Analyse non canonique

2005/2006

1 Rappels

1.1 Analyse LR

Grammaires LR(k)

La classe des grammaires LR(k) est la classe de grammaires la plus large analysable de gauche à droite avec un automate à pile déterministe.

Exercice 1.1. Montrer que la grammaire suivante n'est pas LR(k) pour aucune valeur de k :

$$S \rightarrow BC \,|\, AD,\ A \rightarrow a,\ B \rightarrow a,\ C \rightarrow CA \,|\, A,\ D \rightarrow aD \,|\, b.$$

2 Analyse non canonique

2.1 Principes

Au-delà de LR(k)

- Ne plus utiliser un automate à pile déterministe, mais un automate à deux piles déterministe.
- L'entrée sert de seconde pile.
- Les réductions se font en plaçant le symbole réduit sur la pile d'entrée.
- Les symboles non terminaux peuvent être utilisés en fenêtre.
- Les réductions de syntagmes qui ne sont pas des poignées sont permises.

Exemple 2.1. L'analyse non canonique pour

$$S \rightarrow BC \,|\, AD,\ A \rightarrow a,\ B \rightarrow a,\ C \rightarrow CA \,|\, A,\ D \rightarrow aD \,|\, b$$

va permettre la réduction de b à D dans la forme sententielle aa^nb.

Exemple d'analyse

Exemple 2.2. Soit la grammaire

$$S \rightarrow BC \,|\, AD,\ A \rightarrow a,\ B \rightarrow a,\ C \rightarrow CA \,|\, A,\ D \rightarrow aD \,|\, b.$$

L'analyse non canonique de la phrase aaa se fait par les étapes :

$$
\begin{aligned}
& \$[\varepsilon]\|aaa\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][a]\|aa\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][a][\![aa]\!]\|a\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][a][\![aa]\!][\![aaa]\!]\|\$ \\
\models_{A\to a}\ & \$[\varepsilon][a][\![aa]\!]\|A\$ \\
\models_{A\to a}\ & \$[\varepsilon][a]\|AA\$ \\
\models_{B\to a}\ & \$[\varepsilon]\|BAA\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][B]\|AA\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][B][BA]\|A\$ \\
\models_{C\to A}\ & \$[\varepsilon][B]\|CA\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][B][BC]\|A\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][B][BC][BCA]\|\$ \\
\models_{C\to CA}\ & \$[\varepsilon][B]\|C\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][B][BC]\|\$ \\
\models_{S\to BC}\ & \$[\varepsilon]\|S\$ \\
\models_{\text{shift}}\ & \$[\varepsilon][S]\|\$
\end{aligned}
$$

Les méthodes d'analyse non canonique

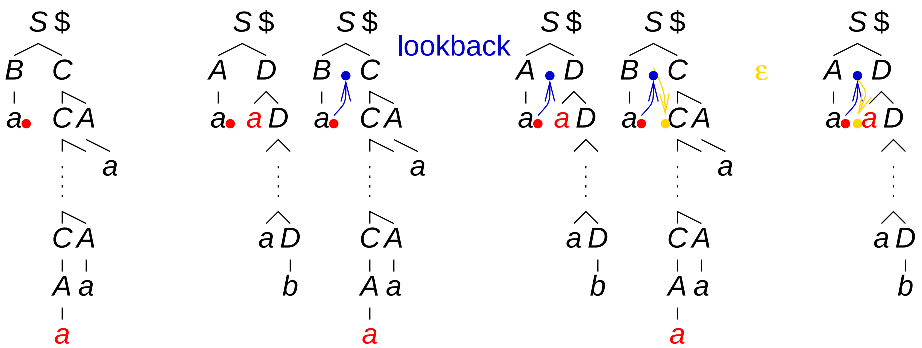
- LR(k,t) permet de réduire le t^{e} syntagme depuis l'extrémité gauche d'une forme sententielle [Knu65] ;
- *Total Precedence* étend l'analyse ascendante par précédence [Col70] ;
- BC(m,n) étend les méthodes à contexte borné [Wil72] ;
- LR(k,∞) et FSPA(k) généralisent LR(k,t) [SW76] ;
- NSLR(1) étend l'analyse SLR(1) [Tai79] ;
- NDR est une extension de DR(k) [FFG04].

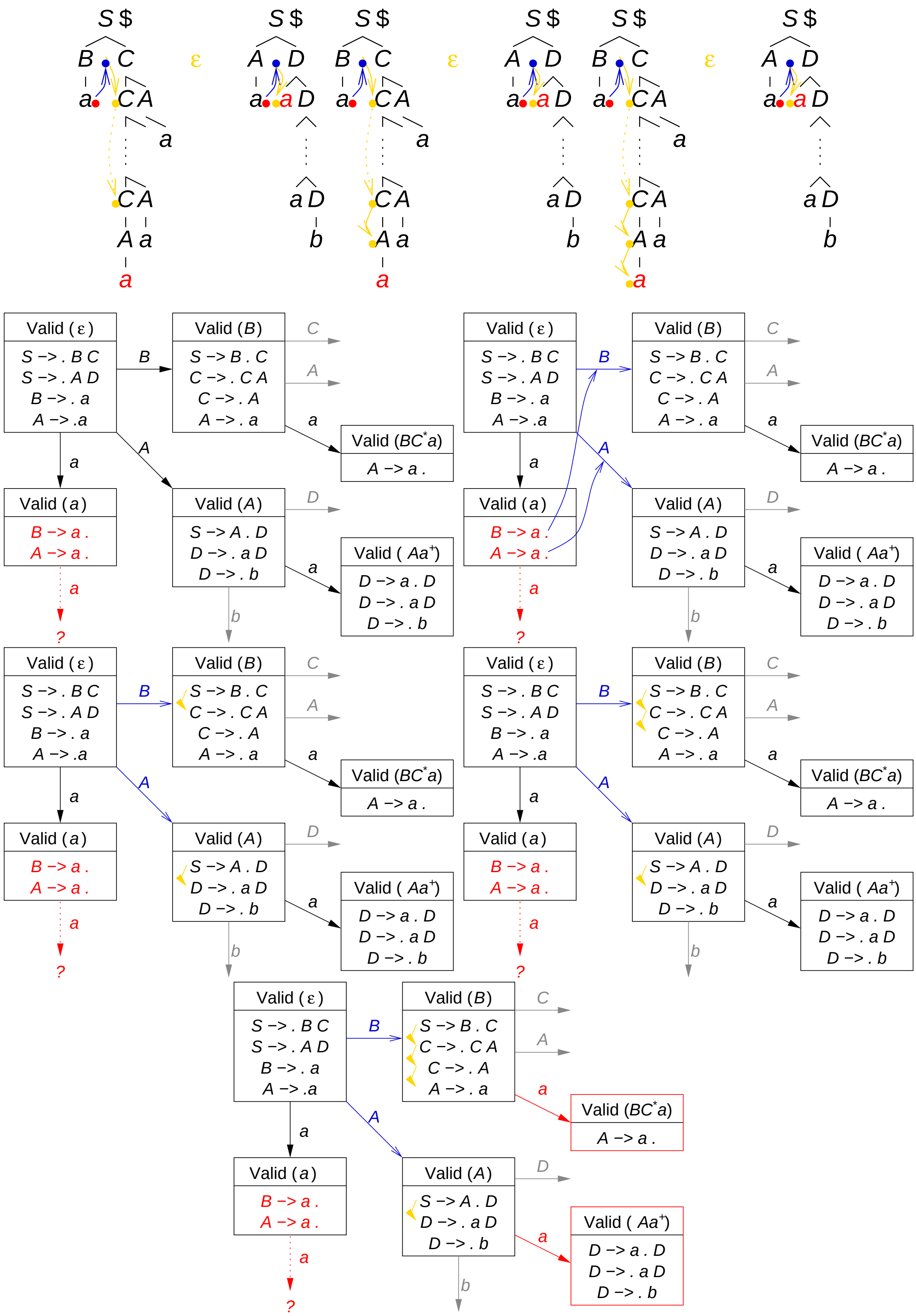
Nous allons présenter une analyse LALR(1) non canonique.

2.2 Analyse NLALR(1)

Calcul des états non canoniques

Comment passer des positions $B\to a\bullet$ et $A\to a\bullet$ de l'état $[a]$ aux positions $A\to a\bullet$ et $D\to a\bullet D$ de l'état $[\![aa]\!]$?





L'état $[\![aa]\!]$ est l'*union* des états LR(0) $[BC^*a]$ et $[Aa^+]$.

Couvertures valides

On dit que les préfixes valides BC^*a et Aa^+ *couvrent* le préfixe de forme sententielle aa, et effectivement $Ba \Rightarrow^* aa$ et $Aa \Rightarrow^* aa$.

Définition 2.3. La chaîne γ est une *couverture valide* dans $\mathcal{G}$ pour la chaîne δ si et seulement si γ est un préfixe valide et $\gamma \Rightarrow^* \delta$.

On écrit $\hat{\delta}$ pour dénoter une couverture de δ et Cover(δ) pour dénoter l'ensemble de toutes les couvertures valides pour δ.

Théorème 2.4. *Soit α un syntagme tel que $S' \Rightarrow^* \delta A X \omega \Rightarrow \delta \alpha X \omega$ avec $q = [\hat{\delta}\alpha]$, alors*

$$\mathit{Cover}(\hat{\delta} A X) = \{\gamma C X \mid (q, A \rightarrow \alpha) \ \textbf{\textit{lookback}} \ \circ \ \textbf{\textit{includes}}^* \circ \ \textbf{\textit{reads}}^*([\gamma], C)\}.$$

Fenêtres NLALR(1)

Définition 2.5. L'ensemble des *fenêtres réduites* pour une réduction par $A \rightarrow \alpha$ dans l'état LR(0) q est défini par

$$\text{RLA}(q, A \rightarrow \alpha) = \{X \mid S' \underset{\text{lm}}{\Rightarrow^*} z A \gamma X \omega, \gamma \Rightarrow^* \varepsilon, X \Rightarrow^* a x, \text{ et } q = [\hat{z}\alpha]\}. \tag{1}$$

Définition 2.6. L'ensemble des *fenêtres dérivées* pour une réduction par $A \rightarrow \alpha$ dans l'état LR(0) q est défini par

$$\text{DLA}(q, A \rightarrow \alpha) = \{X \mid S' \Rightarrow^* \delta A X \omega, X \Rightarrow^* a x, \text{ et } q = [\hat{\delta}\alpha]\}. \tag{2}$$

Définition 2.7. L'ensemble des *fenêtres en conflit* pour une réduction par $A \rightarrow \alpha$ dans un ensemble d'états LR(0) s est défini par

$$\begin{aligned}\text{CLA}(s, A \rightarrow \alpha) = \{X \in \text{DLA}(q, A \rightarrow \alpha) \mid q \in s, ((q, X) \text{ ou } \exists p \in s,\\ X \in \text{DLA}(p, B \rightarrow \beta)) \text{ et } X \notin \text{RLA}(q, A \rightarrow \alpha)\}.\end{aligned}$$

L'ensemble de *fenêtres non canoniques* pour une réduction par $A \rightarrow \alpha$ dans un ensemble d'états LR(0) s est défini par

$$\text{NLA}(s, A \rightarrow \alpha) = \big(\bigcup_{q \in s} \text{DLA}(q, A \rightarrow \alpha)\big) - \text{CLA}(s, A \rightarrow \alpha). \tag{3}$$

Théorème 2.8.

$$\mathit{RLA}(q, A \rightarrow \alpha) = \{X \mid X \Rightarrow^* a x, \psi \Rightarrow^* \varepsilon, C \Rightarrow \rho B \bullet \psi X \sigma \in \mathit{Kernel}(\delta \rho B) \ \mathit{et}$$

$$(q, A \rightarrow \alpha) \ \textbf{\textit{lookback}} \ \circ \ \textbf{\textit{includes}}^*([\delta\rho], B)\}.$$

Soit

$$\mathit{DR}([\delta], A) = \{X \mid ([\delta A], X) \ \mathit{et} \ X \Rightarrow^* a x\},$$

alors

$$\mathit{DLA}(q, A \rightarrow \alpha) = \bigcup_{(q, A \rightarrow \alpha) \ \textbf{\textit{lookback}} \ \circ \ \textbf{\textit{includes}}^* \circ \ \textbf{\textit{reads}}^* (r, C)} \mathit{DR}(r, C).$$

Exemple

Exemple 2.9. Les fenêtres non canoniques pour la réduction par $A \rightarrow a$ dans l'état non canonique $[\![aa]\!] = \{[BC^*a], [Aa^+]\}$ sont

$$\begin{aligned}\text{RLA}([BC^*a], A \rightarrow a) &= \{A, \$\},\\ \text{DLA}([BC^*a], A \rightarrow a) &= \{A, a, \$\},\\ \text{CLA}([\![aa]\!], A \rightarrow a) &= \{a\}, \text{ et}\\ \text{NLA}([\![aa]\!], A \rightarrow a) &= \{A, \$\}.\end{aligned}$$

État non canonique NLALR(1)

Définition 2.10. L'*état non canonique* $[\![\delta]\!]$ est l'ensemble d'états LR(0) défini par

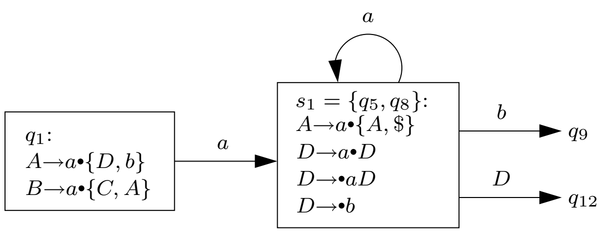
$$[\![\varepsilon]\!] = \{[\varepsilon]\} \text{ et} \tag{4}$$

$$[\![\delta X]\!] = \{[\widehat{\hat{\gamma} A X}] \mid X \in \mathrm{CLA}([\![\delta]\!], A \to \alpha) \text{ et } [\hat{\gamma}\alpha] \in [\![\delta]\!]\} \cup \{[\varphi X] \mid [\varphi] \in [\![\delta]\!]\}. \tag{5}$$

La *transition non canonique* depuis $[\![\delta]\!]$ vers $[\![\delta X]\!]$ sur le symbole X, dénotée par $([\![\delta]\!], X)$, existe si et seulement si $[\![\delta X]\!] \neq \emptyset$.

La réduction $([\![\delta]\!], A \to \alpha)$ existe si et seulement si il existe une réduction $(q, A \to \alpha)$ et si l'état LR(0) q est un élément de $[\![\delta]\!]$.

Exemple


q_1:
$A \to a \bullet \{D, b\}$
$B \to a \bullet \{C, A\}$
a
a
$s_1 = \{q_5, q_8\}$:
$A \to a \bullet \{A, \$\}$
$D \to a \bullet D$
$D \to \bullet aD$
$D \to \bullet b$
b
q_9
D
q_{12}

Automate NLALR(1)

Définition 2.11. Un *automate NLALR(1)* pour la grammaire $\mathcal{G}$ a pour configuration initiale $[\![\varepsilon]\!]\|w\$$ avec w la chaîne d'entrée de Σ^*, pour configuration finale $[\![\varepsilon]\!][\![S]\!]\|\$$, et des règles de réécriture de la forme

- *shift* X dans l'état $[\![\delta]\!]$

$$[\![\delta]\!]\|X \vdash_{\text{shift}} [\![\delta]\!][\![\delta X]\!]\|, \tag{6}$$

défini si il existe une transition $([\![\delta]\!], X)$, ou

- *reduce* par $A{\rightarrow}X_1 \dots X_n$ dans l'état $[\![\delta X_1 \dots X_n]\!]$ avec la fenêtre X

$$[\![\delta X_1]\!] \dots [\![\delta X_1 \dots X_n]\!]\|X \vdash_{A\rightarrow X_1 \dots X_n} \|AX, \tag{7}$$

défini si $A{\rightarrow}X_1 \dots X_n$ est une réduction de $[\![\delta X_1 \dots X_n]\!]$ et si la fenêtre X appartient à $\mathrm{NLA}([\![\delta X_1 \dots X_n]\!], A{\rightarrow}X_1 \dots X_n)$.

3 Conclusion

En bref

- Des méthodes d'analyse déterministes.
- Des méthodes plus puissantes que leurs pendants canoniques.
- Analyse en temps linéaire si la fenêtre a une longueur bornée.
- Peuvent être étendues pour utiliser une fenêtre de longueur non bornée à la LRR !

References

[Col70] Alain Colmerauer. Total precedence relations. *Journal of the ACM*, 17(1):14–30, January 1970.

[FFG04] Jacques Farré and José Fortes Gálvez. Bounded-connect noncanonical discriminating-reverse parsers. *Theoretical Computer Science*, 313(1):73–91, February 2004.

[Knu65] Donald E. Knuth. On the translation of languages from left to right. *Information and Control*, 8:607–639, 1965.

[SW76] Thomas G. Szymanski and John H. Williams. Noncanonical extensions of bottom-up parsing techniques. *SIAM Journal of Computing*, 5(2):231–250, June 1976.

[Tai79] Kuo-Chung Tai. Noncanonical SLR(1) grammars. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 1(2):295–320, 1979.

[Wil72] John H. Williams. Bounded context parsable grammars. Technical Report 72-127, Department of Computer Science, Cornell University, Ithaca, New York, April 1972.