

Corrigé du travail pratique #1

Réponses

Note : les réponses aux différentes questions ne sont pas toujours uniques.

1. (a) ϵ ,

c,
ca,
car,
cart,
carte

(b) Attention! Il fallait éviter les préfixes dégénérés.

er,
ker,
oker,
r

(c) Attention! Il fallait éviter les sous-chaînes dégénérées.

a,
ai,
ain,
i,
in,
m,
ma,
mai,
n

(d) ϵ ,

a,
aa,
aaa,
aaaa,
aaaab,
aaab,
aab,
ab,
b

- (e) abcabcab,
bcabcabc,
cabcabca

2. (a) Une apparition obligatoire de chacun des trois symboles **b**, **c** et **c** est placée suivant un certain ordre puis des symboles additionnels quelconques sont insérés avant, entre et après les apparitions obligatoires. Il y a 3 ordres pour les apparitions obligatoires.

$$\begin{aligned} A &\rightarrow (\mathbf{a} \mid \mathbf{b} \mid \mathbf{c})^* \\ B &\rightarrow A \mathbf{b} A \mathbf{c} A \mathbf{c} A \\ &\quad \mid A \mathbf{c} A \mathbf{b} A \mathbf{c} A \\ &\quad \mid A \mathbf{c} A \mathbf{c} A \mathbf{b} A \end{aligned}$$

- (b) On crée une définition régulière qui comporte d'abord ~~les alias d_0 à d_5~~ les alias d_0 à d_8 . L'alias d_i génère ~~des chaînes~~ les chaînes faites de chiffres dans $\{0, \dots, 4\}$ dont ~~ont~~ la somme des chiffres égale i . — L'alias f génère les chaînes qui se terminent par un chiffres élevé. — Enfin, on crée l'alias principal e qui génère le langage désiré.

$$\begin{aligned} d_0 &\rightarrow 0^* \\ d_1 &\rightarrow d_0 1 d_0 \\ d_2 &\rightarrow d_1 d_1 \mid d_0 2 d_0 \\ d_3 &\rightarrow d_1 d_2 \mid d_2 d_1 \mid d_0 3 d_0 \\ d_4 &\rightarrow d_1 d_3 \mid d_2 d_2 \mid d_3 d_1 \mid d_0 4 d_0 \\ d_5 &\rightarrow \del{d_1 d_4 \mid d_2 d_3 \mid d_3 d_2 \mid d_4 d_1 \mid d_0 5 d_0} \\ e &\rightarrow \del{d_5 (0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9)^*} \\ d_5 &\rightarrow d_1 d_4 \mid d_2 d_3 \mid d_3 d_2 \mid d_4 d_1 \\ d_6 &\rightarrow d_1 d_5 \mid d_2 d_4 \mid d_3 d_3 \mid d_4 d_2 \\ d_7 &\rightarrow d_1 d_6 \mid d_2 d_5 \mid d_3 d_4 \mid d_4 d_3 \\ d_8 &\rightarrow d_1 d_7 \mid d_2 d_6 \mid d_3 d_5 \mid d_4 d_4 \\ f &\rightarrow [0-4]^* [5-9] \\ e &\rightarrow (d_5 \mid d_6 \mid d_7 \mid d_8 \mid f) [0-9]^* \end{aligned}$$

- (c) Les deux sous-chaînes requises peuvent apparaître dans un ordre ou dans l'autre et elles peuvent même se chevaucher.

$$\begin{aligned} A &\rightarrow (\mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \mathbf{d})^* \\ B &\rightarrow A \mathbf{a} \mathbf{b} \mathbf{c} A \mathbf{b} \mathbf{c} \mathbf{d} A \mid A \mathbf{b} \mathbf{c} \mathbf{d} A \mathbf{a} \mathbf{b} \mathbf{c} A \mid A \mathbf{a} \mathbf{b} \mathbf{c} \mathbf{d} A \end{aligned}$$

- (d) A priori, on pourrait envisager de concevoir des mots du langage de la manière suivante. La forme du mot serait initialement ' $\dots c \dots a \dots b \dots$ ', pour s'assurer

d'avoir la sous-séquence 'cab', puis on placerait la sous-chaîne 'abc' en remplaçant l'un des '...' par '...abc...' et, enfin, on remplacerait les '...' par n'importe quelles suites de symboles. Cependant, ce serait sans tenir compte du fait que la sous-chaîne 'abc', de par sa présence, fournit certains symboles nécessaires à la sous-séquence 'cab'. En fait, on n'a que deux formes de chaînes à considérer : il suffit de se poser la question à savoir s'il apparaît un 'c' indépendant avant la première apparition de 'abc'.

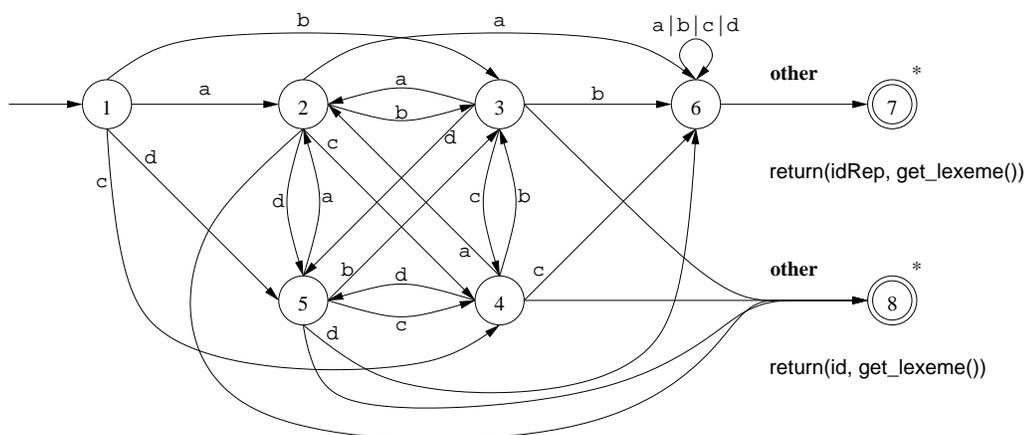
$$A \rightarrow (a | b | c)^*$$

$$B \rightarrow A abc A a A b A | A c A abc A$$

- (e) On peut observer que toute chaîne du langage tombe dans l'un des cas suivants :
 soit la chaîne ne contient pas de 'a' ;
 soit, après le premier 'a', la chaîne ne contient pas de 'c' ;
 soit, après le premier 'c' qui suit le premier 'a', la chaîne ne contient pas de 'd' ;
 etc.

$$\begin{aligned} & (b | c | d)^* \\ | & (b | c | d)^* a (a | b | d)^* \\ | & (b | c | d)^* a (a | b | d)^* c (a | b | c)^* \\ | & (b | c | d)^* a (a | b | d)^* c (a | b | c)^* d (a | b | d)^* \end{aligned}$$

3. Tout d'abord, veuillez noter que la question n'était pas claire en ce qui concerne la question de la fin de l'identificateur. Il aurait fallu avoir un cinquième symbole, par exemple \sqcup , qui aurait permis de terminer les identificateurs sans avoir besoin de consommer tout le texte d'entrée. Peu importe, dans la correction, on va accepter que ayez supposé qu'il y avait un cinquième symbole ou pas. Nous ne présentons qu'un seul diagramme mais celui-ci a deux états acceptants, tout dépendant si une répétition a été aperçue ou pas.



4. Voici le pseudo-code.

```
while (true)
{
  switch(state)
  {
    ...
    case 1:
      c := buffer[forward];
      forward ++;
      if (c == '-')
        state := 3;
      else if (is_digit(c))
        state := 2;
      else
        state := fail();
      break;
    case 2:
      c := buffer[forward];
      forward ++;
      if (c == '.')
        state := 5;
      else if (is_digit(c))
        state := 2;
      else
        state := 4;
      break;
    case 3:
      c := buffer[forward];
      forward ++;
      if (is_digit(c))
        state := 2;
      else
        state := fail();
      break;
    case 4:
      forward --;
      jeton := num;
      attribut := gettoken();
      token_beginning := forward;
      return jeton;
    case 5:
      c := buffer[forward];
      forward ++;
      if (is_digit(c))
        state := 6;
      else
        state := fail();
      break;
    case 6:
      c := buffer[forward];
      forward ++;
      if (is_digit(c))
        state := 6;
      else
        state := 7;
      break;
    case 7:
      forward --;
      jeton := num;
      attribut := gettoken();
      token_beginning := forward;
      return jeton;
    ...
  }
}
```

5. Notons tout d'abord que la fonction de bilan possède une propriété fort utile :

$$f(uv) = f(u) + f(v) .$$

Je fais une preuve par induction sur la longueur des chaînes dans L_5 . Notons que, pour respecter la contrainte sur le nombre de 'a' et de 'b', toute chaîne dans L_5 a une longueur qui est un multiple de 3.

Base de l'induction. Considérons la chaîne w de longueur 0. Cette chaîne est nécessairement $w = \epsilon$. À cause de la première production, G_5 est capable de générer w .

Hypothèse d'induction (HI). Supposons que, pour toute chaîne w dans L_5 de longueur de moins de n symboles, G_5 est capable de générer w .

Pas d'induction. Soit $w \in L_5$ de longueur $n > 0$. Considérons la suite des bilans calculés sur tous les préfixes de $w = c_1 \dots c_n$:

$$0 = f(\epsilon), f(c_1), \dots, f(c_1 \dots c_{n-1}), f(c_1 \dots c_n) = 0 .$$

Premier cas : la suite de bilans passe par zéro en un autre endroit qu'aux extrémités. En d'autres mots, la suite de bilans a la forme :

$$0, \dots, 0, \dots, 0 .$$

Dans ce cas, on peut découper w en deux sous-chaînes non-vides qui ont chacune un bilan nul ; i.e. $w = uv$ et $f(u) = 0 = f(v)$. On note que chacune de u et v est de longueur plus petite que n . On peut donc utiliser l'HI et savoir que $S \Rightarrow^* u$ et $S \Rightarrow^* v$. On peut donc donner une dérivation pour w :

$$S \Rightarrow S S \Rightarrow^* u S \Rightarrow^* u v = w .$$

Deuxième cas : w commence par 'a' et se termine par 'a'. Dans ce cas, la suite de bilan doit avoir la forme suivante :

$$0, +1, \dots, -1, 0 .$$

Or, pour que le bilan soit passé du positif au négatif sans passer par zéro, il faut nécessairement qu'un 'b' ait fait passer le bilan de +1 à -1. Donc, la suite de bilan doit avoir la forme suivante :

$$0, +1, \dots, +1, -1, \dots, -1, 0 .$$

Ceci implique qu'on peut décomposer w en la concaténation $\mathbf{a}ubv\mathbf{a}$ où $f(u) = f(v) = 0$. Bien entendu, u et v sont de longueur plus petite que n , donc l'HI s'applique et on peut donc donner une dérivation de w :

$$S \Rightarrow \mathbf{a}S\mathbf{b}S\mathbf{a} \Rightarrow^* \mathbf{a}ubS\mathbf{a} \Rightarrow^* \mathbf{a}ubv\mathbf{a} = w .$$

Troisième cas : w commence par 'a' et se termine par 'b'. Dans ce cas, la suite de bilan doit avoir la forme suivante :

$$0, +1, \dots, +2, 0 .$$

Or, il faut nécessairement qu'un 'a' ait fait passer le bilan de +1 à +2. Donc, la suite de bilan doit avoir la forme suivante :

$$0, +1, \dots, +1, +2, \dots, +2, 0 .$$

Ceci implique qu'on peut décomposer w en la concaténation $auavb$ où $f(u) = f(v) = 0$. Encore une fois, u et v sont de longueur plus petite que n , donc l'HI s'applique et on peut donc donner une dérivation de w :

$$S \Rightarrow aSaSb \Rightarrow^* auasb \Rightarrow^* auavb = w .$$

Quatrième cas : w commence par 'b' et se termine par 'a'. Ce cas se traite de manière similaire au cas précédent.

Cinquième cas : w commence par 'b' et se termine par 'b'. Dans ce cas, la suite de bilan doit avoir la forme suivante :

$$0, -2, \dots, +2, 0 .$$

Or, il est inutile de considérer ce cas car la suite de bilans doit nécessairement passer par zéro entre le -2 et le $+2$ et cette situation a déjà été captée par le premier cas. Ceci complète le pas d'induction et, par la même occasion, la preuve que $L_5 \subseteq L(G_5)$.

8. Nous commençons avec la grammaire originale.

$$\begin{aligned}
 E &\rightarrow E \text{ ou } T \mid T \\
 T &\rightarrow T \text{ et } F \mid F \\
 F &\rightarrow \text{non } F \mid A \\
 A &\rightarrow (E) \mid \text{faux} \mid \text{vrai} \mid \text{id}
 \end{aligned}$$

Pour $i = 1$: On nettoie les E -productions.

- La boucle sur j est vide. (C'est-à-dire qu'il n'y a aucun non-terminal placé *avant* E dans l'ordre des non-terminaux.)
- Il faut éliminer la récursion à gauche immédiate (c'est-à-dire, toute production de la forme $E \rightarrow E\alpha$). On obtient la grammaire :

$$\begin{aligned}
 E &\rightarrow T E' \\
 E' &\rightarrow \text{ou } T E' \mid \epsilon \\
 T &\rightarrow T \text{ et } F \mid F \\
 F &\rightarrow \text{non } F \mid A \\
 A &\rightarrow (E) \mid \text{faux} \mid \text{vrai} \mid \text{id}
 \end{aligned}$$

Pour $i = 2$: On nettoie les T -productions.

- La boucle sur j se résume à $j = 1$. Il faut éliminer toute production de la forme $T \rightarrow E\alpha$. Rien à faire.
- Il faut éliminer la récursion à gauche immédiate. On obtient la grammaire :

$$\begin{aligned}
 E &\rightarrow T E' \\
 E' &\rightarrow \text{ou } T E' \mid \epsilon \\
 T &\rightarrow F T' \\
 T' &\rightarrow \text{et } F T' \mid \epsilon \\
 F &\rightarrow \text{non } F \mid A \\
 A &\rightarrow (E) \mid \text{faux} \mid \text{vrai} \mid \text{id}
 \end{aligned}$$

Pour $i = 3$: On nettoie les F -productions.

- Pour $j = 1$, il faut éliminer toute production de la forme $F \rightarrow E\alpha$. Rien à faire.
- Pour $j = 2$, il faut éliminer toute production de la forme $F \rightarrow T\alpha$. Rien à faire.
- Il n'y a pas de récursion à gauche immédiate chez F . Rien à faire.

Pour $i = 4$: On nettoie les A -productions.

- Pour $j = 1$, il faut éliminer toute production de la forme $A \rightarrow E\alpha$. Rien à faire.
- Pour $j = 2$, il faut éliminer toute production de la forme $A \rightarrow T\alpha$. Rien à faire.
- Pour $j = 3$, il faut éliminer toute production de la forme $A \rightarrow F\alpha$. Rien à faire.
- Il n'y a pas de récursion à gauche immédiate chez A . Rien à faire.

La dernière grammaire donnée est la grammaire souhaitée.

9. Voici le pseudo-code.

```
void S()
{
  if (peekToken().type == 'a')
  {
    readToken('a');
    S();
    readToken('a');
  }
  else if (peekToken().type == 'b')
  {
    readToken('b');
    S();
    readToken('b');
  }
  else
  {
    readToken('c');
  }
}
```

10.

PILE	ENTRÉE	SORTIE
$S \$$	bbbaabdccddd\$	$S \rightarrow b S d$
$b S d \$$	bbbaabdccddd\$	(Consomme b)
$S d \$$	bbaabdccddd\$	$S \rightarrow b S d$
$b S d d \$$	bbaabdccddd\$	(Consomme b)
$S d d \$$	baabdccddd\$	$S \rightarrow b S d$
$b S d d d \$$	baabdccddd\$	(Consomme b)
$S d d d \$$	aabdccddd\$	$S \rightarrow a S c$
$a S c d d d \$$	aabdccddd\$	(Consomme a)
$S c d d d \$$	abdccddd\$	$S \rightarrow a S c$
$a S c c d d d \$$	abdccddd\$	(Consomme a)
$S c c d d d \$$	bdccddd\$	$S \rightarrow b S d$
$b S d c c d d d \$$	bdccddd\$	(Consomme b)
$S d c c d d d \$$	dccddd\$	$S \rightarrow \epsilon$
$d c c d d d \$$	dccddd\$	(Consomme d)
$c c d d d \$$	ccddd\$	(Consomme c)
$c d d d \$$	cddd\$	(Consomme c)
$d d d \$$	ddd\$	(Consomme d)
$d d \$$	dd\$	(Consomme d)
$d \$$	d\$	(Consomme d)
$\$$	$\$$	(Succès)

11. Posons les contraintes sur les ensembles FIRST qui sont nécessaires au calcul de FIRST(S).

$$\begin{aligned}
 \text{FIRST}(S) &= \text{FIRST}(\epsilon) \cup \text{FIRST}(\mathbf{aSaSb}) \cup \text{FIRST}(\mathbf{aSbSa}) \\
 &\quad \cup \text{FIRST}(\mathbf{bSaSa}) \cup \text{FIRST}(SS) \\
 &= \{\epsilon\} \cup \{\mathbf{a}\} \cup \{\mathbf{a}\} \cup \{\mathbf{b}\} \cup \text{FIRST}(SS) \\
 &= \{\epsilon, \mathbf{a}, \mathbf{b}\} \cup (\text{FIRST}(S) - \{\epsilon\}) \cup \text{FIRST}(S) \\
 &= \{\epsilon, \mathbf{a}, \mathbf{b}\} \cup \text{FIRST}(S)
 \end{aligned}$$

La plus petite solution à cette contrainte nous donne l'ensemble suivant.

	FIRST
S	$\{\mathbf{a}, \mathbf{b}, \epsilon\}$

Maintenant, posons les contraintes sur l'ensemble FOLLOW qui sont nécessaires au calcul de FOLLOW(S).

$\text{FOLLOW}(S) \supseteq \{\mathbf{\$}\}$		symb. de départ
$\text{FOLLOW}(S) \supseteq \text{FIRST}(\mathbf{aSb}) - \{\epsilon\} = \{\mathbf{a}\}$		appar. dans 2e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FIRST}(\mathbf{b}) - \{\epsilon\} = \{\mathbf{b}\}$		appar. dans 2e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FIRST}(\mathbf{bSa}) - \{\epsilon\} = \{\mathbf{b}\}$		appar. dans 3e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FIRST}(\mathbf{a}) - \{\epsilon\} = \{\mathbf{a}\}$		appar. dans 3e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FIRST}(\mathbf{aS a}) - \{\epsilon\} = \{\mathbf{a}\}$		appar. dans 4e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FIRST}(\mathbf{a}) - \{\epsilon\} = \{\mathbf{a}\}$		appar. dans 4e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FIRST}(S) - \{\epsilon\} = \{\mathbf{a}, \mathbf{b}\}$		appar. dans 5e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FIRST}(\epsilon) - \{\epsilon\} = \{\}$		appar. dans 5e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FOLLOW}(S)$		appar. fin 5e S -prod.
$\text{FOLLOW}(S) \supseteq \text{FOLLOW}(S)$		appar. fin 5e S -prod.

La plus petite solution à ces contraintes nous donne l'ensemble suivant.

	FOLLOW
S	$\{\mathbf{a}, \mathbf{b}, \mathbf{\$}\}$

12. Voici les ensembles PREDICT qui justifient les insertions de productions dans la table d'analyse.

PRODUCTION	PREDICT
$A \rightarrow B$	{a, b, c}
$A \rightarrow dB$	{d}
$B \rightarrow cA$	{c}
$B \rightarrow Cb$	{a, b}
$B \rightarrow a$	{a}
$C \rightarrow aB$	{a}
$C \rightarrow \epsilon$	{b}

NON- TERMINAL	SYMBOLE D'ENTRÉE				
	a	b	c	d	\$
A	$A \rightarrow B$	$A \rightarrow B$	$A \rightarrow B$	$A \rightarrow dB$	
B	$B \rightarrow Cb$ $B \rightarrow a$	$B \rightarrow Cb$	$B \rightarrow cA$		
C	$C \rightarrow aB$	$C \rightarrow \epsilon$			