Chapitre 4 : Analyse syntaxique

4.1 Dérivation la plus à gauche et arbre de dérivation

4.1.1 Définition d'une grammaire non contextuelle

Soit une grammaire G (VT, VN, S, P)

VT : ensemble des symboles terminaux de la grammaire (unités lexicales)

VN : ensemble des symboles non terminaux de la grammaire (variables syntaxiques qui dénotent un ensemble de chaînes)

S : l'axiome est un non terminal particulier, tel que l'ensemble des chaînes qu'il dénote est le langage défini par la grammaire.

P : ensemble de règles de production de la grammaire; spécifient la manière dont les terminaux et les non terminaux peuvent être combinés pour former les chaînes.

Exemple

La grammaire constituée des productions suivantes définit des expressions arithmétiques simples :

expr → expr op expr

expr → (expr)

expr → – expr

expr → id

op → +

op → –

op → *

op → /

op →↑

Dans cette grammaire les symboles terminaux sont : id + - * / ↑( ), les symboles non terminaux sont expr, et op et l'axiome est expr

En abrégé:

expr → expr op expr|(expr)|– expr| id

op → +| – | * | / | ↑

4.1.2 Définition d'une Dérivation

Nous disons que $\alpha A\beta \Rightarrow \alpha\gamma\beta$ si $A \rightarrow \gamma$ est une production et α et β sont des chaînes arbitraires de symboles grammaticaux.

Si $\alpha 1 \Rightarrow \alpha 2 \Rightarrow \ldots . \Rightarrow \alpha n$, on dit que $\alpha 1$ se dérive en αn.

Le symbole $\Rightarrow$ signifie "se dérive en une étape"

Pour dire "se dérive en zéro, une ou plusieurs étapes" on peut utiliser le symbole $*\Rightarrow$

Donc

$\alpha *\Rightarrow \alpha$ pour une chaîne quelconque α et

Si $\alpha *\Rightarrow \beta$ et $\beta \Rightarrow \gamma$, alors $\alpha *\Rightarrow \gamma$

$+\Rightarrow$ signifie "se dérive en une ou plusieurs étapes"

Soit une grammaire G et S son axiome; on peut utiliser la relation $+\Rightarrow$ pour définir L(G), le langage engendré par G. On dit qu'une chaîne de terminaux w appartient à L(G) ssi $S +\Rightarrow w$. La chaîne w est appelée phrase de G.

Un langage qui peut être engendré par une grammaire est dit langage non contextuel.

Si $S *\Rightarrow \alpha$, où α peut contenir certains non terminaux, on dit que α est une protophrase de G.

Exemple

Soit la grammaire G (VT, VN, E, P)

$E \rightarrow E+E|E*E|(E)|-E|id$ (3)

La chaîne –(id+id) est une phrase de la grammaire (3) car on a la dérivation:

$E \Rightarrow -E \Rightarrow -(E) \Rightarrow -(E+E) \Rightarrow -(id+E) \Rightarrow -(id+id)$ (4)

Les chaînes E, -E, -(E),…,-(id+id) sont toutes des proto-phrases de cette grammaire.

On écrit $E *\Rightarrow -(id+id)$ pour indiquer que E se dérive en –(id+id)

A chaque étape d'une dérivation, on doit faire deux choix :

- Il faut choisir le non terminal à remplacer
- Quelle alternative utiliser pour ce non terminal

Dérivation gauche : Seul le non terminal le plus à gauche est remplacé à chaque étape. On écrit $\alpha \underset{g}{\Rightarrow} \beta$ on peut alors réécrire (4)

$$E \underset{g}{\Rightarrow} -E \underset{g}{\Rightarrow} -(E) \underset{g}{\Rightarrow} -(E+E) \underset{g}{\Rightarrow} -(id+E) \underset{g}{\Rightarrow} -(id+id) \qquad (4)$$

Chaque étape d'une dérivation gauche peut s'écrire:

$\omega A\gamma \underset{g}{\Rightarrow} \omega\delta\gamma$ ou ω est formé uniquement de terminaux, A → δ est la production utilisée et γ est une chaîne de symboles grammaticaux.

α se dérive en β par une dérivation gauche est notée $\alpha \underset{g}{\Rightarrow} \beta$

Si $S \overset{*}{\underset{g}{\Rightarrow}} \alpha$ on dit que α est une proto-phrase gauche de la grammaire considérée.

Dérivation droite : le non terminal le plus à droite est remplacé à chaque étape.