Caractérisation de PREMIER pour l'analyse $LL(1)$

LEÇONS : 923

RÉFÉRENCES : LEGENDRE, SCHWARZENTRUBER, *Compilation : analyse lexicale et syntaxique* (p.63) [?] et LESESVRE–MONTAGNON–LE BARBENCHON–PIERRON, *131 développements pour l'oral* (p. 783) [?]

Notations :

- $G = (\Sigma, \mathcal{N}, \mathcal{T}, S, \mathcal{R})$ désigne une grammaire où Σ est l'alphabet utilisé, $\mathcal{N}$ est l'ensemble des non-terminaux, $\mathcal{T}$ l'ensemble des terminaux, S l'axiome de G et $\mathcal{R}$ est l'ensemble des règles de production.
- les minuscules seront utilisées pour les terminaux et les majuscules pour les non-terminaux
- $\vec{\alpha}$ désigne une concaténation d'éléments de $\mathcal{N}$ et $\mathcal{T}$
- $\rightarrow^k$ désigne une dérivation de longueur k

Introduction :

On va prouver un théorème qui permet de donner une caractérisation de la fonction PREMIER et grâce à cette caractérisation, on pourra donner un algorithme de saturation qui calcule cette fonction PREMIER. Cette fonction est un outil clé dans l'analyse LL(1), analyse syntaxique descendante durant la compilation d'un programme.

Définition 1. (PREMIER) Pour $\vec{\alpha} \in (\mathcal{N} \cup \mathcal{T})^*$, on pose

$$\textsc{Premier}(\vec{\alpha}) = \begin{cases} \{a, \vec{\alpha} \rightarrow^* a\vec{\beta}\} & \text{si } \vec{\alpha} \not\rightarrow^* \varepsilon \\ \{a, \vec{\alpha} \rightarrow^* a\vec{\beta}\} \cup \{\bowtie\} & \text{si } \vec{\alpha} \rightarrow^* \varepsilon \end{cases}$$

On peut voir PREMIER comme une partie de $(\mathcal{N} \cup \mathcal{T})^* \times (\mathcal{T} \cup \bowtie)$
Si $\textsc{Premier}(\vec{\alpha}) = \{a, +, \bowtie\}$, on écrit alors

$$\textsc{Premier} = \{(\vec{\alpha}, a), (\vec{\alpha}, +), (\vec{\alpha}, \bowtie)\}$$

Théorème 1. La fonction PREMIER est la plus petite partie P de $(\mathcal{N} \cup \mathcal{T})^* \times (\mathcal{T} \cup \bowtie)$ telle que

(i) $a \in P(a)$

(ii) $\bowtie \in P(\varepsilon)$

(iii) si $N \rightarrow \alpha_1\alpha_2 \dots \alpha_n$, alors $P(\alpha_1 \dots \alpha_n) \subset P(N)$

(iv) si $N \rightarrow \varepsilon$, alors $\bowtie \in P(N)$

(v) si $\bowtie \notin P(\alpha_1)$, alors $P(\alpha_1) \subset P(\alpha_1 \dots \alpha_n)$

(vi) si $\bowtie \in P(\alpha_1)$, alors $(P(\alpha_1) \setminus \{\bowtie\}) \cup P(\alpha_2) \subset P(\alpha_1 \dots \alpha_n)$

Démonstration. On procède par double inclusion pour prouver que $P = \textsc{Premier}$
$\boxed{\subset}$ inclusion simple puisque PREMIER vérifie les 6 propriétés du théorème et que P est la plus petite partie qui vérifie ses 6 propriétés.

(i) Comme $a \to^0 a$, on a $a \in \textsc{Premier}(a)$

(ii) Comme $\varepsilon \to^0 \varepsilon$, on est dans le deuxième cas de la définition de Premier et ainsi

$$\bowtie \in \textsc{Premier}(\varepsilon)$$

(iii) Supposons que $N \to \alpha_1\alpha_2 \dots \alpha_n$. Soit $a \in \textsc{Premier}(\alpha_1 \dots \alpha_n)$, ainsi $\alpha_1 \dots \alpha_n \to^* a\vec{\beta}$, ainsi par hypothèse $N \to^* a\vec{\beta}$, ce qui permet de conclure $a \in \textsc{Premier}(N)$

(iv) Supposons que $N \to \varepsilon$. On est dans le deuxième cas de la définition de Premier et ainsi

$$\textsc{Premier}(\varepsilon) = \{\bowtie\} \subset\subset P(N)$$

(v) Supposons que $\bowtie \notin \textsc{Premier}(\alpha_1)$. Si $a \in \textsc{Premier}(\alpha_1)$, alors $\alpha_1 \to^* a\vec{\beta}$. Ainsi

$$\alpha_1\alpha_2 \dots \alpha_n \to^* a\vec{\beta}\alpha_2 \dots \alpha_n$$

Donc $a \in \textsc{Premier}(\alpha_1 \dots \alpha_n)$

(vi) Supposons que $\bowtie \in \textsc{Premier}(\alpha_1)$. Si $a \in \textsc{Premier}(\alpha_1) \setminus \{\bowtie\}$, alors $\alpha_1 \to^* a\vec{\beta}$. Ainsi

$$a \in \textsc{Premier}(\alpha_1 \dots \alpha_n)$$ [1]

Si $a \in \textsc{Premier}(\alpha_2 \dots \alpha_n)$, et comme $\alpha_1 \to^* \varepsilon$ [2], alors

$$\alpha_1\alpha_2 \dots \alpha_n \to^* \varepsilon\alpha_2 \dots \alpha_n$$

Donc $a \in \textsc{Premier}(\alpha_1 \dots \alpha_n)$

$\boxed{\supset}$ On veut montrer que $\textsc{Premier} \subset P$

On va établir ce résultat par récurrence forte sur la longueur de la dérivation

$$HR_n : \begin{array}{ll} \text{“} & \text{si } \vec{\alpha} \to^n a\vec{\beta} \text{ alors } a \in P(\vec{\alpha}) \\ \text{et} & \text{si } \vec{\alpha} \to^n \varepsilon \text{ alors } \bowtie \in P(\vec{\alpha}) \text{”} \end{array}$$

<u>Initialisation :</u>

- Si $\vec{\alpha} \to^0 a\vec{\beta}$, alors $\vec{\alpha} \in \mathcal{T}$ et $\vec{\alpha} = a$. D'où

$$a \in P(\vec{\alpha}) \text{ par (i)}$$

- Si $\vec{\alpha} \to^0 \varepsilon$, alors $\vec{\alpha} = \varepsilon$. D'où

$$\bowtie \in P(\vec{\alpha}) \text{ par (ii)}$$

<u>Hérédité :</u> Soit $n \in \mathbb{N}^*$. Supposons que HR_l soit vraie pour tout $l < n$, montrons HR_n.

- Si $\vec{\alpha} \to^n a\vec{\beta}$, on veut montrer que $a \in P(\vec{\alpha})$
  - ☞ Si le premier caractère de $\vec{\alpha}$ est un terminal, on a forcément $\vec{\alpha}[1] = a$. D'où

$$a \underbrace{\in}_{\text{(i)}} P(a) \underbrace{\subset}_{\text{(v)}} P(\vec{\alpha})$$

  - ☞ Si le premier caractère de $\vec{\alpha}$ est un non-terminal, notons $N := \vec{\alpha}[1]$ et $\vec{\gamma}$ tel que $\vec{\alpha} = N\vec{\gamma}$.

1. pour les mêmes raisons que (v)
2. car $\bowtie \in \textsc{Premier}(\alpha_1)$

✪ Si $N \to^j \varepsilon$, on a $j < n$ [3] et

$$N\vec{\gamma} \to^j \vec{\gamma} \to^{n-j} a\vec{\beta}$$

Donc $\bowtie \in P(N)$ par HR_j et $a \in P(\vec{\gamma})$ par HR_{n-j}. Ainsi

$$a \underbrace{\in}_{HR_{n-j}} P(\vec{\gamma}) \underbrace{\subset}_{\text{(vi)}} P(N\vec{\gamma}) = P(\vec{\alpha})$$

✪ Sinon, on a $N \to x_1 \dots x_k$ avec un des (x_i) qui produit la lettre a. Soit i l'indice minimale tel que $x_i \to^{<n} a$. [4] Ainsi, pour tout $j < i$, $x_j \to^{<n} \varepsilon$. Par $HR_{<n}$, on a $\bowtie \in P(x_j)$ pour tout $j < i$ et $a \in P(x_i)$

$$N\vec{\gamma} \to^{<n} x_i x_{i+1} \dots x_k \vec{\gamma} \to^{<n} a\vec{\beta}$$

$$a \underbrace{\in}_{HR_{<n}} P(x_i) \underbrace{\subset}_{\text{(v) ou (vi)}} P(x_i \dots x_k) \underbrace{\subset}_{\text{(vi)}} P(x_{i-1} \dots x_k) \underbrace{\subset}_{\text{(vi)}} \dots \underbrace{\subset}_{\text{(vi)}} P(x_1 \dots x_k)$$

$$P(x_1 \dots x_k) \underbrace{\subset}_{\text{(iv)}} P(N) \underbrace{\subset}_{\text{(v) ou (vi)}} P(N\vec{\gamma}) = P(\vec{\alpha})$$

- Si $\vec{\alpha} \to^n \varepsilon$, on veut montrer que $\bowtie \in P(\vec{\alpha})$.
  On est forcément dans le cas où $\vec{\alpha}$ commence par un non-terminal [5] que l'on continue d'appeler N et $\vec{\alpha} = N\vec{\gamma}$
  $$N\vec{\gamma} \to x_1 \dots x_k \vec{\gamma} \to^{n-1} \varepsilon$$
  Ainsi $x_1 \dots x_n \to^{<n} \varepsilon$, d'où par $HR_{<n}$, on a
  $$\bowtie \in P(x_1 \dots x_k) \underbrace{\subset}_{\text{(iii)}} P(N)$$
  Or $\vec{\gamma} \to^{<n} \varepsilon$, d'où par $HR_{<n}$, on a
  $$\bowtie \in P(\vec{\gamma}) \underbrace{\subset}_{\text{(vi)}} P(N\vec{\gamma}) = P(\vec{\alpha})$$

Conclusion : On a HR_n pour tout $n \in \mathbb{N}$
Donc ainsi Premier $\subset P$.
Ce qui achève la preuve du théorème

□

Remarques :
Ce théorème permet de donner un algorithme par saturation pour calculer Premier.

Astuces de l'agrégatif :
Je ne fais pas le cas $\vec{\alpha} \to^n \varepsilon$ pendant le développement car il se traite de manière similaire. J'essaye de bien expliquer les cas, en faisant bien référence aux règles à chaque fois que je les utilise.
Je prononce “noeud papillon” pour le symbole $\bowtie$ lors du développement.

3. car il faut que $N\vec{\gamma} \to^n a\vec{\beta}$
4. notation pour dire que la dérivation a strictement moins de n étapes
5. car si on commençait avec un terminal a, on n'engendrerait pas le mot vide à la fin de la dérivation

L'algorithme suivant n'est pas à écrire dans le développement mais est essentiel à comprendre pour savoir à quoi ce théorème peut bien servir. C'est donc cet algorithme qui est utilisé en pratique pour calculer PREMIER.

Algorithme 1 Calcul_PREMIER

```
Procédure Calcul_PREMIER(α⃗)
    Pour a ∈ 𝒯 faire
        PREMIER(a) ← {a}
    Fin Pour
    PREMIER(ε) ← {⋈}
    Pour N → ε ∈ ℛ faire
        PREMIER(N) ← {⋈}
    Fin Pour
    Tant que (PREMIER est modifié) faire
        PREMIER(N) ← PREMIER(N) ∪ ⋃_{N→α1...αn} PREMIER(α1 ... αn)
        Si (⋈ ∉ PREMIER(α1)) alors
            PREMIER(α1 ... αn) ← PREMIER(α1 ... αn) ∪ PREMIER(α1)
        Fin Si
        Si (⋈ ∈ PREMIER(α1)) alors
            PREMIER(α1 ... αn) ← PREMIER(α1 ... αn) ∪ (PREMIER(α1) \ {⋈})
                                  ∪ PREMIER(α2 ... αn)
        Fin Si
    Fait
Fin
```

Questions possibles :

- Calculer la fonction PREMIER pour la grammaire suivante $G = (\Sigma, \mathcal{N}, \mathcal{T}, S, \mathcal{R})$ où $\Sigma = \{a, b, c, d\}$, $\mathcal{N} = \{S, A, B, C\}$, $\mathcal{T} = \Sigma$ et $\mathcal{R}$ est décrit par

$$\begin{aligned} S &\to AB \mid Ca \\ A &\to aAb \mid \varepsilon \\ B &\to bB \mid b \\ C &\to cC \mid d. \end{aligned}$$

- Donner la table d'analyse syntaxique de la grammaire précédente.
- Donner une grammaire qui n'est pas LL(1).
- Calculer la fonction SUIVANT pour la grammaire suivante $G = (\Sigma, \mathcal{N}, \mathcal{T}, S, \mathcal{R})$ où $\Sigma = \{a, b\}$, $\mathcal{N} = \{S\}$, $\mathcal{T} = \Sigma$ et $\mathcal{R}$ est décrit par

$$S \to \varepsilon \mid aSb.$$