Caractérisation d'une grammaire LL(1)

Caractérisation par table d'analyse : une grammaire est LL(1) si chaque case contient exactement une production ou erreur.

Caractérisation « par contre-exemple » : une grammaire n'est pas LL(1) s'il existe 2 productions $X \to \alpha$ et $X \to \beta$ telles que :

1. soit $Premier(\alpha) \cap Premier(\beta) \neq \emptyset$;
   Ex : $S \to aS \,|\, A,\ A \to a$
2. soit $Eps(\alpha) = vrai$ et $Premier(\beta) \cap Suivant(X) \neq \emptyset$;
   Ex : $S \to aS \,|\, Ab,\ A \to \epsilon \,|\, b$
3. soit $Eps(\alpha) = vrai$ et $Eps(\beta) = vrai$ (la grammaire est ambiguë)
   Ex : $S \to A \,|\, B,\ A \to \epsilon,\ B \to \epsilon$

Cas classiques non LL(1)

Dans les cas suivants, la grammaire n'est pas LL(1) :

- ambiguïté ;
- récursivité gauche : $A \rightarrow Aa \mid \epsilon$;
  - intuitivement récursivité infinie de `A()`.
- non factorisation gauche : $S \rightarrow aA \mid aB$

Solutions : :

- factorisation à gauche (parfois) ;
- suppression de la récursivité gauche (parfois) ;
- utiliser un générateur de parser plus puissant !

Factorisation à gauche : exemple - 2

$X \rightarrow ab \mid abbX \mid abbbX$

Factorisation de ab : on prend le plus grand préfixe commun.
$X \rightarrow abY$
$Y \rightarrow \epsilon \mid bX \mid bbX$

Puis à nouveau factorisation de b.

Factorisation à gauche - algorithme

On remplace les règles de la forme :

$$X \to \alpha\beta_1 \mid \ldots \mid \alpha\beta_n \mid \gamma_1 \mid \ldots \mid \gamma_m$$

où

- $\alpha \in (V_T \cup V_N)^+$ et $\beta_i, \gamma_j \in (V_T \cup V_N)*$;
- le préfixe commun α est choisi le plus grand possible ;
- α n'est pas préfixe de γ_j.

par les règles :

$$\begin{array}{c} X \to \alpha X' \mid \gamma_1 \mid \ldots \mid \gamma_m \\ X' \to \beta_1 \mid \ldots \mid \beta_n \end{array}$$

où X' est un nouveau non-terminal.
On réitère ce processus tant que nécessaire.

Suppression de la récursivité à gauche

Récursivité gauche :

- immédiate : production $A \rightarrow A\alpha$, $\alpha \in (V_T \cup V_N)^+$;
- générale : il existe une dérivation $A \Rightarrow^* A\alpha$, $\alpha \in (V_T \cup V_N)^+$.

Il est possible de supprimer les deux cas.

On ne verra que la récursivité immédiate.

Suppression de la récursivité gauche immédiate

On remplace les règles de la forme

$$X \rightarrow X\alpha_1 \,|\, \ldots \,|\, X\alpha_n \,|\, \beta_1 \,|\, \ldots \,|\, \beta_m$$

où

- $\alpha_i \in (V_T \cup V_N)^+$ et $\beta_j \in (V_T \cup V_N)*$;
- les β_j ne commencent pas par X.

par les règles :

$$\begin{array}{c} X \rightarrow \beta_1 X' \,|\, \ldots \,|\, \beta_m X' \\ X' \rightarrow \alpha_1 X' \,|\, \ldots \,|\, \alpha_n X' \,|\, \epsilon \end{array}$$

où X' est un nouveau non-terminal.

Suppression de la récursivité gauche : exemple

Grammaire non ambiguë
des expressions arithmétiques :

$$E \to E + T \mid T$$
$$T \to T * F \mid F$$
$$F \to \mathtt{i} \mid (E)$$

$$\begin{aligned} X \to \quad & X\alpha_1 \mid \ldots \mid X\alpha_n \\ & \mid \beta_1 \mid \ldots \mid \beta_m \end{aligned}$$

Après suppression de la rec gauche :

$$E \to TE'$$
$$E' \to +TE' \mid \epsilon$$
$$T \to FT'$$
$$T' \to *FT' \mid \epsilon$$
$$F \to \mathtt{i} \mid (E)$$

$$X \to \beta_1 X' \mid \ldots \mid \beta_m X'$$
$$X' \to \alpha_1 X' \mid \ldots \mid \alpha_n X' \mid \epsilon$$

Parfois ça ne suffit pas

La grammaire $(\{a, b\}, \{S, A\}, S, P)$ avec

$$P = \{S \to aSb \,|\, A,\ \ A \to aA \,|\, \epsilon\}$$

- n'est pas ambiguë ;
- n'est pas récursive gauche ;
- est factorisée à gauche ;

mais elle n'est pas LL(1).

Au delà des grammaires LL(1) - 1

La grammaire $(\{a, b\}, \{A, B\}, A, P)$ avec

$$P = \{A \rightarrow abB \mid \epsilon,\ \ B \rightarrow Aaa \mid b\}$$

n'est pas LL(1).

En effet $Eps(A) = vrai$ et $a \in Premier(A) \cap Suivant(A)$.