

S.V.P. ME SIGNALER TOUTE FAUTE DE FRAPPE, GRAMMAIRE, ORTHOGRAPHE OU LOGIQUE; LES FAUTES DE LOGIQUE VOUS APPORTENT DES POINTS SUPPLÉMENTAIRES POUR LA NOTE FINALE. TOUTE SUGGESTION D'UNE MEILLEURE TRADUCTION D'UN TERME ANGLAIS SERA ÉGALEMENT BIENVENUE.

On veut une grammaire vérifiant

- toute règle est soit de la forme  $A \rightarrow a$ ,  $A \in V$ ,  $a \in \Sigma$ , soit de la forme  $A \rightarrow BC$ ,  $A, B, C \in V$ ;
- le symbole (variable) de départ n'apparaît jamais à droite d'une règle (i.e. il n'y a pas de règles  $A \rightarrow SB$  ou  $A \rightarrow BS$ );
- toutes les variables sont utiles, i.e. pour tout  $A \in V$  il existe  $\alpha, \beta \in (V \cup \Sigma)^*$  et  $w \in \Sigma^*$  tels que  $S \xRightarrow{*} \alpha A \beta \xRightarrow{*} \alpha w \beta$ .

## 1 Version Hopcroft-Ullman

qui utilise la notion de  $A \xRightarrow{*} \varepsilon$  et  $A \xRightarrow{*} B$ .

Soit  $G = (V, \Sigma, R, S)$  la grammaire définie par

$$V = \{S, T\}$$

$$\Sigma = \{a, b\}$$

R:

$$S \rightarrow aSa \mid aTa \mid S \mid T$$

$$T \rightarrow aTa \mid bTb \mid \varepsilon$$

Pour la convertir en forme normale de Chomsky, on construit d'abord l'ensemble  $E = \{A \in V : A \xRightarrow{*} \varepsilon\}$ .

$$E_0 = \{T\} = \{A \in V : A \rightarrow \varepsilon\}$$

$$E_1 = \{T, S\} = \{A \in V : A \rightarrow X_1 \dots X_k, k \geq 1, X_i \in E_0\} \cup E_0$$

$$E_2 = \{A \in V : A \rightarrow X_1 \dots X_k, k \geq 1, X_i \in E_1\} \cup E_1 = E_1 = E$$

La première chose à faire est d'ajouter une nouvelle variable  $S'$  avec de nouvelles règles:

$$V' = V \cup \{S'\}, S' \notin V$$

$$R' = (R \cup \{S' \rightarrow \alpha : S \rightarrow \alpha, \alpha \neq \varepsilon\}) =$$

$$\{S \rightarrow aSa \mid aTa \mid S \mid T,$$

$$T \rightarrow aTa \mid bTb \mid \varepsilon,$$

$$S' \rightarrow aSa \mid aTa \mid S \mid T\}$$

La règle  $S' \rightarrow \varepsilon$  sera ajoutée à la fin pour remettre  $\varepsilon$  dans le langage (car  $S \in E$ , il le faut). On enlève maintenant les règles de la forme  $A \rightarrow \varepsilon$ . On remplace chaque règle  $r = A \rightarrow \alpha$  de  $R$  par un ensemble de règles obtenu en remplaçant certaines variables de  $\alpha$  par  $\varepsilon$ . "Certaines" veut dire "tous les choix possibles".

$S \rightarrow aSa$  devient  $\{S \rightarrow aSa \mid aa\}$ ,  $S \rightarrow aTa$  devient  $\{S \rightarrow aTa \mid aa\}$ , etc. pour nous donner

$$R'' = (R' \cup \{S \rightarrow aa, T \rightarrow aa \mid bb, S' \rightarrow aa \mid bb \mid S \mid T\}) \setminus \{T \rightarrow \varepsilon\} = \\ \{S \rightarrow aa \mid aSa \mid aTa \mid S \mid T, \\ T \rightarrow aa \mid bb \mid aTa \mid bTb \mid, \\ S' \rightarrow aa \mid aSa \mid aTa \mid S \mid T\}$$

Ensuite on enlève les règles unitaires, i.e. de la forme  $A \rightarrow B$ ,  $A, B \in V$ . On a que  $S \xrightarrow{*} T \mid S$  et  $T \xrightarrow{*} S \mid T$ .

(enlever  $S \xrightarrow{*} S \mid T$ )

$$R_{\{S\}} = \{ \\ S \rightarrow aa \mid bb \mid aSa \mid aTa \mid bTb, \\ T \rightarrow aa \mid bb \mid aTa \mid bTb, \\ S' \rightarrow aa \mid bb \mid aSa \mid aTa \mid T\}$$

(enlever  $S' \rightarrow S \mid T$ )

$$R_{\{S, S'\}} = \{ \\ S \rightarrow aa \mid bb \mid aSa \mid aTa \mid bTb, \\ T \rightarrow aa \mid bb \mid aTa \mid bTb, \\ S' \rightarrow aa \mid bb \mid aSa \mid aTa \mid bTb, \}$$

On ajoute les nouvelles variables  $X_x$  pour chaque  $x \in \Sigma$  :

$$V_1 = \{S, T, S', X_a, X_b\} \\ R_1 = \{X_a \rightarrow a, \\ X_b \rightarrow b, \\ S \rightarrow X_a X_a \mid X_a S X_a \mid X_a T X_a \mid X_b T X_b, \\ T \rightarrow X_a X_a \mid X_b X_b \mid X_a T X_a \mid X_b T X_b, \\ S' \rightarrow X_a X_a \mid X_a S X_a \mid X_a T X_a \mid X_b T X_b \\ \}$$

On convertit chaque règle avec plus que deux variables à droite en une série de de règles avec juste deux variables droite; il faut ajouter des nouvelles variables. On ajoute aussi  $S' \rightarrow \varepsilon$ . La grammaire finale est  $\hat{G} = \hat{V}, \Sigma, \hat{R}, S'$  avec

$$\hat{V} = \{S, T, S', X_a, X_b, X_S, X_T, Y_T\} \\ \hat{R} = \{ \\ X_a \rightarrow a, \\ X_b \rightarrow b, \\ S \rightarrow X_a X_a \mid X_a X_S \mid X_a X_T \mid X_b Y_T, \\ T \rightarrow X_a X_a \mid X_b X_b \mid X_a X_T \mid X_b Y_T, \\ S' \rightarrow X_a X_a \mid X_a X_S \mid X_a X_T \mid X_b Y_T, \\ X_S \rightarrow S X_a, \\ X_T \rightarrow T X_a, \\ Y_T \rightarrow T X_b, \\ S' \rightarrow \varepsilon\}$$

## 2 Version Sipser

Mais on veut toujours que toutes les variables soient utiles. On construit d'abord une grammaire équivalente qui ne génère pas  $\varepsilon$ . Si le mot vide appartient au langage, on ajoutera à la fin.

Soit  $G = (V, \Sigma, R, S)$  la grammaire définie par

$$V = \{S, A, B, C, D\}$$

$$\Sigma = \{a, b\}$$

R:

$$S \longrightarrow ABAC \mid SS \mid \varepsilon$$

$$A \longrightarrow aBa \mid C \mid a \mid \varepsilon$$

$$B \longrightarrow bAb \mid CC \mid b \mid \varepsilon$$

$$C \longrightarrow CC \mid \varepsilon$$

$$D \longrightarrow ASB \mid BB \mid C \mid abba$$

Mettons  $G_0 = G$ .

1. Les variables doivent mener aux mots.

On a que  $A \longrightarrow a$ ,  $B \longrightarrow b$  et  $S, C, D \longrightarrow \varepsilon$ . De plus si on cherche l'ensemble de variables accessibles à partir de  $S$ , on trouve :

$$V_0 = \{S\}$$

$$V_1 = \{S, A, B, C\}$$

$$V_2 = V_1$$

et on conclut que  $D$  est inutile. La grammaire  $G_1$  est alors :

$$V = \{S, A, B, C\}$$

$$\Sigma = \{a, b\}$$

R:

$$S \longrightarrow ABAC \mid SS \mid \varepsilon$$

$$A \longrightarrow aBa \mid C \mid a \mid \varepsilon$$

$$B \longrightarrow bAb \mid CC \mid b \mid \varepsilon$$

$$C \longrightarrow CC \mid \varepsilon$$

2. Eliminons  $\varepsilon$ .

- $S \longrightarrow \varepsilon$  : puisque  $S \longrightarrow SS$ , on devrait ajouter  $S \longrightarrow S$ , mais cela n'a aucun effet. Les règles pour  $S$  deviennent  $S \longrightarrow ABAC \mid SS$  et les autres ne changent pas; on a  $G_1$ .
- $A \longrightarrow \varepsilon$  : les règles avec  $A$  à droite sont  $S \longrightarrow ABAC$  et  $B \longrightarrow bAb$ . On obtient donc les règles de la grammaire  $G_2$  :
 
$$S \longrightarrow ABAC \mid BAC \mid ABC \mid BC \mid SS$$

$$A \longrightarrow aBa \mid C \mid a$$

$$B \longrightarrow bAb \mid bb \mid CC \mid b \mid \varepsilon$$

$$C \longrightarrow CC \mid \varepsilon$$

- $B \rightarrow \varepsilon$  : cette fois beaucoup de changements pour obtenir  $G_3$ :  
 $S \rightarrow ABAC \mid AAC \mid BAC \mid ABC \mid AC \mid BC \mid SS$   
 $A \rightarrow aBa \mid aa \mid C \mid a$   
 $B \rightarrow bAb \mid bb \mid CC \mid b$   
 $C \rightarrow CC \mid \varepsilon$
- $C \rightarrow \varepsilon$  : on enlève  $C \rightarrow \varepsilon$  et n'ajoute pas  $C \rightarrow C$ . La grammaire  $G_4$  devient  
 $S \rightarrow ABAC \mid ABA \mid AAC \mid AA \mid BAC \mid BA \mid ABC \mid AB \mid AC \mid A \mid BC \mid B \mid SS$   
 $A \rightarrow aBa \mid aa \mid C \mid a$   
 $B \rightarrow bAb \mid bb \mid CC \mid b$   
 $C \rightarrow CC$

Ici, on se rend compte que  $C$  ne sert plus à rien. Il serait donc mieux d'éliminer les variables inutiles à la fin car en ce moment de la procédure on a ce  $C$  inutile qui restera!

### Recommandons.

Soit  $G = (V, \Sigma, R, S)$  la grammaire définie par

$$V = \{S, A, B, C, D\}$$

$$\Sigma = \{a, b\}$$

R:

$$S \rightarrow ABAC \mid SS \mid \varepsilon$$

$$A \rightarrow aBa \mid C \mid a \mid \varepsilon$$

$$B \rightarrow bAb \mid CC \mid b \mid \varepsilon$$

$$C \rightarrow CC \mid \varepsilon$$

$$D \rightarrow ASB \mid BB \mid C \mid abba$$

Mettons  $G_0 = G$ .

#### 1. Eliminons $\varepsilon$ .

- $S \rightarrow \varepsilon$  : puisque  $S \rightarrow SS$ , on devrait ajouter  $S \rightarrow S$ , mais cela n'a aucun effet. Les règles pour  $S$  deviennent  
 $S \rightarrow ABAC \mid SS$  et les autres ne changent pas; on a  $G_1$ .
- $A \rightarrow \varepsilon$  : les règles avec  $A$  à droite sont  $S \rightarrow ABAC$  et  $B \rightarrow bAb$ . On obtient donc les règles de la grammaire  $G_2$  :  
 $S \rightarrow ABAC \mid BAC \mid ABC \mid BC \mid SS$   
 $A \rightarrow aBa \mid C \mid a$   
 $B \rightarrow bAb \mid bb \mid CC \mid b \mid \varepsilon$   
 $C \rightarrow CC \mid \varepsilon$   
 $D \rightarrow ASB \mid AB \mid BB \mid C \mid abba$
- $B \rightarrow \varepsilon$  : cette fois beaucoup de changements pour obtenir  $G_3$ :  
 $S \rightarrow ABAC \mid AAC \mid BAC \mid ABC \mid AC \mid BC \mid SS$   
 $A \rightarrow aBa \mid aa \mid C \mid a$

$$\begin{aligned}
B &\longrightarrow bAb \mid bb \mid CC \mid b \\
C &\longrightarrow CC \mid \varepsilon \\
D &\longrightarrow ASB \mid AB \mid BB \mid B \mid C \mid abba
\end{aligned}$$

- $C \longrightarrow \varepsilon$  : on enlève  $C \longrightarrow \varepsilon$  et n'ajoute pas  $C \longrightarrow C$ . La grammaire  $G_4$  devient
$$\begin{aligned}
S &\longrightarrow ABAC \mid ABA \mid AAC \mid AA \mid BAC \mid BA \mid ABC \mid AB \mid AC \mid A \mid BC \mid B \mid SS \\
A &\longrightarrow aBa \mid aa \mid C \mid a \\
B &\longrightarrow bAb \mid bb \mid CC \mid b \\
C &\longrightarrow CC \\
D &\longrightarrow ASB \mid AB \mid BB \mid B \mid C \mid abba
\end{aligned}$$

2. Eliminons les règles unitaires.

- $A \longrightarrow C$ . On l'enlève et on ajoute ce qu'il faut pour obtenir (car  $A \longrightarrow C$  et  $C \longrightarrow CC$ )  $G_5$ 

$$\begin{aligned}
S &\longrightarrow ABAC \mid ABA \mid AAC \mid AA \mid BAC \mid BA \mid ABC \mid AB \mid AC \mid A \mid BC \mid B \mid SS \\
A &\longrightarrow aBa \mid aa \mid CC \mid a \\
B &\longrightarrow bAb \mid bb \mid CC \mid b \\
C &\longrightarrow CC \\
D &\longrightarrow ASB \mid AB \mid BB \mid B \mid C \mid abba
\end{aligned}$$
- $S \longrightarrow A$ , après son enlèvement, on a  $G_6$ 

$$\begin{aligned}
S &\longrightarrow ABAC \mid ABA \mid AAC \mid AA \mid BAC \mid BA \mid ABC \mid AB \mid AC \mid BC \mid B \mid SS \\
S &\longrightarrow aBaaBa \mid aaaa \mid CCCC \mid aa \\
A &\longrightarrow aBa \mid aa \mid CC \mid a \\
B &\longrightarrow bAb \mid bb \mid CC \mid b \\
C &\longrightarrow CC \\
D &\longrightarrow ASB \mid AB \mid BB \mid B \mid C \mid abba
\end{aligned}$$
- $S \longrightarrow B$  donne  $G_7$ 

$$\begin{aligned}
S &\longrightarrow ABAC \mid ABA \mid AAC \mid AA \mid BAC \mid BA \mid ABC \mid AB \mid AC \mid BC \mid SS \\
S &\longrightarrow aBaaBa \mid aaaa \mid CCCC \mid aa \\
S &\longrightarrow bAbbAB \mid bbbb \mid bb \\
A &\longrightarrow aBa \mid aa \mid CC \mid a \\
B &\longrightarrow bAb \mid bb \mid CC \mid b \\
C &\longrightarrow CC \\
D &\longrightarrow ASB \mid AB \mid BB \mid B \mid C \mid abba
\end{aligned}$$
- $D \longrightarrow B \mid C$  - leur enlèvement n'a aucun effet sur le reste et  $G_8$  est.
$$\begin{aligned}
S &\longrightarrow ABAC \mid ABA \mid AAC \mid AA \mid BAC \mid BA \mid ABC \mid AB \mid AC \mid BC \mid SS \\
S &\longrightarrow aBaaBa \mid aaaa \mid CCCC \mid aa \\
S &\longrightarrow bAbbAb \mid bbbb \mid bb \\
A &\longrightarrow aBa \mid aa \mid CC \mid a \\
B &\longrightarrow bAb \mid bb \mid CC \mid b \\
C &\longrightarrow CC \\
D &\longrightarrow ASB \mid AB \mid BB \mid abba
\end{aligned}$$

3. Eliminons les variables inutiles. Pour les variables qui mènent aux mots, on a

$$V_0 = \{A, B\}$$

$$V_1 = \{A, B, S\} \quad V_2 = V_1$$

Donc la grammaire  $G_9$  est

$$S \longrightarrow ABA \mid AA \mid BA \mid AB \mid SS$$

$$S \longrightarrow aBaaBa \mid aaaa \mid aa$$

$$S \longrightarrow bAbbAb \mid bbbb \mid bb$$

$$A \longrightarrow aBa \mid aa \mid a$$

$$B \longrightarrow bAb \mid bb \mid b$$

$$D \longrightarrow ASB \mid AB \mid BB \mid abba$$

Pour les variables inaccessibles :

$$V_0\{S\}$$

$$V_1 = \{S, A, B\} = V_2$$

et donc la grammaire presque finale  $G_{10}$  est

$$S \longrightarrow ABA \mid AA \mid BA \mid AB \mid SS$$

$$S \longrightarrow aBaaBa \mid aaaa \mid aa$$

$$S \longrightarrow bAbbAb \mid bbbb \mid bb$$

$$A \longrightarrow aBa \mid aa \mid a$$

$$B \longrightarrow bAb \mid bb \mid b$$

4. Il reste à s'assurer que le symbole de départ n'est jamais à droite dans un règle. On fait donc les changements suivants pour avoir  $G_{11}$

$$V' = \{S', S, A, B\}$$

$$S' \longrightarrow ABA \mid AA \mid BA \mid AB \mid SS$$

$$S \longrightarrow ABA \mid AA \mid BA \mid AB \mid SS$$

$$S \longrightarrow aBaaBa \mid aaaa \mid aa$$

$$S \longrightarrow bAbbAb \mid bbbb \mid bb$$

$$A \longrightarrow aBa \mid aa \mid a$$

$$B \longrightarrow bAb \mid bb \mid b$$

5. Puisque  $\varepsilon$  est dans le langage de  $G$ , il faut l'ajouter. Ceci est très facile, on ajoute la règle  $S' \longrightarrow \varepsilon$ . La grammaire finale  $\hat{G}$  est donc

$$V' = \{S', S, A, B\}$$

$$S' \longrightarrow ABA \mid AA \mid BA \mid AB \mid SS \mid \varepsilon$$

$$S \longrightarrow ABA \mid AA \mid BA \mid AB \mid SS$$

$$S \longrightarrow aBaaBa \mid aaaa \mid aa$$

$$S \longrightarrow bAbbAb \mid bbbb \mid bb$$

$$A \longrightarrow aBa \mid aa \mid a$$

$$B \longrightarrow bAb \mid bb \mid b$$

**Exercice 1** *On a ici un exemple qui montre que les variables inutiles doivent être enlevées à la fin. Trouver un exemple pour montrer qu'il faut enlever les règles  $\varepsilon$  avant les règles unitaires.*