Compiler Construction

Lecture 10: Syntax Analysis VI ($LR(1)$ and $LALR(1)$ Parsing)

Summer Semester 2016

Thomas Noll
Software Modeling and Verification Group
RWTH Aachen University

https://moves.rwth-aachen.de/teaching/ss-16/cc/

Software Modeling and Verification Chair | RWTH AACHEN UNIVERSITY

$LR(0)$ Items and Sets

Definition ($LR(0)$ items and sets)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$ be start separated by $S' \to S$ and $S' \Rightarrow_r^* \alpha A w \Rightarrow_r \alpha\beta_1\beta_2 w$ (i.e., $A \to \beta_1\beta_2 \in P$).

- $[A \to \beta_1 \cdot \beta_2]$ is called an $LR(0)$ item for $\alpha\beta_1$.
- Given $\gamma \in X^*$, $LR(0)(\gamma)$ denotes the set of all $LR(0)$ items for γ, called the $LR(0)$ set (or: $LR(0)$ information) of γ.
- $LR(0)(G) := \{LR(0)(\gamma) \mid \gamma \in X^*\}$.

Corollary

1. *For every $\gamma \in X^*$, $LR(0)(\gamma)$ is finite.*
2. *$LR(0)(G)$ is finite.*
3. *The item $[A \to \beta\cdot] \in LR(0)(\gamma)$ indicates the possible reduction $(w, \alpha\beta, z) \vdash (w, \alpha A, zi)$ where $\pi_i = A \to \beta$ and $\gamma = \alpha\beta$.*
4. *The item $[A \to \beta_1 \cdot Y\beta_2] \in LR(0)(\gamma)$ indicates an incomplete handle β_1 (to be completed by shift operations or ε-reductions).*

$LR(0)$ Conflicts

Definition ($LR(0)$ conflicts)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$ and $I \in LR(0)(G)$.

- I has a shift/reduce conflict if there exist $A \to \alpha_1 a \alpha_2, B \to \beta \in P$ such that
$$[A \to \alpha_1 \cdot a\alpha_2], [B \to \beta \cdot] \in I.$$
- I has a reduce/reduce conflict if there exist $A \to \alpha, B \to \beta \in P$ with $A \neq B$ or $\alpha \neq \beta$ such that
$$[A \to \alpha \cdot], [B \to \beta \cdot] \in I.$$

Lemma

$G \in LR(0)$ iff no $I \in LR(0)(G)$ contains conflicting items.

Proof.

omitted □

The goto Function

Observation: if $G \in LR(0)$, then $LR(0)(\gamma)$ yields deterministic shift/reduce decision for $NBA(G)$ in a configuration with pushdown γ
$\implies$ new pushdown alphabet: $LR(0)(G)$ in place of X

Moreover $LR(0)(\gamma Y)$ is determined by $LR(0)(\gamma)$ and Y but independent from γ in the following sense:

$$LR(0)(\gamma) = LR(0)(\gamma') \implies LR(0)(\gamma Y) = LR(0)(\gamma' Y)$$

Definition ($LR(0)$ goto function)

The function $\mathrm{goto} : LR(0)(G) \times X \to LR(0)(G)$ is determined by

$$\mathrm{goto}(I, Y) = I' \quad \text{iff} \quad \text{there exists } \gamma \in X^* \text{ such that } I = LR(0)(\gamma) \text{ and } I' = LR(0)(\gamma Y).$$

The $LR(0)$ Action Function

The parsing automaton will be defined using another table, the action function, which determines the shift/reduce decision (reminder: $\pi_0 = S' \to S$).

Definition ($LR(0)$ action function)

The $LR(0)$ action function $\mathrm{act} : LR(0)(G) \to \{\text{red } i \mid i \in [p]\} \cup \{\text{shift}, \text{accept}, \text{error}\}$ is defined by

$$\mathrm{act}(I) := \begin{cases} \text{red } i & \text{if } i \neq 0, \pi_i = A \to \alpha \text{ and } [A \to \alpha \cdot] \in I \\ \text{shift} & \text{if } [A \to \alpha_1 \cdot a\alpha_2] \in I \\ \text{accept} & \text{if } [S' \to S \cdot] \in I \\ \text{error} & \text{if } I = \emptyset \end{cases}$$

Corollary

For every $G \in CFG_\Sigma$, $G \in LR(0)$ iff act is well defined.

Together, act and goto form the $LR(0)$ parsing table of G.

The $LR(0)$ Parsing Automaton

Definition ($LR(0)$ parsing automaton)

Let $G = \langle N, \Sigma, P, S\rangle \in LR(0)$. The (deterministic) $LR(0)$ parsing automaton of G is defined by the following components:

- Input alphabet Σ
- Pushdown alphabet $\Gamma := LR(0)(G)$
- Output alphabet $\Delta := [p] \cup \{0, \text{error}\}$
- Configurations $\Sigma^* \times \Gamma^* \times \Delta^*$
- Initial configuration (w, I_0, ε) where $I_0 := LR(0)(\varepsilon)$
- Final configurations $\{\varepsilon\} \times \{\varepsilon\} \times \Delta^*$
- Transitions:

$$\begin{aligned}
\text{shift: } & (aw, \alpha I, z) \vdash (w, \alpha I I', z) \text{ if } \mathrm{act}(I) = \text{shift and } \mathrm{goto}(I, a) = I' \\
\text{reduce: } & (w, \alpha I I_1 \ldots I_n, z) \vdash (w, \alpha I I', zi) \text{ if } \mathrm{act}(I_n) = \text{red } i,\ \pi_i = A \to Y_1 \ldots Y_n,\ \mathrm{goto}(I, A) = I' \\
\text{accept: } & (\varepsilon, I_0 I, z) \vdash (\varepsilon, \varepsilon, z\,0) \text{ if } \mathrm{act}(I) = \text{accept} \\
\text{error: } & (w, \alpha I, z) \vdash (\varepsilon, \varepsilon, z\,\text{error}) \text{ if } \mathrm{act}(I) = \text{error}
\end{aligned}$$

The $SLR(1)$ Action Function

Definition ($SLR(1)$ action function)

The $SLR(1)$ action function

$$\mathrm{act} : LR(0)(G) \times \Sigma_\varepsilon \to \{\text{red } i \mid i \in [p]\} \cup \{\text{shift}, \text{accept}, \text{error}\}$$

is defined by

$$\mathrm{act}(I, x) := \begin{cases} \text{red } i & \text{if } i \neq 0, \pi_i = A \to \alpha, [A \to \alpha\cdot] \in I, \\ & \text{and } x \in \mathrm{fo}(A) \\ \text{shift} & \text{if } [A \to \alpha_1 \cdot x\alpha_2] \in I \text{ and } x \in \Sigma \\ \text{accept} & \text{if } [S' \to S\cdot] \in I \text{ and } x = \varepsilon \\ \text{error} & \text{otherwise} \end{cases}$$

Definition ($SLR(1)$ grammar)

A grammar $G \in CFG_\Sigma$ has the $SLR(1)$ property (notation: $G \in SLR(1)$) if its $SLR(1)$ action function is well defined.

act and the $LR(0)$ goto function (Definition 9.3) form the $SLR(1)$ parsing table of G.

The $SLR(1)$ Parsing Automaton

Definition ($SLR(1)$ parsing automaton)

The $SLR(1)$ parsing automaton is defined as in the $LR(0)$ case (see Definition 9.8), except for the transition relation:

shift: $(aw, \alpha I, z) \vdash (w, \alpha I I', z)$ if $\mathrm{act}(I, a) = \text{shift}$ and $\mathrm{goto}(I, a) = I'$

reduce$_a$: $(aw, \alpha I I_1 \ldots I_n, z) \vdash (aw, \alpha I I', zi)$ if $\mathrm{act}(I_n, a) = \text{red } i$, $\pi_i = A \to Y_1 \ldots Y_n$, and $\mathrm{goto}(I, A) = I'$

reduce$_\varepsilon$: $(\varepsilon, \alpha I I_1 \ldots I_n, z) \vdash (\varepsilon, \alpha I I', zi)$ if $\mathrm{act}(I_n, \varepsilon) = \text{red } i$, $\pi_i = A \to Y_1 \ldots Y_n$, and $\mathrm{goto}(I, A) = I'$

accept: $(\varepsilon, I_0 I, z) \vdash (\varepsilon, \varepsilon, z\,0)$ if $\mathrm{act}(I, \varepsilon) = \text{accept}$

error$_a$: $(aw, \alpha I, z) \vdash (\varepsilon, \varepsilon, z\,\text{error})$ if $\mathrm{act}(I, a) = \text{error}$

error$_\varepsilon$: $(\varepsilon, \alpha I, z) \vdash (\varepsilon, \varepsilon, z\,\text{error})$ if $\mathrm{act}(I, \varepsilon) = \text{error}$

$SLR(1)$ Conflicts

Problem: not all conflicts can be resolved using fo sets

Example 10.1

$G_{LR}: S' \to S \quad S \to L{=}R \mid R \quad L \to *R \mid \mathrm{a} \quad R \to L$

$LR(0)(G_{LR})$:

$I_0 := LR(0)(\varepsilon)$: $[S' \to \cdot S]$ $[S \to \cdot L{=}R]$ $[S \to \cdot R]$ $[L \to \cdot *R]$ $[L \to \cdot \mathrm{a}]$ $[R \to \cdot L]$

$I_1 := LR(0)(S)$: $[S' \to S\cdot]$

$I_2 := LR(0)(L)$: $[S \to L \cdot {=}R]$ $[R \to L\cdot]$

$I_3 := LR(0)(R)$: $[S \to R\cdot]$

$I_4 := LR(0)(*)$: $[L \to * \cdot R]$ $[R \to \cdot L]$ $[L \to \cdot *R]$ $[L \to \cdot \mathrm{a}]$

$I_5 := LR(0)(\mathrm{a})$: $[L \to \mathrm{a}\cdot]$

$I_6 := LR(0)(L{=})$: $[S \to L{=} \cdot R]$ $[R \to \cdot L]$ $[L \to \cdot *R]$ $[L \to \cdot \mathrm{a}]$

$I_7 := LR(0)(*R)$: $[L \to *R\cdot]$

$I_8 := LR(0)(*L)$: $[R \to L\cdot]$

$I_9 := LR(0)(L{=}R)$: $[S \to L{=}R\cdot]$

But: conflict in I_2 not $SLR(1)$-solvable since $= \in \mathrm{fo}(R)$

$LR(1)$ Items and Sets I

Observation: not every element of $\mathrm{fo}(A)$ can follow every occurrence of A
$\Longrightarrow$ refinement of $LR(0)$ items by adding possible lookahead symbols

Definition 10.2 ($LR(1)$ items and sets)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$ be start separated by $S' \to S$.

- If $S' \Rightarrow_r^* \alpha Aaw \Rightarrow_r \alpha\beta_1\beta_2 aw$, then $[A \to \beta_1 \cdot \beta_2, a]$ is called an $LR(1)$ item for $\alpha\beta_1$.
- If $S' \Rightarrow_r^* \alpha A \Rightarrow_r \alpha\beta_1\beta_2$, then $[A \to \beta_1 \cdot \beta_2, \varepsilon]$ is called an $LR(1)$ item for $\alpha\beta_1$.
- Given $\gamma \in X^*$, $LR(1)(\gamma)$ denotes the set of all $LR(1)$ items for γ, called the $LR(1)$ set (or: $LR(1)$ information) of γ.
- $LR(1)(G) := \{LR(1)(\gamma) \mid \gamma \in X^*\}$.

$LR(1)$ Items and Sets II

Corollary 10.3

1. *For every* $\gamma \in X^*$, $LR(1)(\gamma)$ *is finite.*
2. $LR(1)(G)$ *is finite.*
3. *For every* $\gamma \in X^*$, $LR(1)(\gamma)$ *"contains"* $LR(0)(\gamma)$, *i.e.,*

$$\{[A \to \beta_1 \cdot \beta_2] \mid [A \to \beta_1 \cdot \beta_2, x] \in LR(1)(\gamma)\} = LR(0)(\gamma).$$

4. $[A \to \beta_1 \cdot \beta_2, x] \in I \in LR(1)(G) \implies x \in \mathrm{fo}(A)$

$LR(1)$ Conflicts

Definition 10.4 ($LR(1)$ conflicts)

Let $G = \langle N, \Sigma, P, S \rangle \in CFG_\Sigma$ and $I \in LR(1)(G)$.

- I has a shift/reduce conflict if there exist $A \to \alpha_1 a \alpha_2, B \to \beta \in P$ and $x \in \Sigma_\varepsilon$ such that

$$[A \to \alpha_1 \cdot a\alpha_2, x], [B \to \beta\cdot, a] \in I.$$

- I has a reduce/reduce conflict if there exist $x \in \Sigma_\varepsilon$ and $A \to \alpha, B \to \beta \in P$ with $A \neq B$ or $\alpha \neq \beta$ such that

$$[A \to \alpha\cdot, x], [B \to \beta\cdot, x] \in I.$$

Lemma 10.5

$G \in LR(1)$ iff no $I \in LR(1)(G)$ contains conflicting items.

Computing $LR(1)$ Sets I

The computation of $LR(0)$ sets (cf. Theorem 8.15) can be extended to cover right contexts:

Theorem 10.6 (Computing $LR(1)$ sets)

Let $G = \langle N, \Sigma, P, S\rangle \in CFG_\Sigma$ be start separated by $S' \to S$ and reduced.

1. $LR(1)(\varepsilon)$ *is the least set such that*
   - $[S' \to \cdot S, \varepsilon] \in LR(1)(\varepsilon)$ *and*
   - *if* $[A \to \cdot B\gamma, x] \in LR(1)(\varepsilon)$, $B \to \beta \in P$, *and* $y \in \mathrm{fi}(\gamma x)$, *then* $[B \to \cdot\beta, y] \in LR(1)(\varepsilon)$.
2. $LR(1)(\alpha Y)$ $(\alpha \in X^*, Y \in X)$ *is the least set such that*
   - *if* $[A \to \gamma_1 \cdot Y\gamma_2, x] \in LR(1)(\alpha)$, *then* $[A \to \gamma_1 Y \cdot \gamma_2, x] \in LR(1)(\alpha Y)$ *and*
   - *if* $[A \to \gamma_1 \cdot B\gamma_2, x] \in LR(1)(\alpha Y)$, $B \to \beta \in P$, *and* $y \in \mathrm{fi}(\gamma_2 x)$, *then* $[B \to \cdot\beta, y] \in LR(1)(\alpha Y)$.

Computing $LR(1)$ Sets II

Example 10.7 (cf. Example 10.1)

$LR(1)(G_{LR})$ for $G_{LR}: S' \to S \quad S \to L{=}R \mid R \quad L \to *R \mid \mathrm{a} \quad R \to L$

$[S' \to \cdot S, \varepsilon] \in LR(1)(\varepsilon)$ $[A \to \cdot B\gamma, x] \in LR(1)(\varepsilon), B \to \beta \in P, y \in \mathrm{fi}(\gamma x) \implies [B \to \cdot\beta, y] \in LR(1)(\varepsilon)$ $[A \to \gamma_1 \cdot Y\gamma_2$

$I'_0 := LR(1)(\varepsilon)$: $[S' \to \cdot S, \varepsilon]$ $[S \to \cdot L{=}R, \varepsilon]$ $[S \to \cdot R, \varepsilon]$ $[L \to \cdot *R, =]$ $[L \to \cdot \mathrm{a}, =]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_1 := LR(1)(S)$: $[S' \to S\cdot, \varepsilon]$

$I'_2 := LR(1)(L)$: $[S \to L \cdot {=}R, \varepsilon]$ $[R \to L\cdot, \varepsilon]$

$I'_3 := LR(1)(R)$: $[S \to R\cdot, \varepsilon]$

$I'_4 := LR(1)(*)$: $[L \to * \cdot R, =]$ $[L \to * \cdot R, \varepsilon]$ $[R \to \cdot L, =]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, =]$ $[L \to \cdot \mathrm{a}, =]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_5 := LR(1)(\mathrm{a})$: $[L \to \mathrm{a}\cdot, =]$ $[L \to \mathrm{a}\cdot, \varepsilon]$

$I'_6 := LR(1)(L{=})$: $[S \to L{=} \cdot R, \varepsilon]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_7 := LR(1)(*R)$: $[L \to *R\cdot, =]$ $[L \to *R\cdot, \varepsilon]$

$I'_8 := LR(1)(*L)$: $[R \to L\cdot, =]$ $[R \to L\cdot, \varepsilon]$

$I'_9 := LR(1)(L{=}R)$: $[S \to L{=}R\cdot, \varepsilon]$

$I'_{10} := LR(1)(L{=}L)$: $[R \to L\cdot, \varepsilon]$

$I'_{11} := LR(1)(L{=}*)$: $[L \to * \cdot R, \varepsilon]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_{12} := LR(1)(L{=}\mathrm{a})$: $[L \to \mathrm{a}\cdot, \varepsilon]$

$I'_{13} := LR(1)(L{=}*R)$: $[L \to *R\cdot, \varepsilon]$

$I'_{14} := \emptyset$

In I'_2: shift on =/reduce on $\varepsilon \implies G_{LR} \in LR(1)$

The $LR(1)$ Action Function

Definition 10.8 ($LR(1)$ action function)

The $LR(1)$ action function

$$\mathrm{act} : LR(1)(G) \times \Sigma_\varepsilon \to \{\text{red } i \mid i \in [p]\} \cup \{\text{shift}, \text{accept}, \text{error}\}$$

is defined by

$$\mathrm{act}(I, x) := \begin{cases} \text{red } i & \text{if } i \neq 0, \pi_i = A \to \alpha \text{ and } [A \to \alpha\cdot, x] \in I \\ \text{shift} & \text{if } [A \to \alpha_1 \cdot x\alpha_2, y] \in I \text{ and } x \in \Sigma \\ \text{accept} & \text{if } [S' \to S\cdot, \varepsilon] \in I \text{ and } x = \varepsilon \\ \text{error} & \text{otherwise} \end{cases}$$

Corollary 10.9

For every $G \in CFG_\Sigma$, $G \in LR(1)$ iff its $LR(1)$ action function is well defined.

The $LR(1)$ goto Function

The goto function is defined in analogy to the $LR(0)$ case (Definition 9.3).

Definition 10.10 ($LR(1)$ goto function)

The function $\mathrm{goto} : LR(1)(G) \times X \to LR(1)(G)$ is determined by

$$\mathrm{goto}(I, Y) = I' \quad \text{iff} \quad \text{there exists } \gamma \in X^* \text{ such that } I = LR(1)(\gamma) \text{ and } I' = LR(1)(\gamma Y).$$

Again, act and goto form the $LR(1)$ parsing table of G.

The $LR(1)$ Parsing Table

Example 10.11 (cf. Example 10.7)

$LR(1)(G_{LR})$	$\mathrm{act/goto}\|_\Sigma$				$\mathrm{goto}\|_N$		
	$*$	$=$	a	ε	S	L	R
I'_0	shift$/I'_4$		shift$/I'_5$		I'_1	I'_2	I'_3
I'_1				accept			
I'_2		shift$/I'_6$		red 5			
I'_3				red 2			
I'_4	shift$/I'_4$		shift$/I'_5$			I'_8	I'_7
I'_5		red 4		red 4			
I'_6	shift$/I'_{11}$		shift$/I'_{12}$			I'_{10}	I'_9
I'_7		red 3		red 3			
I'_8		red 5					
I'_9				red 1			
I'_{10}				red 5			
I'_{11}	shift$/I'_{11}$		shift$/I'_{12}$			I'_{10}	I'_{13}
I'_{12}				red 4			
I'_{13}				red 3			

(empty = error$/\emptyset$)

The $LR(1)$ Parsing Automaton I

Definition 10.12 ($LR(1)$ parsing automaton)

The $LR(1)$ parsing automaton is defined as in the $LR(0)$ case (see Definition 9.8), except for the transition relation:

shift: $(aw, \alpha I, z) \vdash (w, \alpha I I', z)$ if $\mathrm{act}(I, a) = \text{shift}$ and $\mathrm{goto}(I, a) = I'$

reduce$_a$: $(aw, \alpha I I_1 \ldots I_n, z) \vdash (aw, \alpha I I', zi)$ if $\mathrm{act}(I_n, a) = \text{red } i$, $\pi_i = A \to Y_1 \ldots Y_n$, and $\mathrm{goto}(I, A) = I'$

reduce$_\varepsilon$: $(\varepsilon, \alpha I I_1 \ldots I_n, z) \vdash (\varepsilon, \alpha I I', zi)$ if $\mathrm{act}(I_n, \varepsilon) = \text{red } i$, $\pi_i = A \to Y_1 \ldots Y_n$, and $\mathrm{goto}(I, A) = I'$

accept: $(\varepsilon, I_0 I, z) \vdash (\varepsilon, \varepsilon, z\,0)$ if $\mathrm{act}(I, \varepsilon) = \text{accept}$

error$_a$: $(aw, \alpha I, z) \vdash (\varepsilon, \varepsilon, z\,\text{error})$ if $\mathrm{act}(I, a) = \text{error}$

error$_\varepsilon$: $(\varepsilon, \alpha I, z) \vdash (\varepsilon, \varepsilon, z\,\text{error})$ if $\mathrm{act}(I, \varepsilon) = \text{error}$

The $LR(1)$ Parsing Automaton II

Example 10.13 (cf. Example 10.7)

$G_{LR}: S' \to S\ (0) \quad S \to L{=}R \mid R\ (1,2) \quad L \to {*}R \mid \mathtt{a}\ (3,4) \quad R \to L\ (5)$

$LR(1)(G_{LR})$	act/goto$\vert_\Sigma$ $*$	$=$	a	ε	goto$\vert_N$ S	L	R
I'_0	shift/I'_4		shift/I'_5		I'_1	I'_2	I'_3
I'_1				accept			
I'_2		shift/I'_6		red 5			
I'_3				red 2			
I'_4	shift/I'_4		shift/I'_5			I'_8	I'_7
I'_5		red 4		red 4			
I'_6	shift/I'_{11}		shift/I'_{12}			I'_{10}	I'_9
I'_7		red 3		red 3			
I'_8		red 5					
I'_9				red 1			
I'_{10}				red 5			
I'_{11}	shift/I'_{11}		shift/I'_{12}			I'_{10}	I'_{13}
I'_{12}				red 4			
I'_{13}				red 3			

(empty = error/$\emptyset$)

$LR(1)$ parsing of a=*a:

$$
\begin{aligned}
& (\mathtt{a{=}{*}a},\ I'_0,\ \varepsilon) \\
\vdash\ & (\mathtt{{=}{*}a},\ I'_0 I'_5,\ \varepsilon) \\
\vdash\ & (\mathtt{{=}{*}a},\ I'_0 I'_2,\ 4) \\
\vdash\ & (\mathtt{{*}a},\ I'_0 I'_2 I'_6,\ 4) \\
\vdash\ & (\mathtt{a},\ I'_0 I'_2 I'_6 I'_{11},\ 4) \\
\vdash\ & (\varepsilon,\ I'_0 I'_2 I'_6 I'_{11} I'_{12},\ 4) \\
\vdash\ & (\varepsilon,\ I'_0 I'_2 I'_6 I'_{11} I'_{10},\ 44) \\
\vdash\ & (\varepsilon,\ I'_0 I'_2 I'_6 I'_{11} I'_{13},\ 445) \\
\vdash\ & (\varepsilon,\ I'_0 I'_2 I'_6 I'_{10},\ 4453) \\
\vdash\ & (\varepsilon,\ I'_0 I'_2 I'_6 I'_9,\ 44535) \\
\vdash\ & (\varepsilon,\ I'_0 I'_1,\ 445351) \\
\vdash\ & (\varepsilon,\ \varepsilon,\ 4453510)
\end{aligned}
$$

$LALR(1)$ Parsing

- **Motivation:** resolving conflicts using $LR(1)$ too expensive
- Example 10.1/10.7: $|LR(0)(G_{LR})| = 11$, $|LR(1)(G_{LR})| = 15$
- Empirical evaluations:
  - A. Johnstone, E. Scott: *Generalised Reduction Modified LR Parsing for Domain Specific Language Prototyping*, HICSS '02, IEEE, 2002
  - X. Chen, D. Pager: *Full LR(1) Parser Generator Hyacc and Study on the Performance of LR(1) Algorithms*, C3S2E '11, ACM, 2011

Grammar	$\lvert LR(0)(G)\rvert$	$\lvert LR(1)(G)\rvert$
Pascal	368	1395
Ansi-C	381	1788
C++	1236	9723

$LR(0)$ Equivalence I

Observation: potential redundancy by containment of $LR(0)$ sets in $LR(1)$ sets (cf. Corollary 10.3)

Definition 10.14 ($LR(0)$ equivalence)

Let $\mathrm{lr}_0 : LR(1)(G) \to LR(0)(G)$ be defined by

$$\mathrm{lr}_0(I) := \{[A \to \beta_1 \cdot \beta_2] \mid [A \to \beta_1 \cdot \beta_2, x] \in I\}.$$

Two sets $I_1, I_2 \in LR(1)(G)$ are called $LR(0)$-equivalent (notation: $I_1 \sim_0 I_2$) if $\mathrm{lr}_0(I_1) = \mathrm{lr}_0(I_2)$.

$LR(0)$ Equivalence II

Example 10.15 (cf. Example 10.1/10.7)

G_{LR} : $S' \to S \qquad S \to L{=}R \mid R$
$\qquad L \to *R \mid \mathrm{a} \qquad R \to L$

$LR(0)(G_{LR})$:

$I_0(\varepsilon)$: $[S' \to \cdot S]$ $[S \to \cdot L{=}R]$ $[S \to \cdot R]$ $[L \to \cdot *R]$ $[L \to \cdot \mathrm{a}]$ $[R \to \cdot L]$

$I_1(S)$: $[S' \to S\cdot]$

$I_2(L)$: $[S \to L \cdot {=}R]$ $[R \to L\cdot]$

$I_3(R)$: $[S \to R\cdot]$

$I_4(*)$: $[L \to * \cdot R]$ $[R \to \cdot L]$ $[L \to \cdot *R]$ $[L \to \cdot \mathrm{a}]$

$I_5(\mathrm{a})$: $[L \to \mathrm{a}\cdot]$

$I_6(L{=})$: $[S \to L{=} \cdot R]$ $[R \to \cdot L]$ $[L \to \cdot *R]$ $[L \to \cdot \mathrm{a}]$

$I_7(*R)$: $[L \to *R\cdot]$

$I_8(*L)$: $[R \to L\cdot]$

$I_9(L{=}R)$: $[S \to L{=}R\cdot]$

$\Longrightarrow \quad I'_4 \sim_0 I'_{11} \quad I'_5 \sim_0 I'_{12}$
$\qquad I'_7 \sim_0 I'_{13} \quad I'_8 \sim_0 I'_{10}$

$LR(1)(G_{LR})$:

$I'_0(\varepsilon)$: $[S' \to \cdot S, \varepsilon]$ $[S \to \cdot L{=}R, \varepsilon]$ $[S \to \cdot R, \varepsilon]$ $[L \to \cdot *R, =]$ $[L \to \cdot \mathrm{a}, =]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_1(S)$: $[S' \to S\cdot, \varepsilon]$

$I'_2(L)$: $[S \to L \cdot {=}R, \varepsilon]$ $[R \to L\cdot, \varepsilon]$

$I'_3(R)$: $[S \to R\cdot, \varepsilon]$

$I'_4(*)$: $[L \to * \cdot R, =]$ $[L \to * \cdot R, \varepsilon]$ $[R \to \cdot L, =]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, =]$ $[L \to \cdot \mathrm{a}, =]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_5(\mathrm{a})$: $[L \to \mathrm{a}\cdot, =]$ $[L \to \mathrm{a}\cdot, \varepsilon]$

$I'_6(L{=})$: $[S \to L{=} \cdot R, \varepsilon]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_7(*R)$: $[L \to *R\cdot, =]$ $[L \to *R\cdot, \varepsilon]$

$I'_8(*L)$: $[R \to L\cdot, =]$ $[R \to L\cdot, \varepsilon]$

$I'_9(L{=}R)$: $[S \to L{=}R\cdot, \varepsilon]$

$I'_{10}(L{=}L)$: $[R \to L\cdot, \varepsilon]$

$I'_{11}(L{=}*)$: $[L \to * \cdot R, \varepsilon]$ $[R \to \cdot L, \varepsilon]$ $[L \to \cdot *R, \varepsilon]$ $[L \to \cdot \mathrm{a}, \varepsilon]$

$I'_{12}(L{=}\mathrm{a})$: $[L \to \mathrm{a}\cdot, \varepsilon]$

$I'_{13}(L{=}*R)$: $[L \to *R\cdot, \varepsilon]$

$LALR(1)$ Sets I

Corollary 10.16

For every $G \in CFG_\Sigma$, $|LR(1)(G)/\sim_0| = |LR(0)(G)|$.

Idea: merge $LR(0)$-equivalent $LR(1)$ sets
(maintaining the lookahead information, but possibly introducing conflicts)

Definition 10.17 ($LALR(1)$ sets)

Let $G \in CFG_\Sigma$.

- An information $I \in LR(1)(G)$ determines the $LALR(1)$ set

$$\bigcup [I]_{\sim_0} = \bigcup \{I' \in LR(1)(G) \mid I' \sim_0 I\}.$$

- The set of all $LALR(1)$ sets of G is denoted by $LALR(1)(G)$.

Remark: by Corollary 10.16, $|LALR(1)(G)| = |LR(0)(G)|$
(but $LALR(1)$ sets provide additional lookahead information)

$LALR(1)$ Sets II

Example 10.18 (cf. Example 10.15)

$G_{LR}: S' \to S \quad S \to L{=}R \mid R \quad L \to *R \mid \mathrm{a} \quad R \to L$

$LR(0)(G_{LR})$:			$LALR(1)(G_{LR})$:		
$I_0(\varepsilon)$:	$[S' \to \cdot S]$	$[S \to \cdot L{=}R]$	$I''_0 := I'_0$:	$[S' \to \cdot S, \varepsilon]$	$[S \to \cdot L{=}R, \varepsilon]$
	$[S \to \cdot R]$	$[L \to \cdot * R]$		$[S \to \cdot R, \varepsilon]$	$[L \to \cdot * R, {=}/\varepsilon]$
	$[L \to \cdot \mathrm{a}]$	$[R \to \cdot L]$		$[L \to \cdot \mathrm{a}, {=}/\varepsilon]$	$[R \to \cdot L, \varepsilon]$
$I_1(S)$:	$[S' \to S \cdot]$		$I''_1 := I'_1$:	$[S' \to S \cdot, \varepsilon]$	
$I_2(L)$:	$[S \to L \cdot {=}R]$	$[R \to L \cdot]$	$I''_2 := I'_2$:	$[S \to L \cdot {=}R, \varepsilon]$	$[R \to L \cdot, \varepsilon]$
$I_3(R)$:	$[S \to R \cdot]$		$I''_3 := I'_3$:	$[S \to R \cdot, \varepsilon]$	
$I_4(*)$:	$[L \to * \cdot R]$	$[R \to \cdot L]$	$I''_4 := I'_4 \cup I'_{11}$:	$[L \to * \cdot R, {=}/\varepsilon]$	$[R \to \cdot L, {=}/\varepsilon]$
	$[L \to \cdot * R]$	$[L \to \cdot \mathrm{a}]$		$[L \to \cdot * R, {=}/\varepsilon]$	$[L \to \cdot \mathrm{a}, {=}/\varepsilon]$
$I_5(\mathrm{a})$:	$[L \to \mathrm{a} \cdot]$		$I''_5 := I'_5 \cup I'_{12}$:	$[L \to \mathrm{a} \cdot, {=}/\varepsilon]$	
$I_6(L{=})$:	$[S \to L{=} \cdot R]$	$[R \to \cdot L]$	$I''_6 := I'_6$:	$[S \to L{=} \cdot R, \varepsilon]$	$[R \to \cdot L, \varepsilon]$
	$[L \to \cdot * R]$	$[L \to \cdot \mathrm{a}]$		$[L \to \cdot * R, \varepsilon]$	$[L \to \cdot \mathrm{a}, \varepsilon]$
$I_7(*R)$:	$[L \to * R \cdot]$		$I''_7 := I'_7 \cup I'_{13}$:	$[L \to * R \cdot, {=}/\varepsilon]$	
$I_8(*L)$:	$[R \to L \cdot]$		$I''_8 := I'_8 \cup I'_{10}$:	$[R \to L \cdot, {=}/\varepsilon]$	
$I_9(L{=}R)$:	$[S \to L{=}R \cdot]$		$I''_9 := I'_9$:	$[S \to L{=}R \cdot, \varepsilon]$	

The $LALR(1)$ Action Function

The $LALR(1)$ action function is defined in analogy to the $LR(1)$ case (Definition 10.8).

Definition 10.19 ($LALR(1)$ action function)

The $LALR(1)$ action function

$$\mathrm{act} : LALR(1)(G) \times \Sigma_\varepsilon \to \{\text{red } i \mid i \in [p]\} \cup \{\text{shift}, \text{accept}, \text{error}\}$$

is defined by

$$\mathrm{act}(I, x) := \begin{cases} \text{red } i & \text{if } i \neq 0, \pi(i) = A \to \alpha \text{ and } [A \to \alpha\cdot, x] \in I \\ \text{shift} & \text{if } [A \to \alpha_1 \cdot x\alpha_2, y] \in I \text{ and } x \in \Sigma \\ \text{accept} & \text{if } [S' \to S\cdot, \varepsilon] \in I \text{ and } x = \varepsilon \\ \text{error} & \text{otherwise} \end{cases}$$

Definition 10.20 ($LALR(1)$ grammar)

A grammar $G \in CFG_\Sigma$ has the $LALR(1)$ property (notation: $G \in LALR(1)$) if its $LALR(1)$ action function is well defined.

The $LALR(1)$ goto Function

Example 10.21 (cf. Example 10.18)

$G_{LR} \in LALR(1)$

Also the $LR(1)$ goto function (Definition 10.10) carries over to the $LALR(1)$ case.
Reason:

Lemma 10.22

Let $G \in CFG_\Sigma$ and $I_1, I_2 \in LR(1)(G)$ such that $I_1 \sim_0 I_2$. Then, for every $Y \in X$, $\mathrm{goto}(I_1, Y) \sim_0 \mathrm{goto}(I_2, Y)$.

Again, act and goto form the $LALR(1)$ parsing table of G.

The $LALR(1)$ Parsing Table

Example 10.23 (cf. Example 10.18)

$LALR(1)(G_{LR})$	$\mathrm{act}/\mathrm{goto}\vert_\Sigma$				$\mathrm{goto}\vert_N$		
	$*$	$=$	a	ε	S	L	R
I''_0	shift$/I''_4$		shift$/I''_5$		I''_1	I''_2	I''_3
I''_1				accept			
I''_2		shift$/I''_6$		red 5			
I''_3				red 2			
I''_4	shift$/I''_4$		shift$/I''_5$			I''_8	I''_7
I''_5		red 4		red 4			
I''_6	shift$/I''_4$		shift$/I''_5$			I''_8	I''_9
I''_7		red 3		red 3			
I''_8		red 5		red 5			
I''_9				red 1			

(empty = error$/\emptyset$)

LALR(1) Conflicts

But: merging of *LR*(1) sets can produce new conflicts (also see exercises):

Example 10.24

$G: S' \to S \quad S \to \mathrm{a}A\mathrm{d} \mid \mathrm{b}B\mathrm{d} \mid \mathrm{a}B\mathrm{e} \mid \mathrm{b}A\mathrm{e} \quad A \to \mathrm{c} \quad B \to \mathrm{c}$

$LR(1)(\varepsilon)$:	$[S' \to \cdot S, \varepsilon]$	$[S \to \cdot\mathrm{a}A\mathrm{d}, \varepsilon]$	$[S \to \cdot\mathrm{b}B\mathrm{d}, \varepsilon]$ $[S \to \cdot\mathrm{a}B\mathrm{e}, \varepsilon]$	$[S \to \cdot\mathrm{b}A\mathrm{e}, \varepsilon]$
$LR(1)(S)$:	$[S' \to S\cdot, \varepsilon]$			
$LR(1)(\mathrm{a})$:	$[S \to \mathrm{a} \cdot A\mathrm{d}, \varepsilon]$	$[S \to \mathrm{a} \cdot B\mathrm{e}, \varepsilon]$	$[A \to \cdot\mathrm{c}, \mathrm{d}]$	$[B \to \cdot\mathrm{c}, \mathrm{e}]$
$LR(1)(\mathrm{b})$:	$[S \to \mathrm{b} \cdot B\mathrm{d}, \varepsilon]$	$[S \to \mathrm{b} \cdot A\mathrm{e}, \varepsilon]$	$[B \to \cdot\mathrm{c}, \mathrm{d}]$	$[A \to \cdot\mathrm{c}, \mathrm{e}]$
$LR(1)(\mathrm{a}A)$:	$[S \to \mathrm{a}A \cdot \mathrm{d}, \varepsilon]$		$LR(1)(\mathrm{a}B)$:	$[S \to \mathrm{a}B \cdot \mathrm{e}, \varepsilon]$
$LR(1)(\mathrm{ac})$:	$[A \to \mathrm{c}\cdot, \mathrm{d}]$	$[B \to \mathrm{c}\cdot, \mathrm{e}]$		
$LR(1)(\mathrm{b}B)$:	$[S \to \mathrm{b}B \cdot \mathrm{d}, \varepsilon]$		$LR(1)(\mathrm{b}A)$:	$[S \to \mathrm{b}A \cdot \mathrm{e}, \varepsilon]$
$LR(1)(\mathrm{bc})$:	$[B \to \mathrm{c}\cdot, \mathrm{d}]$	$[A \to \mathrm{c}\cdot, \mathrm{e}]$		
$LR(1)(\mathrm{a}A\mathrm{d})$:	$[S \to \mathrm{a}A\mathrm{d}\cdot, \varepsilon]$		$LR(1)(\mathrm{a}B\mathrm{e})$:	$[S \to \mathrm{a}B\mathrm{e}\cdot, \varepsilon]$
$LR(1)(\mathrm{b}B\mathrm{d})$:	$[S \to \mathrm{b}B\mathrm{d}\cdot, \varepsilon]$		$LR(1)(\mathrm{b}A\mathrm{e})$:	$[S \to \mathrm{b}A\mathrm{e}\cdot, \varepsilon]$

No conflicts $\Longrightarrow$ $G \in LR(1)$

$LR(1)(\mathrm{ac}) \sim_0 LR(1)(\mathrm{bc})$, but $LR(1)(\mathrm{ac}) \cup LR(1)(\mathrm{bc})$ has conflicts $\Longrightarrow$ $G \notin LALR(1)$