# Compiler Construction

## Lecture 10: Syntax Analysis VI ($LR(0)$ Parsing)

Thomas Noll

Lehrstuhl für Informatik 2
(Software Modeling and Verification)

RWTH Aachen University

noll@cs.rwth-aachen.de

http://www-i2.informatik.rwth-aachen.de/i2/cc12/

Summer Semester 2012

# Outline

1. Repetition: $LR(0)$ Grammars

2. $LR(0)$ Parsing

# $LR(0)$ Grammars

The case $k = 0$ is relevant (in contrast to $LL(0)$): here the decision is just based on the contents of the pushdown, without any lookahead.

## Corollary ($LR(0)$ grammar)

*$G \in CFG_\Sigma$ has the $LR(0)$ property if for all rightmost derivations of the form*

$$S \begin{cases} \Rightarrow_r^* \alpha A w \Rightarrow_r \alpha\beta w \\ \Rightarrow_r^* \gamma B x \Rightarrow_r \alpha\beta y \end{cases}$$

*it follows that $\alpha = \gamma$, $A = B$, and $x = y$.*

**Goal:** derive a finite information from the pushdown which suffices to resolve the nondeterminism (similar to abstraction of right context in LL parsing by $\mathrm{fo}$-sets)

# $LR(0)$ 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 $\gamma$, called the $LR(0)$ set (or: $LR(0)$ information) of $\gamma$.
- $LR(0)(G) := \{LR(0)(\gamma) \mid \gamma \in X^*\}$.

## 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 $\beta_1$ (to be completed by shift operations or $\varepsilon$-steps).*

# $LR(0)$ 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 $\square$

# Computing $LR(0)$ Sets I

## Theorem (Computing $LR(0)$ sets)

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

1. $LR(0)(\varepsilon)$ *is the least set such that*
   - $[S' \to \cdot S] \in LR(0)(\varepsilon)$ *and*
   - *if* $[A \to \cdot B\gamma] \in LR(0)(\varepsilon)$ *and* $B \to \beta \in P$,
     *then* $[B \to \cdot\beta] \in LR(0)(\varepsilon)$.
2. $LR(0)(\alpha Y)$ $(\alpha \in X^*, Y \in X)$ *is the least set such that*
   - *if* $[A \to \gamma_1 \cdot Y\gamma_2] \in LR(0)(\alpha)$,
     *then* $[A \to \gamma_1 Y \cdot \gamma_2] \in LR(0)(\alpha Y)$ *and*
   - *if* $[A \to \gamma_1 \cdot B\gamma_2] \in LR(0)(\alpha Y)$ *and* $B \to \beta \in P$,
     *then* $[B \to \cdot\beta] \in LR(0)(\alpha Y)$.

# Computing $LR(0)$ Sets II

## Example (cf. Example 9.6)

$$
\begin{aligned}
G:\quad S' &\to S \\
S &\to B \mid C \\
B &\to aB \mid b \\
C &\to aC \mid c
\end{aligned}
\qquad [S' \to \cdot S] \in
$$

$$
LR(0)(\varepsilon) \quad \frac{[A \to \cdot B\gamma] \in LR(0)(\varepsilon),\, B \to \beta \in P}{\Longrightarrow\ [B \to \cdot\beta] \in LR(0)(\varepsilon)} \qquad \frac{[A \to \gamma_1 \cdot Y\gamma_2] \in LR(0)(\alpha)}{\Longrightarrow\ [A \to \gamma_1 Y \cdot \gamma_2] \in LR(0)(\alpha Y)}
$$

| | | | | |
|---|---|---|---|---|
| $I_0 := LR(0)(\varepsilon)$ : | $[S' \to \cdot S]$ | $[S \to \cdot B]$ | $[S \to \cdot C]$ | $[B \to \cdot aB]$ |
| | $[B \to \cdot b]$ | $[C \to \cdot aC]$ | $[C \to \cdot c]$ | |
| $I_1 := LR(0)(S)$ : | $[S' \to S \cdot]$ | | | |
| $I_2 := LR(0)(B)$ : | $[S \to B \cdot]$ | | | |
| $I_3 := LR(0)(C)$ : | $[S \to C \cdot]$ | | | |
| $I_4 := LR(0)(a)$ : | $[B \to a \cdot B]$ | $[C \to a \cdot C]$ | $[B \to \cdot aB]$ | $[B \to \cdot b]$ |
| | $[C \to \cdot aC]$ | $[C \to \cdot c]$ | | |
| $I_5 := LR(0)(b)$ : | $[B \to b \cdot]$ | | | |
| $I_6 := LR(0)(c)$ : | $[C \to c \cdot]$ | | | |
| $I_7 := LR(0)(aB)$ : | $[B \to aB \cdot]$ | | | |
| $I_8 := LR(0)(aC)$ : | $[C \to aC \cdot]$ | | | |

($LR(0)(aa) = LR(0)(a) = I_4$, $LR(0)(ab) = LR(0)(b) = I_5$,
$LR(0)(ac) = LR(0)(c) = I_6$, $I_9 := LR(0)(\gamma) = \emptyset$ in all remaining cases)

# Reduce/Reduce Conflicts

## Example

$$
\begin{array}{lrcl}
G: & S' & \to & S \\
 & S & \to & Aa \mid Bb \\
 & A & \to & a \\
 & B & \to & a
\end{array}
$$

$$
\begin{array}{lllllll}
LR(0)(\varepsilon): & [S' \to \cdot S] & [S \to \cdot Aa] & [S \to \cdot Bb] & [A \to \cdot a] & [B \to \cdot a] \\
LR(0)(S): & [S' \to S \cdot] \\
LR(0)(A): & [S \to A \cdot a] \\
LR(0)(B): & [S \to B \cdot b] \\
LR(0)(a): & [A \to a \cdot] & [B \to a \cdot] \\
LR(0)(Aa): & [S \to Aa \cdot] \\
LR(0)(Bb): & [S \to Bb \cdot]
\end{array}
$$

**Note:** $G$ is unambiguous

# Shift/Reduce Conflicts

## Example

$$
\begin{aligned}
G: \quad S' &\to S \\
S &\to aS \mid a
\end{aligned}
$$

$$
\begin{array}{lllll}
LR(0)(\varepsilon): & [S' \to \cdot S] & [S \to \cdot aS] & [S \to \cdot a] & \\
LR(0)(S): & [S' \to S\cdot] & & & \\
LR(0)(a): & [S \to a \cdot S] & [S \to \cdot aS] & \color{red}{[S \to \cdot a]} & \color{red}{[S \to a\cdot]} \\
LR(0)(aS): & [S \to aS\cdot] & & &
\end{array}
$$

**Note:** $G$ is unambiguous

# Outline


1. Repetition: $LR(0)$ Grammars

2. $LR(0)$ Parsing

# The goto Function I

**Observation:** if $G \in LR(0)$, then $LR(0)(\gamma)$ yields deterministic shift/reduce decision for $\mathrm{NBA}(G)$ in a configuration with pushdown $\gamma$
$\implies$ 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 $\gamma$ in the following sense:

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

**Definition 10.1 ($LR(0)$ goto function)**

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

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

## Example 10.2 (cf. Example 9.12)

$I_0 := LR(0)(\varepsilon) : \quad [S' \to \cdot S]$
$[S \to \cdot B] \quad [S \to \cdot C]$
$[B \to \cdot aB] \quad [B \to \cdot b]$
$[C \to \cdot aC] \quad [C \to \cdot c]$

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

$I_2 := LR(0)(B) : \quad [S \to B \cdot]$

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

$I_4 := LR(0)(a) : \quad [B \to a \cdot B] \quad [C \to a \cdot C]$
$[B \to \cdot aB] \quad [B \to \cdot b]$
$[C \to \cdot aC] \quad [C \to \cdot c]$

$I_5 := LR(0)(b) : \quad [B \to b \cdot]$

$I_6 := LR(0)(c) : \quad [C \to c \cdot]$

$I_7 := LR(0)(aB) : \quad [B \to aB \cdot]$

$I_8 := LR(0)(aC) : \quad [C \to aC \cdot]$

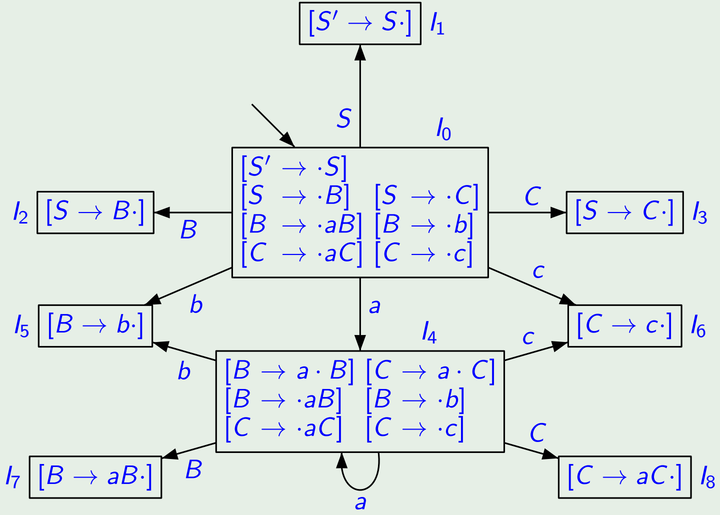
$I_9 := \emptyset$

| goto | $S$ | $B$ | $C$ | $a$ | $b$ | $c$ |
|---|---|---|---|---|---|---|
| $I_0$ | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ |
| $I_1$ | | | | | | |
| $I_2$ | | | | | | |
| $I_3$ | | | | | | |
| $I_4$ | | $I_7$ | $I_8$ | $I_4$ | $I_5$ | $I_6$ |
| $I_5$ | | | | | | |
| $I_6$ | | | | | | |
| $I_7$ | | | | | | |
| $I_8$ | | | | | | |
| $I_9$ | | | | | | |

(empty $= I_9$)

## Example 10.2 (continued)

Representation of goto funtion as finite automaton:



# 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 10.3 ($LR(0)$ action function)

The $LR(0)$ action function

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

is defined by

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

## Corollary 10.4

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

# The $LR(0)$ Parsing Table

## Example 10.5 (cf. Example 10.2)

$$
\begin{array}{lllll}
G: & S' & \to & S & (0) \\
 & S & \to & B \mid C & (1,2) \\
 & B & \to & aB \mid b & (3,4) \\
 & C & \to & aC \mid c & (5,6)
\end{array}
$$

| $LR(0)(G)$ | act | goto | | | | | |
|---|---|---|---|---|---|---|---|
| | | $S$ | $B$ | $C$ | $a$ | $b$ | $c$ |
| $I_0$ | shift | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ |
| $I_1$ | accept | | | | | | |
| $I_2$ | red 1 | | | | | | |
| $I_3$ | red 2 | | | | | | |
| $I_4$ | shift | | $I_7$ | $I_8$ | $I_4$ | $I_5$ | $I_6$ |
| $I_5$ | red 4 | | | | | | |
| $I_6$ | red 6 | | | | | | |
| $I_7$ | red 3 | | | | | | |
| $I_8$ | red 5 | | | | | | |
| $I_9$ | error | | | | | | |

(empty $= I_9$)

$$
\begin{array}{lll}
I_0 := LR(0)(\varepsilon): & [S' \to \cdot S] & \\
 & [S \to \cdot B] & [S \to \cdot C] \\
 & [B \to \cdot aB] & [B \to \cdot b] \\
 & [C \to \cdot aC] & [C \to \cdot c] \\
I_1 := LR(0)(S): & [S' \to S \cdot] & \\
I_2 := LR(0)(B): & [S \to B \cdot] & \\
I_3 := LR(0)(C): & [S \to C \cdot] & \\
I_4 := LR(0)(a): & [B \to a \cdot B] & [C \to a \cdot C] \\
 & [B \to \cdot aB] & [B \to \cdot b] \\
 & [C \to \cdot aC] & [C \to \cdot c] \\
I_5 := LR(0)(b): & [B \to b \cdot] & \\
I_6 := LR(0)(c): & [C \to c \cdot] & \\
I_7 := LR(0)(aB): & [B \to aB \cdot] & \\
I_8 := LR(0)(aC): & [C \to aC \cdot] & \\
I_9 := \emptyset & &
\end{array}
$$

# The $LR(0)$ Parsing Automaton I

## Definition 10.6 ($LR(0)$ 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 $\Sigma$
- Pushdown alphabet $\Gamma := LR(0)(G)$
- Output alphabet $\Delta := [p] \cup \{0, \text{error}\}$
- Configurations $\Sigma^* \times \Gamma^* \times \Delta^*$
- Initial configuration $(w, I_0, \varepsilon)$ where $I_0 := LR(0)(\varepsilon)$
- Final configurations $\{\varepsilon\} \times \{\varepsilon\} \times \Delta^*$
- Transitions:

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

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

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

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

# The $LR(0)$ Parsing Automaton II

## Example 10.7 (cf. Example 10.5)

$$
\begin{array}{llll}
G: & S' & \to S & (0) \\
 & S & \to B \mid C & (1,2) \\
 & B & \to aB \mid b & (3,4) \\
 & C & \to aC \mid c & (5,6)
\end{array}
$$

| $LR(0)(G)$ | act | goto | | | | | |
|---|---|---|---|---|---|---|---|
| | | $S$ | $B$ | $C$ | $a$ | $b$ | $c$ |
| $I_0$ | shift | $I_1$ | $I_2$ | $I_3$ | $I_4$ | $I_5$ | $I_6$ |
| $I_1$ | accept | | | | | | |
| $I_2$ | red 1 | | | | | | |
| $I_3$ | red 2 | | | | | | |
| $I_4$ | shift | | $I_7$ | $I_8$ | $I_4$ | $I_5$ | $I_6$ |
| $I_5$ | red 4 | | | | | | |
| $I_6$ | red 6 | | | | | | |
| $I_7$ | red 3 | | | | | | |
| $I_8$ | red 5 | | | | | | |
| $I_9$ | error | | | | | | |

(empty $= I_9$)

$LR(0)$ parsing of $aac$:

$$
\begin{array}{ll}
 & (aac, I_0, \varepsilon) \\
\vdash & (ac, I_0 I_4, \varepsilon) \\
\vdash & (c, I_0 I_4 I_4, \varepsilon) \\
\vdash & (\varepsilon, I_0 I_4 I_4 I_6, \varepsilon) \\
\vdash & (\varepsilon, I_0 I_4 I_4 I_8, 6) \\
(*) & \\
\vdash & (\varepsilon, I_0 I_4 I_8, 65) \\
\vdash & (\varepsilon, I_0 I_3, 655) \\
\vdash & (\varepsilon, I_0 I_1, 6552) \\
\vdash & (\varepsilon, \varepsilon, 65520)
\end{array}
$$

Check by rightmost derivation (on the board)

**Remark:** in the corresponding computation of $\mathrm{NBA}(G)$, $(*)$ is nondeterministic

**Theorem 10.8 (Correctness of $LR(0)$ Parsing Automaton)**

*If $G \in LR(0)$, then the $LR(0)$ parsing automaton of $G$ is deterministic, and for every $w \in \Sigma^*$ and $z \in \{0, \ldots, p\}^*$:*

$$(w, I_0, \varepsilon) \vdash^* (\varepsilon, \varepsilon, z) \quad \textit{iff} \quad \overleftarrow{z} \textit{ is a rightmost analysis of } w$$

**Proof.**

omitted $\square$