Compilers

First Sets

Alex Aiken

First Sets

- Consider non-terminal A, production $A \rightarrow \alpha$, & token t
- $T[A,t] = \alpha$ in two cases:

- If $\alpha \rightarrow^* t\ \beta$
  - α can derive a t in the first position
  - We say that $t \in \text{First}(\alpha)$

$\alpha \not\xrightarrow{*} t$

$t \notin \text{First}(\alpha)$

$\alpha \xrightarrow{*} \epsilon$

- If $A \rightarrow \alpha$ and $\alpha \rightarrow^* \varepsilon$ and $S \rightarrow^* \beta\ A\ t\ \delta$
  - Useful if stack has A, input is t, and A cannot derive t
  - In this case only option is to get rid of A (by deriving ε)
    - Can work only if t can follow A in at least one derivation
  - We say $t \in \text{Follow}(A)$

First Sets

Definition

$$\text{First}(X) = \{\, t \mid X \rightarrow^* t\alpha \,\} \cup \{\, \varepsilon \mid X \rightarrow^* \varepsilon \,\}$$

Algorithm sketch:

1. $\text{First}(t) = \{\, t \,\}$ t is a terminal
2. $\varepsilon \in \text{First}(X)$
   - if $X \rightarrow \varepsilon$
   - if $X \rightarrow A_1 \ldots A_n$ and $\varepsilon \in \text{First}(A_i)$ for $1 \le i \le n$ ε
3. $\text{First}(\alpha) \subseteq \text{First}(X)$ if $X \rightarrow A_1 \ldots A_n\, \alpha$
   – and $\varepsilon \in \text{First}(A_i)$ for $1 \le i \le n$

$$X \rightarrow \underbrace{A_1 \ldots A_n}_{\alpha} \xrightarrow{*} t$$

$$X \xrightarrow{*} \alpha$$

First Sets

- Recall the grammar

$E \rightarrow T X$ $\qquad$ $X \rightarrow + E \mid \varepsilon$

$T \rightarrow (E) \mid \text{int } Y$ $\qquad$ $Y \rightarrow * T \mid \varepsilon$

First(+) = { + }
First(*) = { * }
⋮
{ (}
{) }
{ int }

First(E) ⊇ First(T)
First(T) = { (, int }
First(X) = { +, ε }
First(Y) = { *, ε }