = k\, n$ et $|y_1 \ldots y_p| = k\, p$ donc $p = n$. Pour tout i entre 1 et n, les lettres d'indices compris entre $k(i-1)+1$ et ki dans $x_1 \ldots x_n$ et $y_1 \ldots y_n$ sont égales, on en déduit que $x_i = y_i$.
3. Soit X un code préfixe. Montrons par récurrence sur n que si $x_1, \ldots, x_n, y_1, \ldots, y_p \in X$ tels que $x_1 \ldots x_n = y_1 \ldots y_p$ alors $n = p$ et $x_i = y_i$ pour tout i entre 1 et p.
   - Si $n = 1$, $x_1 = y_1 \ldots y_p$ donc y_1 est un préfixe de x_1. Comme X est un code préfixe, on en déduit que $x_1 = y_1$ et donc $p = 1$.
   - On suppose la propriété vraie au rang n et montrons que cette propriété reste vraie au rang suivant. Considérons $x_1, \ldots, x_{n+1}, y_1, \ldots, y_p \in X$ tels que $x_1 \ldots x_{n+1} = y_1 \ldots y_p$. On a deux possibilités, soit x_1 est préfixe de y_1, soit y_1 est préfixe de x_1. Dans les deux cas on en déduit que $x_1 = y_1$ car X est préfixe. On a alors $x_2 \ldots x_{n+1} = y_2 \ldots y_p$ et en appliquant l'hypothèse de récurrence, on obtient que $n+1 = p$ et $x_i = y_i$ pour tout i entre 2 et p.
4. Le même type de résultat fonctionne pour les codes suffixes.

Correction 7 1. $\mathcal{M}.(\mathcal{M}^*.\mathcal{N}) \cup \mathcal{N} = (\mathcal{M}.\mathcal{M}^* \cup \varepsilon).\mathcal{N} = \mathcal{M}^*.\mathcal{N}$.

On a $\mathcal{N} \subset \mathcal{M}^*.\mathcal{N}$ et $\mathcal{M}.\mathcal{M}^*.\mathcal{N} \subset \mathcal{M}^*.\mathcal{N}$ donc $\mathcal{M}.\mathcal{M}^*.\mathcal{N} \cup \mathcal{N} \subset \mathcal{M}^*.\mathcal{N}$.

2. Soit $\mathcal{L}$ une solution, montrons par récurrence que pour tout $n \in \mathbb{N}$ on a
$$\mathcal{L} = \mathcal{M}^{n+1}.\mathcal{L} \cup \mathcal{M}^n.\mathcal{N} \cup \mathcal{M}^{n-1}.\mathcal{N} \cup \cdots \cup \mathcal{M}\mathcal{N} \cup \mathcal{N}.$$
Pour $n = 1$, la propriété découle que $\mathcal{L}$ est solution de l'équation.
On suppose la propriété vraie au rang n. Comme $\mathcal{L}$ est solution de l'équation on a :
$$\begin{aligned} \mathcal{L} &= \mathcal{M}.\mathcal{L} \cup \mathcal{N} \\ &= \mathcal{M}.(\mathcal{M}^{n+1}.\mathcal{L} \cup \mathcal{M}^n.\mathcal{N} \cup \mathcal{M}^{n-1}.\mathcal{N} \cup \cdots \cup \mathcal{M}\mathcal{N} \cup \mathcal{N}) \cup \mathcal{N} \\ &= \mathcal{M}^{n+2}.\mathcal{L} \cup \mathcal{M}^n.\mathcal{N} \cup \mathcal{M}^{n-1}.\mathcal{N} \cup \cdots \cup \mathcal{M}\mathcal{N} \cup \mathcal{N} \end{aligned}$$
3. On a $\varepsilon.\mathcal{N} = \mathcal{N} \subset \mathcal{M}.\mathcal{L} \cup \mathcal{N} = \mathcal{L}$.
Pour $n \in \mathbb{N}$, on a $\mathcal{M}^n.\mathcal{N} \subset \mathcal{M}^{n+1}.\mathcal{L} \cup \mathcal{M}^n.\mathcal{N} \cup \mathcal{M}^{n-1}.\mathcal{N} \cup \cdots \cup \mathcal{M}\mathcal{N} \cup \mathcal{N} = \mathcal{L}$.
On en déduit que $\mathcal{M}^*.\mathcal{N} \subset \mathcal{L}$
4. Supposons que $\varepsilon \notin \mathcal{M}$. On veut montrer que $\mathcal{L} \subset \mathcal{M}^*.\mathcal{N}$. Soit $w \in \mathcal{L}$ de longueur n. Alors $w \in \mathcal{M}^{n+1}.\mathcal{L} \cup \mathcal{M}^n.\mathcal{N} \cup \mathcal{M}^{n-1}.\mathcal{N} \cup \cdots \cup \mathcal{M}\mathcal{N} \cup \mathcal{N} = \mathcal{L}$. Comme $\mathcal{M}^{n+1}.\mathcal{L}$ contient uniquement des mots de longueurs plus grandes que $n+1$, on en déduit que $w \in \mathcal{M}^n.\mathcal{N} \cup \mathcal{M}^{n-1}.\mathcal{N} \cup \cdots \cup \mathcal{M}\mathcal{N} \cup \mathcal{N} \subset \mathcal{M}^*.\mathcal{N}$. Ainsi $\mathcal{L} \subset \mathcal{M}^*.\mathcal{N}$ et donc $\mathcal{L} = \mathcal{M}^*.\mathcal{N}$.

Correction 8 (1) $\Longrightarrow$ (3) Supposons que $u.v = v.u$ alors $u^{|v|}v^{|u|} = v^{|u|}u^{|v|}$. Comme $|v^{|u|}| = |u^{|v|}|$ on en déduit que $v^{|u|} = u^{|v|}$.

(3) $\Longrightarrow$ (2) Montrons par récurrence sur n que si $|u| + |v| \leq n$ et s'il existe $i, j \in \mathbb{N}^*$ tels que $u^i = v^j$ alors il existe $w \in \mathcal{A}^*$ et $p, q \in \mathbb{N}^*$ vérifiant $u = w^p$ et $v = w^q$.

- Si $|u| + |v| = 0$, cela signifie que $u = v = \varepsilon$ et la propriété est vérifiée.
- Supposons que la propriété soit vraie pour $n \in \mathbb{N}$ et vérifions là au rang $n+1$. Soit $|u| + |v| \leq n+1$ et $i, j \in \mathbb{N}^*$ tels que $u^i = v^j$. Quitte à inverser les rôle, on peut supposer que $|u| \geq |v|$. On réalise la division euclidienne $|u| = q|v| + r$ avec $r \in [0, |v|-1]$, on a deux possibilités :
  - Si $r = 0$ alors $u = v^q$.
  - Sinon $u = v^q.w$ avec $|w| = r \neq 0$. Comme $u^i = v^j$, en considérant le dernier mot u on voit qu'on peut le décomposer de la forme $u = w'v^q$ où $|w| = |w'|$. Or w est un préfixe de v est donc de u. Comme $|w| = |w'|$, on en déduit que $w = w'$ donc on a $wv^q = v^q w$. D'après l'implication (1) $\Longrightarrow$ (3), il existe $i', j' \in \mathbb{N}^*$ tels que $v^{i'} = w^{j'}$ et $|w| + |v| \leq n$. Par hypothèse de récurrence, il existe un mot x tel que $v = x^r$ et $w = x^s$, on a alors $u = v^q.w = x^{rq+s}$.

  Par récurrence, on en déduit la décomposition demandée.

(2) $\Longrightarrow$ (1) Supposons qu'il existe $w \in \mathcal{A}^*$ et $p, q \in \mathbb{N}^*$ tels que $u = w^p$ et $v = w^q$. On a $uv = w^{p+q} = vu$.