

Simplifications & Formes Normales

Simplifications et formes normales

- Si $\varepsilon \notin L$, on n'a pas besoin de production de la forme $A \rightarrow \varepsilon$
- Toutes les productions peuvent être de la forme $A \rightarrow BC \mid a$ ($A, B, C \in N$ et $a \in T$)
(Forme Normale de Chomsky)
- Toutes les productions peuvent être de la forme $A \rightarrow a\alpha$ ($a \in T$ et $\alpha \in N^*$)
(Forme Normale de Greibach)
- Supprimer les règles de renommage (productions de la forme $A \rightarrow B$)

Suppression des ε -productions

Règles de la forme $A \rightarrow \varepsilon$

Suppression des ε -productions

- Observation: si $\varepsilon \in L(G)$, il sera impossible de supprimer toutes les ε -productions de G ; Si $\varepsilon \notin L(G)$, aucune règle de la forme $A \rightarrow \varepsilon$ n'est utile.
- Principe:
 - Pour chaque variable A , déterminer si $A \rightarrow^* \varepsilon$. Si c'est le cas, on dit que A est «**effaçable**»
 - Remplacer toute production de la forme $B \rightarrow X_1 X_2 \dots X_n$ par toutes les productions où on a effacé les variables X_i effaçables sans ajouter $B \rightarrow \varepsilon$, même si tous les X_i sont effaçables

Variables effaçables

- Une variable A est effaçable si $A \rightarrow^* \varepsilon$
- L'ensemble des variables effaçables de G est
$$\text{Eff}(G) = \{X \in N : X \rightarrow^* \varepsilon\}$$

$$X \in \text{Eff}(G)$$

SSI

$$X \rightarrow \varepsilon \in R$$

ou

$$X \rightarrow X_1 X_2 \dots X_n \in R \text{ pour } X_i \in \text{Eff}(G) \ 0 \leq i \leq n$$

Variables effaçables

Construction inductive:

Base : $\text{Eff}_0(G) = \{X : X \rightarrow \varepsilon \in R\}$

Induction :

$\text{Eff}_i(G) = \text{Eff}_{i-1}(G) \cup \{Y : Y \rightarrow Z_1 Z_2 \dots Z_n \in R \text{ pour } Z_k \in \text{Eff}_{i-1}(G) \ 0 \leq k \leq n+1\}$

On s'arrête dès que $\text{Eff}_i(G) = \text{Eff}_{i-1}(G)$

Exemple

$S \rightarrow aAb \mid AB \mid a$

$A \rightarrow \varepsilon \mid AAB$

$B \rightarrow AC \mid b$

$C \rightarrow \varepsilon \mid aba$

Initialisation : variables de la forme $A \rightarrow \varepsilon$:

$$\text{Eff}_0(G) = \{A, C\}$$

variables de la forme

$$A \rightarrow \alpha, \alpha \in (\text{Eff}_0(G))^*$$

$$\text{Eff}_1(G) = \{A, C\} \cup \{B\}$$

variables de la forme

$$A \rightarrow \alpha, \alpha \in (\text{Eff}_1(G))^*$$

$$\text{Eff}_2(G) = \{B, A, C\} \cup \{S\}$$

Toutes les variables sont effaçables et $\varepsilon \in L(G)$

Supprimer les ε -productions

- On construit un nouvel ensemble de règles R' .
- $X \rightarrow X_1 X_2 \dots X_n \in R$, $X_i \in (N \cup T)$, remplacée par

$$X \rightarrow \alpha_1 \alpha_2 \dots \alpha_n$$

Si $X_i \notin \text{Eff}(G)$, alors $\alpha_i = X_i$

Si $X_i \in \text{Eff}(G)$, alors $\alpha_i = X_i$ et $\alpha_i = \varepsilon$

Les α_i sont non tous nuls

(sauf lorsque toutes les variables sont effacées)

Exemple

- $S \rightarrow aAb | AB | a | ab | A | B$
- $A \rightarrow \cancel{x} | AAB | A | B | AB | AA$
- $B \rightarrow AC | b | A | C$
- $C \rightarrow \cancel{x} | aba$
- Dans G , $S \rightarrow \varepsilon$;
si on veut engendrer le même langage, on ajoute $S \rightarrow \varepsilon$
- $\text{Eff}(G) = \{B, A, C, S\}$
- En retirant les ε -productions, on obtiendra $L \setminus \{\varepsilon\}$

Suppression des renommages

Règles de la forme $X \rightarrow Y$ pour $X, Y \in N$

Les variables de renommage

Les variables en lesquelles X peut être renommée sont : $\text{Ren}(X) = \{Z \in N : X \rightarrow^* Z\}$

Construction inductive, pour chaque variable X , calculer $\text{Ren}(X)$

Base : $\text{Ren}_0(X) = \{X\}$

Induction :

$\text{Ren}_i(X) = \text{Ren}_{i-1}(X) \cup \{Z : Y \rightarrow Z \in R \text{ pour } Y \in \text{Ren}_{i-1}(X)\}$

On s'arrête dès que $\text{Ren}_i(X) = \text{Ren}_{i-1}(X)$

Supprimer les règles de renommage : pour chaque $Y \in \text{Ren}(X)$ et chaque règle $Y \rightarrow \alpha$ qui n'est pas renommage, remplacer $X \rightarrow Y$ par $X \rightarrow \alpha$

Example $\text{Ren}_i(X) = \text{Ren}_{i-1}(X) \cup \{Z: Y \rightarrow Z \in R, Y \in \text{Ren}_{i-1}(X)\}$

$S \rightarrow ab | aAb | A | B | AB | a$
 $A \rightarrow A | AA | AB | B | AAB$
 $B \rightarrow A | C | AC | b$

$C \rightarrow aba$

- $\text{Ren}_0(S) = \{S\}$
- $\text{Ren}_1(S) = \{Z: S \rightarrow Z \in R\} \cup \{S\} = \{S, A, B\}$
- $\text{Ren}_2(S) = \{Z: Y \rightarrow Z \in R, Y \in \{S, A, B\}\} \cup \{S, A, B\} = \{S, A, B, C\}$
- $\text{Ren}_0(A) = \{A\}$
- $\text{Ren}_1(A) = \{Z: A \rightarrow Z \in R\} \cup \{A\} = \{A, B\}$
- $\text{Ren}_2(A) = \{Z: Y \rightarrow Z \in R, Y \in \{A, B\}\} \cup \{A, B\} = \{A, B, C\}$
- $\text{Ren}_0(B) = \{B\}$
- $\text{Ren}_1(B) = \{Z: B \rightarrow Z \in R\} \cup \{B\} = \{A, B, C\}$
- $\text{Ren}_2(B) = \{Z: Y \rightarrow Z \in R, Y \in \{A, B, C\}\} \cup \{A, B\} = \{A, B, C\}$
- $\text{Ren}_0(C) = \{C\}$
- $\text{Ren}_1(C) = \{Z: C \rightarrow Z \in R\} \cup \{C\} = \{C\}$

Exemple

- $\text{Ren}(S) = \{S, A, B, C\}$
- $\text{Ren}(A) = \{A, B, C\}$
- $\text{Ren}(B) = \{A, B, C\}$
- $\text{Ren}(C) = \{C\}$

$S \rightarrow ab | aAb | \cancel{X} | \cancel{B} | AB | a$

$A \rightarrow \cancel{X} | AA | AB | \cancel{B} | AAB$

$B \rightarrow \cancel{X} | \cancel{X} | AC | b$

$C \rightarrow aba$

$\left\{ \begin{array}{l} \cdot S \rightarrow ab | aAb | AB | a | AA | AB | AAB | AC | b | aba \\ \cdot A \rightarrow AA | AB | AAB | AC | b | aba \\ \cdot B \rightarrow AA | AB | AAB | AC | b | aba \end{array} \right\}$ A et B sont identiques
 $\cdot C \rightarrow aba$

$\left\{ \begin{array}{l} \cdot S \rightarrow ab | aAb | AA | a | AAA | AC | b | aba \\ \cdot A \rightarrow AA | AAA | AC | b | aba \\ \cdot C \rightarrow aba \end{array} \right.$ On peut remonter $C \rightarrow aba$
 $\left\{ \begin{array}{l} \cdot S \rightarrow ab | aAb | AA | a | AAA | Aaba | b | aba \\ \cdot A \rightarrow AA | AAA | Aaba | b | aba \end{array} \right.$

Conclusion

Tout langage algébrique ne contenant pas le mot vide peut être engendré par une grammaire sans symbole inutile ni ε -production, ni règle de renommage.

Formes normales

- On peut mettre toutes les productions sous la forme $A \rightarrow BC|a$ avec $A, B, C \in N$ et $a \in T$ (FNC)

Forme normale de Chomsky

Forme normale de Chomsky

- **Théorème** : Toute grammaire algébrique sans renommage ni ε -production est équivalente à une grammaire dont les productions sont de la forme
$$X \rightarrow a, \text{ pour } a \in T \text{ et } X \rightarrow YZ, \text{ pour } Y, Z \in N$$
- Il faut coder un arbre d'arité finie quelconque en un arbre binaire.
- Si une règle est de la forme $X \rightarrow \alpha_1 \alpha_2 \dots \alpha_n$ on regroupe les variables du membre droit en deux paquets en introduisant des productions supplémentaires

Démonstration

Dans chaque production où a est dans le membre droit : $X \rightarrow aY$ ou $X \rightarrow Ya$

On remplace a par C_a : avec $C_a \rightarrow a$; $X \rightarrow C_aY$ ou $X \rightarrow YC_a$

Les règles sont de la forme (G sans règle de renommage)

$$X \rightarrow \alpha_1 \alpha_2 \dots \alpha_n \qquad \alpha_i \in N$$

$$X \rightarrow a \qquad a \in T$$

Démonstration (2)

- On introduit de nouvelles variables pour obtenir à partir de

$$X \rightarrow \alpha_1 \underbrace{\alpha_2 \dots \alpha_{n-1} \alpha_n}$$

$$Y_1 \rightarrow \alpha_2 \underbrace{\alpha_3 \dots \alpha_{n-1} \alpha_n}$$

$$Y_2 \rightarrow \alpha_3 \dots \alpha_{n-1} \alpha_n$$

...

$$Y_{n-2} \rightarrow \alpha_{n-1} \alpha_n$$

- A présent toutes les règles sont de la forme

$$X \rightarrow \alpha\beta \quad \alpha, \beta \in N \text{ ou } X \rightarrow a \quad a \in T$$

Exemple

- La grammaire ETF

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid a$$

- $\text{Ren}(E) = \{E, T, F\}$

- $\text{Ren}(T) = \{T, F\}$

- $\text{Ren}(F) = \{F\}$

hypothèse de mise sous FNC: pas de règle de renommage

On supprime les règles de renommage

$$E \rightarrow E + T$$

$$T \rightarrow T * F$$

$$F \rightarrow (E) \mid a$$

$$E \rightarrow E + T \mid T * F \mid (E) \mid a$$

$$T \rightarrow T * F \mid (E) \mid a$$

$$F \rightarrow (E) \mid a$$

On peut maintenant mettre cette grammaire sous FNC

Exemple

$$E \rightarrow E + T$$

$$E \rightarrow T * F$$

$$E \rightarrow (E)$$

$$E \rightarrow a$$

$$E \rightarrow EX_1 \text{ et } X_1 \rightarrow C_+ T$$

$$E \rightarrow TX_2 \text{ et } X_2 \rightarrow C_* F$$

$$E \rightarrow C_-(X_3 \text{ et } X_3 \rightarrow E C_)$$

$$E \rightarrow a$$

$$C_-(\rightarrow ($$

$$C_) \rightarrow)$$

$$C_* \rightarrow *$$

$$C_+ \rightarrow +$$

$$E \rightarrow EX_1 | TX_2 | C_-(X_3 | a$$

$$X_1 \rightarrow C_+ T$$

$$X_2 \rightarrow C_* F$$

$$X_3 \rightarrow E C_)$$

Exemple

$$T \rightarrow T^*F$$

$$T \rightarrow (E)$$

$$T \rightarrow a$$

$$T \rightarrow TY_1 \text{ et } Y_1 \rightarrow C^*F$$

$$T \rightarrow C(Y_2 \text{ et } Y_2 \rightarrow E C)$$

$$T \rightarrow a$$

$$T \rightarrow TY_1 \mid C(Y_2 \mid a$$

$$Y_1 \rightarrow C^*F$$

$$Y_2 \rightarrow E C$$

$$F \rightarrow (E)$$

$$F \rightarrow a$$

$$F \rightarrow C(Z_1 \text{ et } Z_1 \rightarrow EC)$$

$$F \rightarrow a$$

Exemple

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid a$$

$$C_{(} \rightarrow (\quad C_{)} \rightarrow) \quad C_{*} \rightarrow * \quad C_{+} \rightarrow +$$

$$E \rightarrow EX_1 \mid TX_2 \mid C_{(}X_3 \mid a$$

$$X_1 \rightarrow C_{+}T$$

$$X_2 \rightarrow C_{*}F$$

$$X_3 \rightarrow EC_{)}$$

$$T \rightarrow TY_1 \mid C_{(}Y_2 \mid a$$

$$Y_1 \rightarrow C_{*}F$$

$$Y_2 \rightarrow EC_{)}$$

$$F \rightarrow C_{(}Z_1$$

$$Z_1 \rightarrow EC_{)}$$

$$F \rightarrow a$$

Exemple

$$E \rightarrow E + T \mid T$$

$$T \rightarrow T * F \mid F$$

$$F \rightarrow (E) \mid a$$

$$C_{(} \rightarrow ($$

$$C_{)} \rightarrow)$$

$$C_{*} \rightarrow *$$

$$C_{+} \rightarrow +$$

$$E \rightarrow EX_1 \mid TX_2 \mid C_{(}X_3 \mid a$$

$$X_1 \rightarrow C_{+}T$$

$$X_2 \rightarrow C_{*}F$$

$$X_3 \rightarrow EC_{)}$$

$$T \rightarrow TX_2 \mid C_{(}X_3 \mid a$$

$$F \rightarrow C_{(}X_3$$

$$F \rightarrow a$$

On peut supprimer
les variables
identiques (ayant
les mêmes règles)

Suppression de la récursivité gauche

Pour la forme normale de Greibach on a besoin de la suppression de la récursivité gauche

Suppression récursivité gauche

Lemme : Pour toute grammaire algébrique, il existe une grammaire algébrique équivalente sans récursivité gauche i.e. sans règle de la forme

$$A \rightarrow A\alpha \quad \alpha \in (N \cup T)^*$$

- Transformer la récursivité gauche en récursivité droite.
- On suppose que G ne contient que des règles de la forme $A \rightarrow A\alpha \mid \beta$ avec $\alpha, \beta \in (N \cup T)^*$ où β ne commence pas par A .
- On peut dériver $\beta, \beta\alpha, \beta\alpha\alpha, \beta\alpha\alpha\alpha, \beta\alpha\alpha\alpha\alpha, \dots$
- Ainsi, ce couple de règles se dérive en $\beta\alpha^*$
- Qu'on peut engendrer par les règles sans récursivité gauche
 - $A \rightarrow \beta \mid \beta X$
 - $X \rightarrow \alpha \mid \alpha X$

Plus généralement

- On remplace toute règle de la forme

$$A \rightarrow A\alpha_1 \mid A\alpha_2 \mid \dots \mid A\alpha_n \mid \beta_1 \mid \beta_2 \mid \dots \mid \beta_m$$

- Qui engendre une expression de la forme

$$(\beta_1 + \beta_2 + \dots + \beta_m)(\alpha_1 + \alpha_2 + \dots + \alpha_n)^*$$

- Par les règles

- $A \rightarrow \beta_1 \mid \beta_2 \mid \dots \mid \beta_m \mid \beta_1 B \mid \beta_2 B \mid \dots \mid \beta_m B$

- $B \rightarrow \alpha_1 \mid \alpha_2 \mid \dots \mid \alpha_n \mid \alpha_1 B \mid \alpha_2 B \mid \dots \mid \alpha_n B$

Exemple

$$\begin{aligned} A &\rightarrow \beta | \beta X \\ X &\rightarrow \alpha | \alpha X \end{aligned}$$

- $E \rightarrow EX_1 | TX_2 | C(X_3 | a$

$$A \rightarrow A\alpha_1 | A\alpha_2 | \dots | A\alpha_n | \beta_1 | \beta_2 | \dots | \beta_m$$

- est transformée en

- $E \rightarrow TX_2 | C(X_3 | a | TX_2 B | C(X_3 B | aB$

- $B \rightarrow X_1 | X_1 B$

$$A \rightarrow \beta_1 | \beta_2 | \dots | \beta_m | \beta_1 B | \beta_2 B | \dots | \beta_m B$$

$$B \rightarrow \alpha_1 | \alpha_2 | \dots | \alpha_n | \alpha_1 B | \alpha_2 B | \dots | \alpha_n B$$

Forme normale de Greibach

Théorème

Théorème : Toute grammaire algébrique, sans renommage ni ε -production est équivalente à une grammaire dont toutes les productions sont de la forme

$$X \rightarrow a\gamma, \quad \gamma \in (N \cup T)^* \text{ et } a \in T$$

On cherche

- à supprimer la récursivité gauche des règles, i.e. à éviter des règles de la forme $A \rightarrow A\alpha$, $\alpha \in (N \cup T)^*$
 - à faire commencer toute règle par un terminal.
- pour faciliter l'analyse d'une chaîne de caractères

Mise sous FNG (1)

1. On fixe un ordre sur les variables de $G: N=(A_1, A_2, \dots, A_n)$ et on dit que la grammaire est **montante** si toute règle est de la forme $A_j \rightarrow A_k v$ avec $j < k$ ou $A_j \rightarrow av$ avec $a \in T$
2. On suppose les variables de G ordonnées et ses règles de la forme $A \rightarrow A_{i1} A_{i2} \dots A_{in}$ tq tous les $A_{ij} \in N$ ou $A \rightarrow b$ tq $b \in T$

Mise sous FNG (2)

3. On pose $i = 1$. Tant que $i < n$, on considère les productions P de la forme $A_i \rightarrow \alpha$:
- Si $i > 1$, on remplace toutes les occurrences de A_k , $k < i$ dans les productions $A_i \rightarrow A_k \alpha$ avec $k < i$ et $\alpha \in V^*$
 - Soit P est montante OK
 - Soit P réursive gauche
 - on retire la réc. gauche en ajoutant un nouveau non-terminal numéroté à la suite de ceux de N' et on incrémente n de 1
4. pour i de $n-1$ à 1 on substitue dans les $A_i \rightarrow A_j \alpha$ avec $j > i$ les A_j par leurs parties droites.

Exemple

$$\underline{A_1 \rightarrow A_2 A_2 | 0}$$

$$\underline{A_2 \rightarrow A_1 A_2 | 1}$$

$$\underline{A_2 \rightarrow A_2 A_2 A_2 | 0 A_2 | 1}$$

$$A_2 \rightarrow 0 A_2 | 1 | 0 A_2 A_3 | 1 A_3$$

$$\underline{A_3 \rightarrow A_2 A_2 | A_2 A_2 A_3}$$

$$\begin{array}{l} A_3 \rightarrow 0 A_2 A_2 | 0 A_2 A_2 A_3 \\ \quad | 1 A_2 | 1 A_2 A_3 \\ \quad | 0 A_2 A_3 A_2 | 0 A_2 A_3 A_2 A_3 \\ \quad | 1 A_3 A_2 | 1 A_3 A_2 A_3 \end{array}$$

$$A_1 \rightarrow 0 A_2 A_2 | 1 A_2 | 0 A_2 A_3 A_2 | 1 A_3 A_2 | 0$$

A_1 montante. On passe à A_2

On copie le membre droit de A_1 à la place de A_1 dans $A_2 \rightarrow A_1$ v:

A_2 est réursive gauche, on la transforme grâce au lemme:

On copie le membre droit de A_2 à la place de A_2 dans $A_3 \rightarrow A_2$ v:

A_3 est sous FNG

A_2 est sous FNG

On copie le membre droit de A_2 à la place de A_2 dans $A_1 \rightarrow A_2$ v:

FNG