LL(1) Grammars

LL(1) Grammar

- A grammar is LL(1) if it is possible to choose the next production by looking at only the next token in the input string.
- The first "L" in LL(1) stands for scanning the input from left to right.
- The second "L" stands for producing a leftmost derivation.
- The "1" stands for using one input symbol of lookahead at each step in making parsing action decisions.

Definition of LL(1) Grammar

- Formally, grammar G is LL(1) if and only if
  - For all productions $A \rightarrow \alpha_1 \mid \alpha_2 \mid \ldots \mid \alpha_n$, $First(\alpha_i) \cap First(\alpha_j) = \emptyset$, $1 \leq i,j \leq n$, $i \neq j$.
  - For every nonterminal A such that First(A) contains ε, $First(A) \cap Follow(A) = \emptyset$.

Or, stated another way…

- A grammar G is LL(1) if and only if all pairs of productions $A \rightarrow \alpha$ and $A \rightarrow \beta$ satisfy the following conditions:
  - First(α) and First(β) are disjoint, and
  - if one of α or β derives ε (assume $\alpha \Rightarrow$ ε), then First(β) and Follow(A) are disjoint.

Example: The Expression Grammar

$E \rightarrow TE'$

$E' \rightarrow +TE' \mid \varepsilon$

$T \rightarrow FT'$

$T' \rightarrow *FT' \mid \varepsilon$

$F \rightarrow (E) \mid$ **id**

- First (E) = First (T) = First (F) = { (, **id** }
- First (E') = { +, ε }
- First (T') = { *, ε }
- Follow (E) = Follow (E') = { $,) }
- Follow (T) = Follow (T') = { +, $,) }
- Follow (F) = { *, +, $,) }

Example: The Expression Grammar

- To determine if a grammar is LL(1), we need look at only nonterminals which have multiple productions.
- $E' \rightarrow +TE' \mid \varepsilon$
  - First(+TE') = {+}; First(ε) = {ε}.
    First(+TE') and First(ε) are disjoint
  - Since First(E') includes ε, we must check if First(E') and Follow(E') are disjoint.
    First(E') = {+, ε} and Follow(E') = {$,)} are disjoint.

Example: The Expression Grammar

- T' $\rightarrow$ *FT' | ε
  - First(*FT') = {*}; First() = {ε}. These sets are disjoint.
  - Since First(T') = {*, ε} includes ε , we must check if First(T') and Follow(T') are disjoint. Since Follow(T') = { +, $,) }, it is clear that the two sets are disjoint.

Example: The Expression Grammar

- $F \rightarrow (E) \mid \mathbf{id}$
  - First ((E)) = { (}; First (**id**) = { **id** }. These sets are disjoint.
  - Neither right hand side derives ε, so the second condition is trivially true.
- Since all pairs of productions $A \rightarrow \alpha$ and $A \rightarrow \beta$ satisfy both conditions of the definition, the expression grammar is LL(1).

A Grammar That Is Not LL(1)

- The grammar

  $A \rightarrow \mathbf{d}A$

  $A \rightarrow \mathbf{d}B$

  $A \rightarrow \mathbf{f}$

  $B \rightarrow \mathbf{g}$

  is not LL(1): $\text{First}(\mathbf{d}A) \cap \text{First}(\mathbf{d}B) = \{\mathbf{d}\} \neq \emptyset$

Another Grammar That Is Not LL(1)

- The grammar below is not LL(1):
  - $S \rightarrow X\mathbf{d}$
  - $X \rightarrow C$
  - $X \rightarrow B\mathbf{a}$
  - $C \rightarrow \varepsilon$
  - $B \rightarrow \mathbf{d}$
- Consider the productions $X \rightarrow C$ and $X \rightarrow B\mathbf{a}$. Since $C \Rightarrow \varepsilon$, then First(B**a**) and Follow(X) must be disjoint. Since First(B**a**) = {**d**} and Follow(X) = {**d**} are not disjoint, the grammar is not LL(1).