# Compilers

Arthur Hoskey, Ph.D. Farmingdale State College Computer Systems Department





- Follow set. Indicates which terminals can appear after a given nonterminal.
- Follow(A). This is the set of terminals that can appear after A.
- Follow sets help determine which production to use next.



- Follow Set Overview. Check right sides of productions for the target nonterminal you are calculating the follow set of. Check what follows the target nonterminal in each of those productions.
- Concise rules. Calculate Follow(A) where A is some nonterminal in the grammar. Assume a (alpha) and β (beta) are both strings of symbols (either terminals or nonterminals).
  - If A is the starting symbol, then add eof to the Follow(A) set.
  - If there is a production  $Q \rightarrow aA\beta$  then everything in First( $\beta$ ) except for the empty string ( $\lambda$ ) is placed in Follow(A).
  - If there is a production  $Q \rightarrow aA$ , or a production  $Q \rightarrow aA\beta$  where First( $\beta$ ) contains the empty string ( $\lambda$ ) then everything in Follow(Q) is in Follow(A). Basically, if A is at the end of the RHS then add Follow(Q) to Follow(A).

Taken from:

Compilers: Principles, Techniques, and Tools by Aho, Sethi, and Ulman, 1988. Introduction to Compiler Construction by Thomas Parson, 1992.

### **Calculating a Follow Set**

- Here is a more descriptive version of the rules from the previous slide (for calculating a follow set for a target nonterminal):
  - If the target nonterminal is the starting symbol, then add eof to the follow set.
  - If the next symbol is a terminal, then add that terminal to the follow set.
  - If the next symbol is a nonterminal then add the first set of that nonterminal to the follow set (minus the empty string). Do not add the empty string (lambda  $\lambda$ ) as part of this.
  - If the next symbol is a nonterminal and it derives the empty string (lambda  $\lambda$ ), then also add the first set of the symbol that appears after it to the follow set. Keep going to the next nonterminal and adding its first set as long as that nonterminal derives the empty string.
  - If all symbols after the target derive the empty string, then include the follow set of the left side nonterminal.

### **Calculating a Follow Set**

- Calculate the follow set of A for the following productions (A is the start symbol, lowercase are terminals):
- $A \rightarrow stu$
- $B \rightarrow Axyz$





ANSWER Follow(A) = { x, eof } A only appears in one production. x is a terminal that appears to the right of A, so x is in Follow(A). A is the start symbol so eof is in Follow(A).



- Calculate the follow set of A for the following productions (A is the start symbol, lowercase are terminals):
- $A \rightarrow stu$   $B \rightarrow Axyz$  $C \rightarrow mAn$





A appears in the right sides of two productions. x follows A in one and n follows A in the other so add x and n to the follow set. A is the start symbol so eof is included.

#### **ANSWER** Follow(A) = { x, n, eof }

### **Calculating a Follow Set**

A → stu B → ACxyz C → mn Hint: You may need to calculate a First set to calculate Follow(A)



A → stu B → ACxyz  $\lt$ C → mn C appears after A so the First(C) must be included in the follow set of A



### **Calculating a Follow Set**

- Calculate the follow set of A for the following productions (A is the start symbol, lowercase are terminals):
- A → stu B → ACxyz C → mnD D → gAh



#### A appears in two productions



First(C) = { m } ← Calculate First(C) and add it to the follow set

### **Calculating a Follow Set**

- Remember, if a nonterminal can derive an empty string (λ) then we add the first set of the next nonterminal.
- Calculate the follow set of A (A is the start symbol, lowercase are terminals):
- A → stu B → ACmn C →  $\lambda$ C → xyz



- Remember, if a nonterminal can derive an empty string (λ) then we add the first set of the next nonterminal.
- Calculate the follow set of A (A is the start symbol, lowercase are terminals):

A  $\rightarrow$  stu B  $\rightarrow$  ACmn C  $\rightarrow \lambda$ C  $\rightarrow \lambda$ C  $\rightarrow xyz$ C can derive the empty string ( $\lambda$ ) so First(m) must be included in Follow(A) in addition to First(C)- $\lambda$ 

ANSWER Follow(A) = { x, m, eof } First(C) = { x,  $\lambda$  }

The empty string (λ) is NOT included in Follow(A) even though First(C) was added to it. Include eof since A is the start symbol.

### **Calculating a Follow Set**

- Calculate the follow set of A (A is the start symbol, lowercase are terminals):
- $A \rightarrow stu$  $B \rightarrow ACDmn$
- $C \rightarrow \lambda$
- $C \rightarrow xyz$
- $D \rightarrow \lambda$
- $D \rightarrow gh$



 Calculate the follow set of A (A is the start symbol, lowercase are terminals):



#### **ANSWER**

Follow(A) = { x, g, m, eof } First(C) = { x,  $\lambda$  } First(D) = { g,  $\lambda$  } First(C), First(D), and First(m) are all included in Follow(A). The empty string (λ) is NOT included in Follow(A) even though it appears in both First(C) and First(D). Include eof since A is the start symbol.

### **Calculating a Follow Set**

 Calculate the follow set of <u>C</u> (A is the start symbol, lowercase are terminals):

 $\begin{array}{c} A \rightarrow B \\ B \rightarrow w \\ C \rightarrow x \\ D \rightarrow Ey \\ E \rightarrow zC \end{array}$ 

Hint: C appears at the end of the RHS of a production so you may need to calculate another follow set



 Calculate the follow set of <u>C</u> (A is the start symbol, lowercase are terminals):



- Calculate the follow set of <u>B</u> (A is the start symbol, lowercase are terminals):
- $\begin{array}{c} A \rightarrow sBC \\ B \rightarrow p \end{array}$
- $C \rightarrow \lambda$
- $C \rightarrow Bz$

Hint: Begin with calculating the first sets then calculate the follow sets



 Calculate the follow set of <u>B</u> (A is the start symbol, lowercase are terminals):



#### **ANSWER**

 $Follow(B) = {First(C)-\lambda + First(z) + Follow(A)} = {p, z, eof}$ 

```
First(C) = { p, λ }
First(z) = { z }
Follow(A) = { eof }
```

### **Calculating a Follow Set**

- What are the first and follow sets for the nonterminals A, B, and C?
- A is the start symbol  $A \rightarrow BCdef$
- $B \rightarrow mn$
- $B \rightarrow \lambda$
- $C \rightarrow qr$
- $C \to \lambda$

## Hint: Start by calculating first sets

## Follow Set – Example

 What are the first and follow sets for the nonterminals A, B, and C?

A is the start symbol  $A \rightarrow BCdef$   $B \rightarrow mn$   $B \rightarrow \lambda$  $C \rightarrow qr$ 

 $C \rightarrow \lambda$ 

#### **Answer**

First(A) = { m, q, d }  
First(B) = { m, 
$$\lambda$$
 }  
First(C) = { q,  $\lambda$  }  
Follow(A) = { eof }  
Follow(B) = { q, d }  
Follow(C) = { d }

Follow(A) only has eof because it is the start symbol, and it does not appear in any right sides.

Follow(B) includes First(C) and First(d) because C derives the empty string.

Follow(C) only has d because d is a terminal

### Now on to First+ set...



• First+. This is the augmented first set. We have a production and want to know its first set. If the right side can disappear, we need to know what comes after the left side nonterminal (useful when making decisions about which production to apply).

First+(A→β) =  $\begin{bmatrix} First(β) & if empty string not in First(β) \\ First(β) U Follow(A) & otherwise \end{bmatrix}$ 

- If First(β) contains the empty string, then it can disappear.
- Since First(β) can disappear we need to know what can appear directly after the left side (A).
- Follow(A) contains exactly what can appear after A.
- The union of First(β) and Follow(A) has the first tokens.
- Note: First+ sets are also referred to as a Predict sets.
- Taken from:

Engineering a Compiler, 2<sup>nd</sup> Edition by Cooper and Torczon, 2012

### First+ Set





• What are all the First+ sets of the productions below? A is the start symbol  $A \rightarrow BC$   $B \rightarrow m$   $B \rightarrow M$   $B \rightarrow \lambda$  $C \rightarrow q$ 

 $C \to \lambda$ 



• What are all the First+ sets of the productions below? A is the start symbol  $A \rightarrow BC$  $B \rightarrow m$  $B \rightarrow \lambda$  $C \rightarrow q$  $C \rightarrow \lambda$ Answer First(C) = { q,  $\lambda$ , eof } First(B) = { m,  $\lambda$  } First(A) = (First(B)- $\lambda$ ) + (First(C)- $\lambda$ ) +  $\lambda$  = { m, q,  $\lambda$  }  $Follow(A) = \{ eof \}$ Follow(B) = (First(C)- $\lambda$ ) + Follow(A) = { q, eof }  $Follow(C) = \{ eof \}$ First+(A $\rightarrow$ BC) = { m, q,  $\lambda$ , eof }  $First+(B \rightarrow m) = \{m\}$ First+(B $\rightarrow \lambda$ ) = {  $\lambda$  } + Follow(B) = {  $\lambda$ , q, eof } First+( $C \rightarrow q$ ) = { q } First+(C $\rightarrow \lambda$ ) = {  $\lambda$  } + Follow(C) = {  $\lambda$ , eof } First+ Set – More Examp



