LL(1) Parsing Theory

Goal: Formal, rigorous description of those grammars for which "I can figure out how to do a top-down parse by looking ahead just one token", plus corresponding algorithms.

Notation:

T = Set of Terminals (Tokens)

N = Set of Nonterminals

FIRST(α) - First Token From

Definition: For any $\alpha \in (N \cup T)^*$,

FIRST(α) =

$\{ a \in T \mid \alpha \Rightarrow^* a \beta \text{ for some } \beta \in (N \cup T)^* \} \cup$

$\{\varepsilon, \text{ if } \alpha \Rightarrow^* \varepsilon\}$

Computing First

FIRST(ε) = $\{\varepsilon\}$

For all $a \in T$, FIRST(a) = {a}

For all $A \in N$, repeat until no change

If there is a rule $A \to \varepsilon$, add(ε) to FIRST(A)

For all rules $A \to Y_1 \ldots Y_k$ add(FIRST(Y_1) - $\{\varepsilon\}$)

if $\varepsilon \in$ FIRST(Y_1) then add(FIRST(Y_2) - $\{\varepsilon\}$)

if $\varepsilon \in$ FIRST(Y_1Y_2) then add(FIRST(Y_3) - $\{\varepsilon\}$)

…

if $\varepsilon \in$ FIRST($Y_1 Y_2 ... Y_k$) then add(ε)

Computing FIRST (Cont.)

For all any string $Y_1 \ldots Y_k \in (N \cup T)^*$, similar:

add(FIRST(Y_1) - $\{\varepsilon\}$)

if $\varepsilon \in$ FIRST(Y_1) then add(FIRST(Y_2) - $\{\varepsilon\}$)

if $\varepsilon \in$ FIRST(Y_1Y_2) then add(FIRST(Y_3) - $\{\varepsilon\}$)

…

if $\varepsilon \in$ FIRST($Y_1 Y_2 ... Y_k$) then add(ε)

FOLLOW(B) - Next Token After

Definition: For any $B \in N$,

FOLLOW(B) =

$\{ a \in (T \cup \{\$\}) \mid S\$ \Rightarrow^* \alpha\, B\, a\, \beta$

for some $\alpha, \beta \in (N \cup T \cup \{\$\})^* \}$

Computing FOLLOW(B)

Add \$ to FOLLOW(S)

Repeat until no change

For all rules $A \to \alpha B \beta$ [i.e. all rules with a B in r.h.s],

Add (FIRST(β) - $\{\varepsilon\}$) to FOLLOW(B)

If $\varepsilon \in$ FIRST(β) [e.g. if $\beta = \varepsilon$] then

Add FOLLOW(A) to FOLLOW(B)

Assume for all A that $S \Rightarrow^* \alpha A \beta$ for some $\alpha,\beta \in (N \cup T)^*$, else A irrelevant

PREDICT - Given lhs, which rhs?

For all rules $A \rightarrow \alpha$

For all $a \in$ FIRST(α) - {ε}

Add($A \rightarrow \alpha$) to PREDICT(A,a)

If $\varepsilon \in$ FIRST(α) then

For all $b \in$ FOLLOW(A)

Add($A \rightarrow \alpha$) to PREDICT(A,b)

Defn: G is LL(1) iff every cell has $\leq$ 1 entry