Exercices bases formelles du TAL

Pierre-Léo Bégay

pbegay@ens-cachan.fr

1 Théorie des ensembles

Produit cartésien Soit $A = \{a, b, c, d\}$. Calculer A^2 (vous devriez trouver un ensemble à $4^2 = 16$ éléments)

Correction

$$A^2 = A \times A = \{ \begin{array}{cccc} \langle a,a \rangle, & \langle a,b \rangle, & \langle a,c \rangle, & \langle a,d \rangle, \\ \langle b,a \rangle, & \langle b,b \rangle, & \langle b,c \rangle, & \langle b,d \rangle, \\ \langle c,a \rangle, & \langle c,b \rangle, & \langle c,c \rangle, & \langle c,d \rangle, \\ \langle d,a \rangle, & \langle d,b \rangle, & \langle d,c \rangle, & \langle d,d \rangle \end{array} \}$$

Ensemble des parties Calculer $P(A)$ (vous devriez trouver un ensemble à $2^4 = 16$ éléments)

Correction
$P(A) = \{\{\}, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b\}, \{a, c\}, \{a, d\}, \{b, c\}, \{b, d\}, \{c, d\}, \{b, c, d\}, \{a, c, d\}, \{a, b, d\}, \{a, b, c\}, \{a, b, c, d\}\}$

2 Mots et langages

2.1 Facteurs et sous-mots

Soit $u = abbababaa$. Les mots suivants sont-ils des sous-mots et/ou des facteurs de u ?

Correction

	sous-mot	justification	facteur	justif
aaaaa	Yes	abbababaa	No	
aaaaaa	No	Trop de a	No	
aba	Yes	abbababaa (entre autres)	Yes	abbababaa (ou abbababaa)
bab	Yes	abbababaa (idem)	Yes	abbababaa (ou abbababaa)
babaa	Yes	abbababaa (idem)	Yes	abbababaa
bbaab	Yes	abbababaa	No	
bbaabb	No	Les b sont *imposés*, les a *matchent* pas	No	

2.2 Pré/suffixes

Calculer $Pre(u)$ et $Suf(u)$.

Correction

$$Pre(u) = \{\epsilon, a, ab, abb, abba, abbab, abbaba, abbabab, abbababa, abbababaa\}$$

$$Suf(u) = \{\epsilon, a, aa, baa, abaa, babaa, ababaa, bababaa, bbababaa, abbababaa\}$$

2.3 Distributivité

Soient L_1, L_2 et L_3 trois langages rationnels quelconques. Est-ce qu'on a

$L_1.(L_2 \cup L_3) = L_1.L_2 \cup L_1.L_3$?

$L_1 \cup (L_2.L_3) = (L_1 \cup L_2).(L_1 \cup L_3)$?

Correction

$$
\begin{array}{rll}
 & L_1.(L_2 \cup L_3) & \\
= & \{uv \mid u \in L_1 \text{ et } v \in (L_2 \cup L_3)\} & \text{définition du produit de langages} \\
= & \{uv \mid u \in L_1 \text{ et } (v \in L_2 \text{ ou } v \in L_3)\} & \text{définition de } \cup \\
= & \{uv \mid (u \in L_1 \text{ et } v \in L_2) \text{ ou } (u \in L_1 etv \in L_3)\} & \text{distributivité de "ou" et "et"} \\
= & \{uv \mid uv \in (L_1.L_2) \text{ ou } uv \in (L_1.L_3)\} & \text{définition du produit de langages} \\
= & \{w \mid w \in (L_1.L_2) \text{ ou } w \in (L_1.L_3)\} & \text{uv réécrit en w} \\
= & \{w \mid w \in (L_1.L_2)\} \cup \{w|w \in (L_1.L_3)\} & \text{définition du } \cup \\
= & (L_1.L_2) \cup (L_1.L_3)\ (\{w|w \in L\} = L) &
\end{array}
$$

Quant à $L_1 \cup (L_2.L_3) = (L_1 \cup L_2).(L_1 \cup L_3)$, c'est faux. On peut le vérifier avec le contre-exemple $L_1 = \{a\}, L_2 = \{b\}, L_3 = \{c\}$:

- $L_1 \cup (L_2.L_3) = \{a\} \cup \{bc\} = \{a, bc\}$
- $(L_1 \cup L_2).(L_1 \cup L_3) = \{a, b\}.\{a, c\} = \{aa, ac, ba, bc\}$

3 Expressions rationnelles

3.1 Distributivité

Soient e_1, e_2 et e_3 trois ERs quelconques. Est-ce qu'on a

$e_1(e_2 + e_3) =$[1] $e_1e_2 + e_1e_3$?

$e_1 + (e_2e_3) = (e_1 + e_2)(e_1 + e_3)$?

Correction

$$
\begin{array}{rll}
 & [\![e_1(e_2 + e_3)]\!] & \\
= & [\![e_1]\!].[\![e_2 + e_3]\!] & \text{définition de } [\![ee']\!] \\
= & [\![e_1]\!].([\![e_2]\!] \cup [\![e_3]\!]) & \text{définition de } [\![e + e']\!] \\
= & ([\![e_1]\!].[\![e_2]\!]) \cup ([\![e_1]\!].[\![e_3]\!]) & \text{Exercice précédent} \\
= & [\![e_1e_2]\!] \cup [\![e_1e_3]\!] & \text{définition de } [\![ee']\!] \\
= & [\![e_1e_2 + e_1e_3]\!] & \text{définition de } [\![e + e']\!]
\end{array}
$$

Pour la deuxième question, même contre-exemple que dans l'exercice précédent.

[1] Où l'égalité entre ERs désigne l'égalité *sémantique*, c'est à dire que $e = e'$ si et seulement si $[\![e]\!] = [\![e']\!]$ (où le second $=$ désigne donc l'égalité entre ensembles). A distinguer de l'égalité *syntaxique*, qui indique que les deux ERs données sont formées de l'exacte même façon. On a $(e =_{syntaxe} e') \Rightarrow (e =_{sémantique} e')$, mais pas l'inverse. En effet, $(a^*)^* =_{sémantique} a^*$, mais pas $(a^*)^* =_{syntaxe} a^*$

3.2 Version

Soit un alphabet fini A qui contient au moins les lettres a et b. Exprimer en français, de façon la plus naturelle possible, les langages dénotés par les expressions régulières suivantes :

1. aA^*

   a puis n'importe quoi

   $\Rightarrow$ Les mots qui commencent par un a

2. A^*a

   n'importe quoi puis a

   $\Rightarrow$ Les mots qui finissent par un a

3. A^*abbaA^*

   n'importe quoi puis $abba$ puis n'importe quoi

   $\Rightarrow$ Les mots qui contiennent (au moins) un facteur $abba$

4. $A^*aA^*aA^*$

   n'importe quoi puis a puis n'importe quoi puis a puis n'importe quoi

   $\Rightarrow$ Les mots qui contiennent au moins deux a

5. A^*aaA^*

   n'importe quoi puis aa puis n'importe quoi

   $\Rightarrow$ Les mots qui contiennent (au moins) un facteur aa

6. $(a^*b)^*$

   Autant de fois qu'on veut, un b précédé d'autant de a qu'on veut

   On peut donc former tout mot de la forme $a^{n_1}ba^{n_2}b\ldots a^{n_k}b$, avec $\forall i.n_i \in \mathbb{N}$ (ça comprend 0)

   $\Rightarrow$ L'ensemble des mots qui terminent par b (ainsi que ϵ)

7. $(a^*b)^*a^*$

   (L'ensemble des mots qui terminent par b (ainsi que ϵ)) suivis par autant de a qu'on veut

   $\Rightarrow$ Tous les mots

8. a^*b^*

   $\Rightarrow$ Les mots composés de a suivis de b (ou uniquement de a, ou uniquement de b, ou ϵ)

3.3 ERs et mot vide

Préciser, pour chacune des ERs ci-dessus, si le langage dénoté contient le mot vide ϵ.

Correction

1. aA^*

   $\Rightarrow$ Non, car tout mot du langage doit au moins contenir un a

2. A^*a

   $\Rightarrow$ Idem

3. A^*abbaA^*

   $\Rightarrow$ Idem avec $abba$

4. $A^*aA^*aA^*$

   $\Rightarrow$ Idem avec deux a

5. A^*aaA^*

   $\Rightarrow$ Idem avec aa

6. $(a^*b)^*$

   $\Rightarrow$ Oui, il suffit d'*instancier* $*$ par 0

7. $(a^*b)^*a^*$

   $\Rightarrow$ Oui, il suffit d'*instancier* les deux $*$ par 0

8. a^*b^*

   $\Rightarrow$ Idem

Donner une ER équivalente à $(a+\epsilon)(bc+ab)(c+\epsilon)$ qui n'utilise pas ϵ

Correction Par distributivité, on obtient

$$\begin{aligned} & abcc + abc + aabc + aab + bcc + bc + abc + ab \\ = \; & abcc + abc + aabc + aab + bcc + bc + ab \end{aligned}$$

3.4 Inclusions et équivalences entre ERs

Pour chaque paire d'ERs e_1 et e_2, dire si le langage dénoté par la première est inclu dans celui dénoté par la seconde, et inversement. Dit autrement, si on pose $L_1 = [\![e_1]\!]$ et $L_2 = [\![e_2]\!]$, est-ce qu'on a $L_1 \subseteq L_2$ et/ou $L_2 \subseteq L_1$?

1. $e_1 = a(ba)^*$ et $e_2 = (ab)^*a$

   - $L_1 = \{a(ba)^n | n \in \mathbb{N}\}$
     $= \{(ab)^n a | n \in \mathbb{N}\}$
     $= L_2$
   - Donc $L_1 \subseteq L_2$ et $L_2 \subseteq L_1$

2. $e_1 = a^*(ba)^*$ et $e_2 = (ab)^*a^*$

   - $ba \in L_1$ et $\notin L_2$

- On a donc $L_1 \not\subseteq L_2$
- $ab \in L_2$ et $\notin L_1$
- On a donc $L_2 \not\subseteq L_1$

3. $e_1 = a^*(ba^*)^*$ et $e_2 = (a^*b)^*a^*$
    - $L_1 = [\![(a+b)^*]\!] = L_2$ (cf l'exercice précédent)
    - On a donc $L_1 \subseteq L_2$ et $L_2 \subseteq L_1$
4. $e_1 = (a^*b^*)^*$ et $e_2 = (a^*b^*)$
    - $L_2 = [\![(a^*b^*)]\!] = [\![(a^*b^*)^1]\!] \subseteq \bigcup_{n\in\mathbb{N}}[\![(a^*b^*)^n]\!] = [\![(a^*b^*)^*]\!] = L_1$
    - On a donc $L_2 \subseteq L_1$
    - Par contre, on a $abab \in L_1$ et $\notin L_2$
    - On a donc $L_1 \not\subseteq L_2$
5. $e_1 = A^*(a+b)A^*$ et $e_2 = A^*(a+b)A^* + A^*(a+b)A^*$
    - On a $e_2 = e_1 + e_1$, donc $L_2 = [\![e_2]\!] = [\![e_1 + e_1]\!] = [\![e_1]\!] \cup [\![e_1]\!] = [\![e_1]\!] = L_1$