

Langages et Traducteurs

Analyse Descendante

Définitions

Soit $G = (V_t, V_n, Z, P)$ une grammaire hors-contexte. On définit les notions suivantes :

- soit $\alpha \in (V_t \cup V_n)^*$, le prédicat **Vide**(α) vaut vrai ssi ε peut être dérivé de α par G :

$$\mathbf{Vide}(\alpha) \equiv \alpha \Rightarrow_G^* \varepsilon$$

- soit $\alpha \in (V_t \cup V_n)^*$, **Premiers**(α) est l'ensemble des symboles terminaux qui peuvent débiter une dérivation de α par G :

$$\mathbf{Premier}(\alpha) = \{a \in V_t \mid \alpha \Rightarrow_G^* a\beta\}$$

- soit $X \in V_n$, **Suivants**(X) est l'ensemble des symboles terminaux qui peuvent suivre une occurrence de X dans une dérivation de Z par G :

$$\mathbf{Suivants}(X) = \{a \in V_t \mid Z \Rightarrow_G^* \alpha X a \beta\}$$

- soit $X \rightarrow u_1.u_2 \dots u_n$ une règle de P , avec $u_i \in (V_t \cup V_n)$. **Directeurs**($X \rightarrow u_1.u_2 \dots u_n$) est l'ensemble des symboles terminaux qui peuvent débiter une dérivation par cette règle :

$$\mathbf{Directeurs}(X \rightarrow u_1.u_2 \dots u_n) = \left(\bigcup_{i \in V_i} \mathbf{Premiers}(u_i) \right) \cup (\text{si } \mathbf{Vide}(u_1 \dots u_n) \text{ alors } \mathbf{Suivants}(X) \text{ sinon } \emptyset)$$

avec $V_i = \{i \mid \mathbf{Vide}(u_1 \dots u_{i-1})\}$.

Une grammaire hors-contexte $G = (V_t, V_n, Z, P)$ est dite *LL(1)* si et seulement si l'ensemble des règles de P définissant un même symbole non-terminal ont des directeurs disjoints :

$$\forall X \in V_t. \forall X \rightarrow \alpha, X \rightarrow \beta \in P. \mathbf{Directeur}(X \rightarrow \alpha) \cap \mathbf{Directeur}(X \rightarrow \beta) = \emptyset$$

Un langage hors-contexte est dit *LL(1)* si et seulement si il existe une grammaire *LL(1)* qui le reconnaît. L'ensemble des langages *LL(1)* est *strictement inclus* dans l'ensemble des langages hors-contexte.

Exemple 1

On considère la grammaire G_1 suivante :

$$\begin{aligned} Z &\rightarrow S\# \\ S &\rightarrow Xa \\ S &\rightarrow \varepsilon \\ X &\rightarrow Xb \\ X &\rightarrow Y \\ Y &\rightarrow aX \\ Y &\rightarrow \varepsilon \end{aligned}$$

On a alors les résultats suivants :

$$\begin{aligned}
\mathbf{Premier}(Z) &= \mathbf{Premier}(S\#) \\
&= \mathbf{Premier}(S) \cup \{\#\} \quad (\text{car } \mathbf{Vide}(S)) \\
\mathbf{Premier}(S) &= \mathbf{Premier}(X) \cup \{a\} \\
\mathbf{Premier}(X) &= \mathbf{Premier}(X) \cup \mathbf{Premier}(Y) \cup \{b\} = \{a, b\} \\
\mathbf{Premier}(Y) &= \{a\}
\end{aligned}$$

et

$$\begin{aligned}
\mathbf{Suivant}(Z) &= \emptyset \\
\mathbf{Suivant}(S) &= \{\#\} \\
\mathbf{Suivant}(X) &= \{a\} \cup \{b\} \cup \mathbf{Suivant}(Y) \quad (1) \\
\mathbf{Suivant}(Y) &= \mathbf{Suivant}(X) \quad (2)
\end{aligned}$$

En considérant les *plus petits ensembles de solutions* vérifiant (1) et (2) on déduit :

$$\mathbf{Suivant}(X) = \mathbf{Suivant}(Y) = \{a, b\}$$

Enfin,

$$\begin{aligned}
\mathbf{Directeurs}(Z \rightarrow S\#) &= \{a, b, \#\} \\
\mathbf{Directeurs}(S \rightarrow Xa) &= \{a, b\} \\
\mathbf{Directeurs}(S \rightarrow \varepsilon) &= \{\#\} \\
\mathbf{Directeurs}(X \rightarrow Xb) &= \{a, b\} \\
\mathbf{Directeurs}(X \rightarrow Y) &= \{a, b\} \\
\mathbf{Directeurs}(Y \rightarrow aX) &= \{a\} \\
\mathbf{Directeurs}(Y \rightarrow \varepsilon) &= \{a, b\}
\end{aligned}$$

La grammaire G_1 n'est donc pas LL(1). Plus généralement, toute grammaire réursive à gauche n'est pas LL(1).

Exemple 2

On considère la grammaire G_2 suivante :

$$\begin{aligned}
Z &\rightarrow S\# \\
S &\rightarrow XaY \\
X &\rightarrow W \\
X &\rightarrow T \\
W &\rightarrow bc \\
T &\rightarrow ac \\
Y &\rightarrow eY \\
Y &\rightarrow f \\
Y &\rightarrow \varepsilon
\end{aligned}$$

On obtient :

$$\begin{aligned}\textbf{Directeurs}(Z \rightarrow S) &= \{a, b\} \\ \textbf{Directeurs}(S \rightarrow XaY) &= \{a, b\} \\ \textbf{Directeurs}(X \rightarrow W) &= \{b\} \\ \textbf{Directeurs}(X \rightarrow T) &= \{a\} \\ \textbf{Directeurs}(W \rightarrow bc) &= \{b\} \\ \textbf{Directeurs}(T \rightarrow ac) &= \{a\} \\ \textbf{Directeurs}(Y \rightarrow eY) &= \{e\} \\ \textbf{Directeurs}(Y \rightarrow f) &= \{f\} \\ \textbf{Directeurs}(Y \rightarrow \varepsilon) &= \{\#\}\end{aligned}$$

G_2 est donc une grammaire LL(1).

Exemple 3

On considère la grammaire G_3 :

$$\begin{aligned}Z &\rightarrow S\# \\ S &\rightarrow Sa \\ S &\rightarrow b\end{aligned}$$

Cette grammaire est récursive à gauche, donc non LL(1). Par élimination de la récursivité à gauche (et factorisation) on peut toutefois la transformer en une grammaire G'_3 qui reconnaît le même langage que G_3 :

$$\begin{aligned}Z &\rightarrow S\# \\ S &\rightarrow bY \\ Y &\rightarrow aT \\ T &\rightarrow aT \\ T &\rightarrow \varepsilon\end{aligned}$$

On a alors :

$$\begin{aligned}\textbf{Directeurs}(Z \rightarrow S\#) &= \{b, \#\} \\ \textbf{Directeurs}(S \rightarrow bY) &= \{b\} \\ \textbf{Directeurs}(Y \rightarrow aT) &= \{a\} \\ \textbf{Directeurs}(T \rightarrow aT) &= \{a\} \\ \textbf{Directeurs}(T \rightarrow \varepsilon) &= \{\#\}\end{aligned}$$

G'_3 est donc une grammaire LL(1).

Remarque : Éliminer la récursivité à gauche ne permet pas toujours d'obtenir une grammaire LL(1) (voir par exemple la grammaire G_1 , où cette grammaire G_3 dans laquelle “ $\#$ ” est remplacé par “ a ”).