MIT 6.035
Parse Table Construction

Martin Rinard
Laboratory for Computer Science
Massachusetts Institute of Technology

Parse Tables (Review)

	ACTION			Goto
State	**(**	**)**	**$**	**X**
s0	shift to s2	error	error	goto s1
s1	error	error	accept	
s2	shift to s2	shift to s5	error	goto s3
s3	error	shift to s4	error	
s4	reduce (2)	reduce (2)	reduce (2)	
s5	reduce (3)	reduce (3)	reduce (3)	

- Implements finite state control
- At each step, look up
  - Table[top of state stack] [input symbol]
- Then carry out the action

Parse Tables (Review)

	ACTION			Goto
State	**(**	**)**	**$**	**X**
s0	shift to s2	error	error	goto s1
s1	error	error	accept	
s2	shift to s2	shift to s5	error	goto s3
s3	error	shift to s4	error	
s4	reduce (2)	reduce (2)	reduce (2)	
s5	reduce (3)	reduce (3)	reduce (3)	

- Shift to s*n*
  - Push input token into the symbol stack
  - Push s*n* into state stack
  - Advance to next input symbol

Parse Tables (Review)

	ACTION			Goto
State	**(**	**)**	**$**	**X**
s0	shift to s2	error	error	goto s1
s1	error	error	accept	
s2	shift to s2	shift to s5	error	goto s3
s3	error	shift to s4	error	
s4	reduce (2)	reduce (2)	reduce (2)	
s5	reduce (3)	reduce (3)	reduce (3)	

- Reduce (*n*)
  - Pop both stacks as many times as the number of symbols on the RHS of rule *n*
  - Push LHS of rule *n* into symbol stack

Parser Generators and Parse Tables

- Parser generator (YACC, CUP)
  - Given a grammar
  - Produces a (shift-reduce) parser for that grammar
- Process grammar to synthesize a DFA
  - Contains states that the parser can be in
  - State transitions for terminals and non-terminals
- Use DFA to create an parse table
- Use parse table to generate code for parser

Example

- The grammar

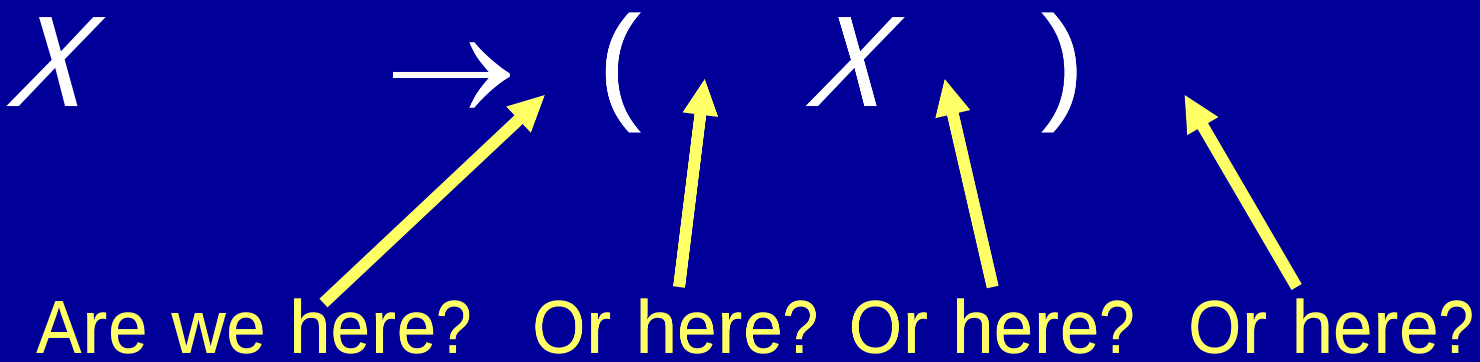
$$S \rightarrow X\ \$ \qquad (1)$$

$$X \rightarrow (X) \qquad (2)$$

$$X \rightarrow (\ ) \qquad (3)$$

DFA States Based on Items

- We need to capture how much of a given production we have scanned so far



Items

- We need to capture how much of a given production we have scanned so far

$$X \rightarrow (\quad X \quad)$$

- Production Generates 4 items
  - $X \rightarrow \bullet (X)$
  - $X \rightarrow (\bullet X)$
  - $X \rightarrow (X \bullet)$
  - $X \rightarrow (X) \bullet$

Example of Items

- The grammar

$S \rightarrow X\ \$$

$X \rightarrow (X)$

$X \rightarrow (\ )$

- Items

$S \rightarrow \bullet X\$$

$S \rightarrow X \bullet \$$

$X \rightarrow \bullet (X)$

$X \rightarrow (\bullet X)$

$X \rightarrow (X \bullet)$

$X \rightarrow (X) \bullet$

$X \rightarrow \bullet (\ )$

$X \rightarrow (\bullet)$

$X \rightarrow (\ ) \bullet$

Notation

- If write production as $A \rightarrow \alpha\ c\ \beta$
  - α is sequence of grammar symbols, can be terminals and nonterminals in sequence
  - c is terminal
  - β is sequence of grammar symbols, can be terminals and nonterminals in sequence
- If write production as $A \rightarrow \alpha \bullet B\ \beta$
  - α, β as above
  - B is a single grammar symbol, either terminal or nonterminal

Key idea behind items

- States correspond to sets of items
- If the state contains the item $A \rightarrow \alpha \bullet c\ \beta$
  - Parser is expecting to eventually reduce using the production $A \rightarrow \alpha\ c\ \beta$
  - Parser has already parsed an α
  - It expects the input may contain c, then β
- If the state contains the item $A \rightarrow \alpha \bullet$
  - Parser has already parsed an α
  - Will reduce using $A \rightarrow \alpha$
- If the state contains the item $S \rightarrow \alpha \bullet \$$ and the input buffer is empty
  - – Parser accepts input

Correlating Items and Actions

- If the current state contains the item $A \rightarrow \alpha \bullet c\ \beta$ and the current symbol in the input buffer is c
  - Parser shifts c onto stack
  - Next state will contain $A \rightarrow \alpha\ c \bullet \beta$
- If the current state contains the item $A \rightarrow \alpha \bullet$
  - Parser reduces using $A \rightarrow \alpha$
- If the current state contains the item $S \rightarrow \alpha \bullet \$$ and the input buffer is empty
  - Parser accepts input

Closure() of a set of items

- Closure finds all the items in the same “state”
- Fixed Point Algorithm for Closure(I)
  - Every item in I is also an item in Closure(I)
  - If $A \rightarrow \alpha \bullet B \beta$ is in Closure(I) and $B \rightarrow \bullet \gamma$ is an item, then add $B \rightarrow \bullet \gamma$ to Closure(I)
  - Repeat until no more new items can be added to Closure(I)

Example of Closure

- Closure($\{X \to (\bullet X)\}$)

$$\left\{ \begin{array}{l} X \to (\bullet X) \\ X \to \bullet (X) \\ X \to \bullet () \end{array} \right\}$$

- Items

$S \to \bullet X\$$

$S \to X \bullet \$$

$X \to \bullet (X)$

$X \to (\bullet X)$

$X \to (X \bullet)$

$X \to (X) \bullet$

$X \to \bullet ()$

$X \to (\bullet)$

$X \to () \bullet$

Another Example

- closure($\{S \rightarrow \bullet X\$\}$)

$$\left\{ \begin{array}{l} S \rightarrow \bullet X\$ \\ X \rightarrow \bullet (X) \\ X \rightarrow \bullet (\ ) \end{array} \right\}$$

- Items

$S \rightarrow \bullet X\$$

$S \rightarrow X \bullet \$$

$X \rightarrow \bullet (X)$

$X \rightarrow (\bullet X)$

$X \rightarrow (X \bullet)$

$X \rightarrow (X) \bullet$

$X \rightarrow \bullet (\ )$

$X \rightarrow (\bullet)$

$X \rightarrow (\ ) \bullet$

Goto() of a set of items

- Goto finds the new state after consuming a grammar symbol while at the current state

- Algorithm for Goto(I, X)
  where I is a set of items
  and X is a grammar symbol

$$\text{Goto}(I, X) = \text{Closure}(\ \{\ A \rightarrow \alpha\ X \bullet \beta \mid A \rightarrow \alpha \bullet X\ \beta \text{ in } I\ \})$$

- goto is the new set obtained by “moving the dot” over X

Example of Goto

- Goto $(\{X \rightarrow (\bullet X)\}, X)$

$$\left\{ X \rightarrow (X \bullet \ ) \right\}$$

- Items

$S \rightarrow \bullet X\$$

$S \rightarrow X \bullet \$$

$X \rightarrow \bullet (X)$

$X \rightarrow (\bullet X)$

$X \rightarrow (X \bullet)$

$X \rightarrow (X) \bullet$

$X \rightarrow \bullet (\ )$

$X \rightarrow (\bullet)$

$X \rightarrow (\ ) \bullet$

Another Example of Goto

- Goto $(\{X \to \bullet(X)\}, ()$

$$\left\{\begin{array}{l} X \to (\ \bullet X) \\ X \to \bullet (X) \\ X \to \bullet (\ ) \end{array}\right\}$$

- Items

$S \to \bullet X\$$

$S \to X \bullet \$$

$X \to \bullet (X)$

$X \to (\bullet X)$

$X \to (X \bullet)$

$X \to (X) \bullet$

$X \to \bullet (\ )$

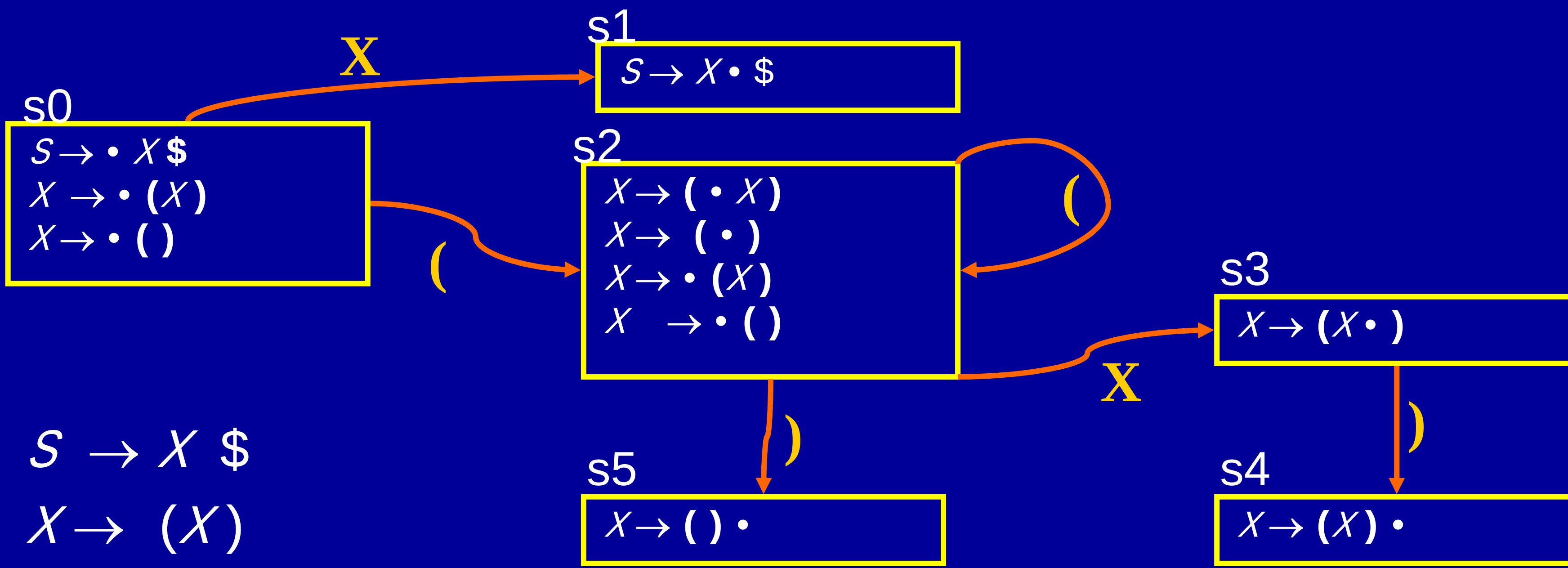
$X \to (\bullet)$

$X \to (\ ) \bullet$

Building the DFA states

- Start with the item $S \rightarrow \bullet \beta \$$
- Create the first state to be Closure({ $S \rightarrow \bullet \beta \$$})
- Pick a state I
  - for each item $A \rightarrow \alpha \bullet X \beta$ in I
    - find Goto(I, X)
    - if Goto(I, X) is not already a state, make one
    - Add an edge X from state I to Goto(I, X) state
- Repeat until no more additions possible

DFA Example



$S \rightarrow X \; \$$

$X \rightarrow (X)$

$X \rightarrow (\;)$

Constructing A Parse Engine

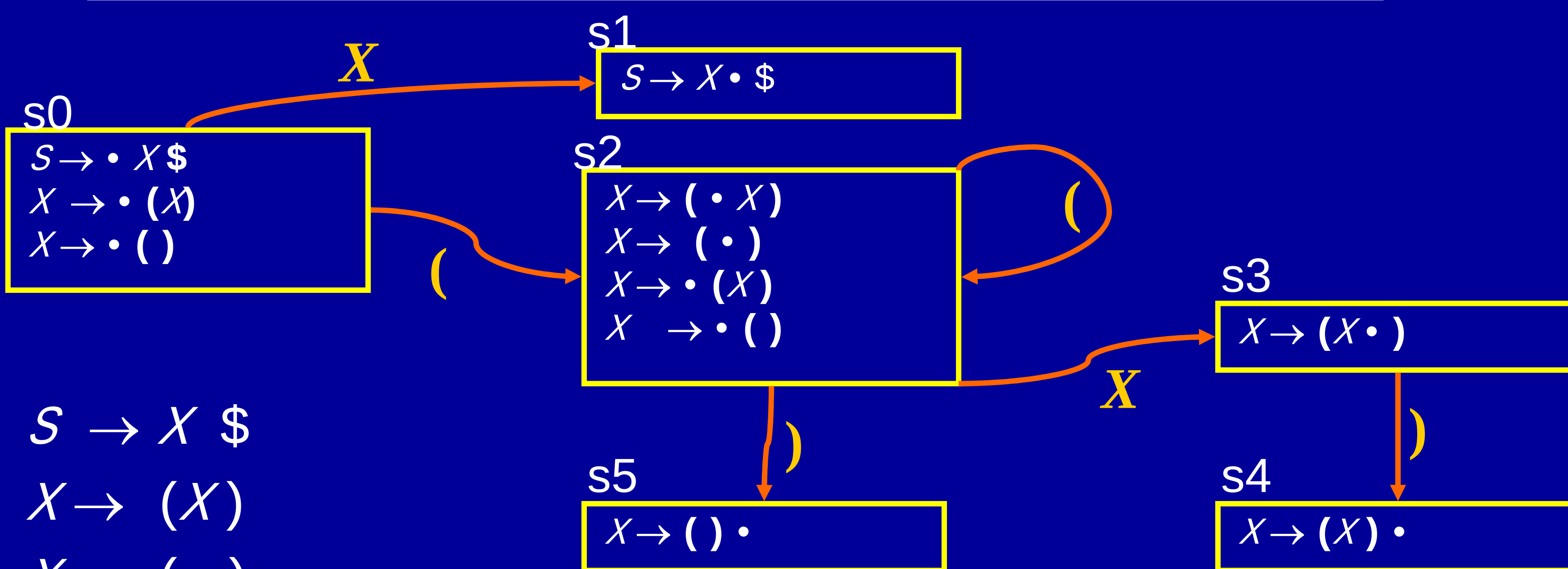
- Build a DFA - DONE
- Construct a parse table using the DFA

Creating the parse tables

- For each state
  - Transition to another state using a terminal symbol is a shift to that state (*shift to sn*)
  - Transition to another state using a non-terminal is a goto to that state (*goto sn*)
  - If there is an item $A \rightarrow \alpha \bullet$ in the state do a reduction with that production for all terminals (*reduce k*)

Building Parse Table Example

	ACTION			Goto
State	**(**	**)**	**$**	**X**
s0	shift to s2	error	error	goto s1
s1	error	error	accept	
s2	shift to s2	shift to s5	error	goto s3
s3	error	shift to s4	error	
s4	reduce (2)	reduce (2)	reduce (2)	
s5	reduce (3)	reduce (3)	reduce (3)	



$S \rightarrow X\ \$$

$X \rightarrow (X)$

$X \rightarrow (\ )$

Potential Problem

- No lookahead
- Vulnerable to unnecessary conflicts
  - Shift/Reduce Conflicts (may reduce too soon in some cases)
  - Reduce/Reduce Conflicts
- Solution: Lookahead
  - Only for reductions - reduce only when next symbol can occur after nonterminal from production
  - Systematic lookahead, split states based on next symbol, action is always a function of next symbol
  - Can generalize to look ahead multiple symbols

Reduction-Only Lookahead Parsing

- If a state contains $A \to \beta \bullet$
- Reduce by $A \to \beta$ only if next input symbol can follow A in some derivation
- Example Grammar

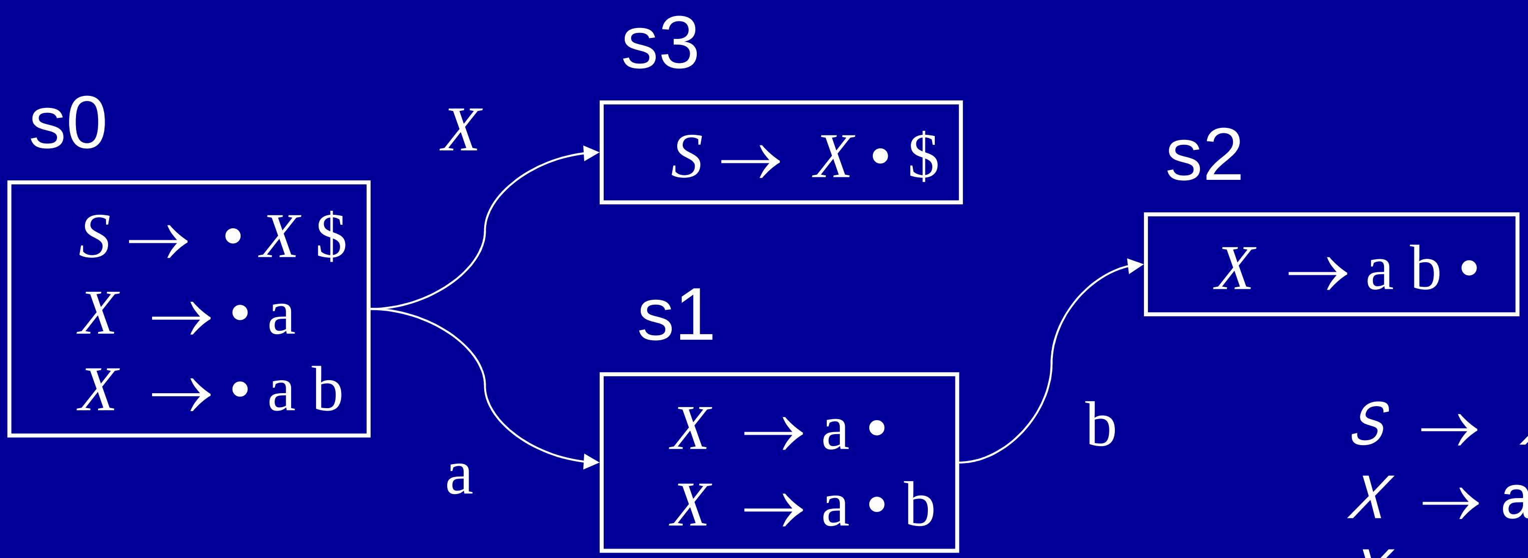
$$S \to X\$$$

$$X \to a$$

$$X \to a\ b$$

Parser Without Lookahead

	ACTION			Goto
State	**a**	**b**	**$**	**X**
s0	shift to s1	error	error	goto s3
s1	reduce(2)	S/R Conflict	reduce(2)	
s2	reduce(3)	reduce(3)	reduce(3)	
s3	error	error	accept	



$S \rightarrow X\,\$$
$X \rightarrow a$
$X \rightarrow a\ b$

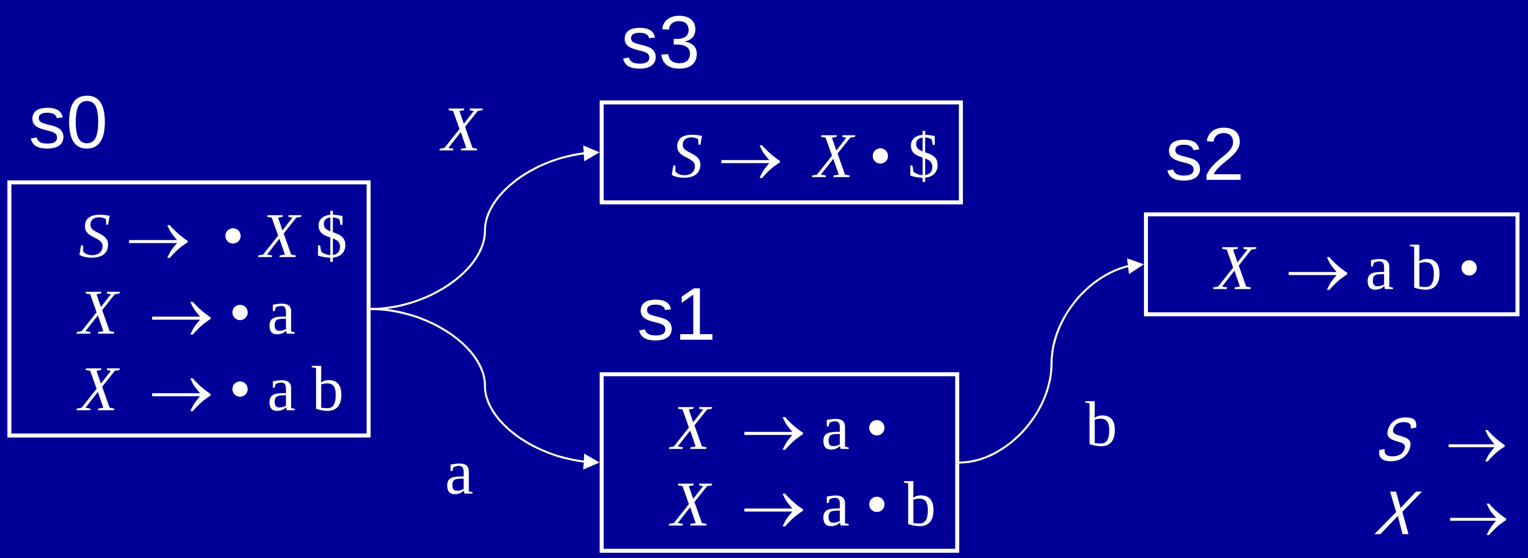
Creating parse tables with reduction-only lookahead

- For each state
  - Transition to another state using a terminal symbol is a shift to that state (*shift to sn*) (same as before)
  - Transition to another state using a non-terminal is a goto that state (*goto sn*) (same as before)
  - If there is an item $X \rightarrow \alpha \bullet$ in the state do a reduction with that production whenever the current input symbol T may follow X in some derivation (more precise than before)
- Eliminates useless reduce actions

New Parse Table

b never follows X in any derivation
resolve shift/reduce conflict to shift

		ACTION		Goto
State	a	b	$	X
s0	shift to s1	error	error	goto s3
s1	reduce(2)	shift to s2	reduce(2)	
s2	reduce(3)	reduce(3)	reduce(3)	
s3	error	error	accept	



$S \rightarrow X \$$
$X \rightarrow a$
$X \rightarrow a b$

More General Lookahead

- Items contain potential lookahead information, resulting in more states in finite state control
- Item of the form [$A \rightarrow \alpha \bullet \beta$ T] says
  - The parser has parsed an α
  - If it parses a β and the next symbol is T
  - Then parser should reduce by $A \rightarrow \alpha\ \beta$

- In addition to current parser state, all parser actions are function of lookahead symbols

Terminology

- Many different parsing techniques
  - Each can handle some set of CFGs
  - Categorization of techniques

Terminology

- Many different parsing techniques
  - Each can handle some set of CFGs
  - Categorization of techniques

□□ (□)

Terminology

- Many different parsing techniques
  - Each can handle some set of CFGs
  - Categorization of techniques

  - **L** - parse from left to right
  - **R** - parse from right to left

Terminology

- Many different parsing techniques
  - Each can handle some set of CFGs
  - Categorization of techniques

  - **L** - leftmost derivation
  - **R** - rightmost derivation

□□ (□)

Terminology

- Many different parsing techniques
  - Each can handle some set of CFGs
  - Categorization of techniques

  - Number of lookahead characters

□□ (□)

Terminology

- Many different parsing techniques
  - Each can handle some set of CFGs
  - Categorization of techniques

  - Examples: LL(0), LR(1)
  - This lecture
    - LR(0) parser
    - SLR parser – LR(0) parser augmented with follow information

L R (k)

Summary

- Parser generators – given a grammar, produce a parser
- Standard technique
  - Automatically build a pushdown automaton
  - Obtain a shift-reduce parser
    - Finite state control plus push down stack
    - Table driven implementation
- Conflicts: Shift/Reduce, Reduce/Reduce
- Use of lookahead to eliminate conflicts
  - SLR parsing (eliminates useless reduce actions)
  - LR(k) parsing (lookahead throughout parser)

Follow() sets in SLR Parsing

For each non-terminal *A*, Follow(*A*) is the set of terminals that can come after *A* in some derivation

Constraints for Follow()

- $ $\in$ Follow(S), where S is the start symbol
- If $A \rightarrow \alpha B \beta$ is a production then First(β) $\subseteq$ Follow(B)
- If $A \rightarrow \alpha B$ is a production then Follow(A) $\subseteq$ Follow(B)
- If $A \rightarrow \alpha B \beta$ is a production and β derives ε
  then Follow(A) $\subseteq$ Follow(B)

Algorithm for Follow

for all nonterminals NT

  Follow(NT) = { }

Follow(S) = { \$ }

while Follow sets keep changing

  for all productions $A \to \alpha B \beta$

    Follow(B) = Follow(B) $\cup$ First(β)

    if (β derives ε) Follow(B) = Follow(B)$\cup$Follow(A)

  for all productions $A \to \alpha B$

    Follow(B) = Follow(B)$\cup$Follow(A)

Augmenting Example with Follow

- Example Grammar for Follow

$$S \rightarrow X\$$$

$$X \rightarrow \text{a}$$

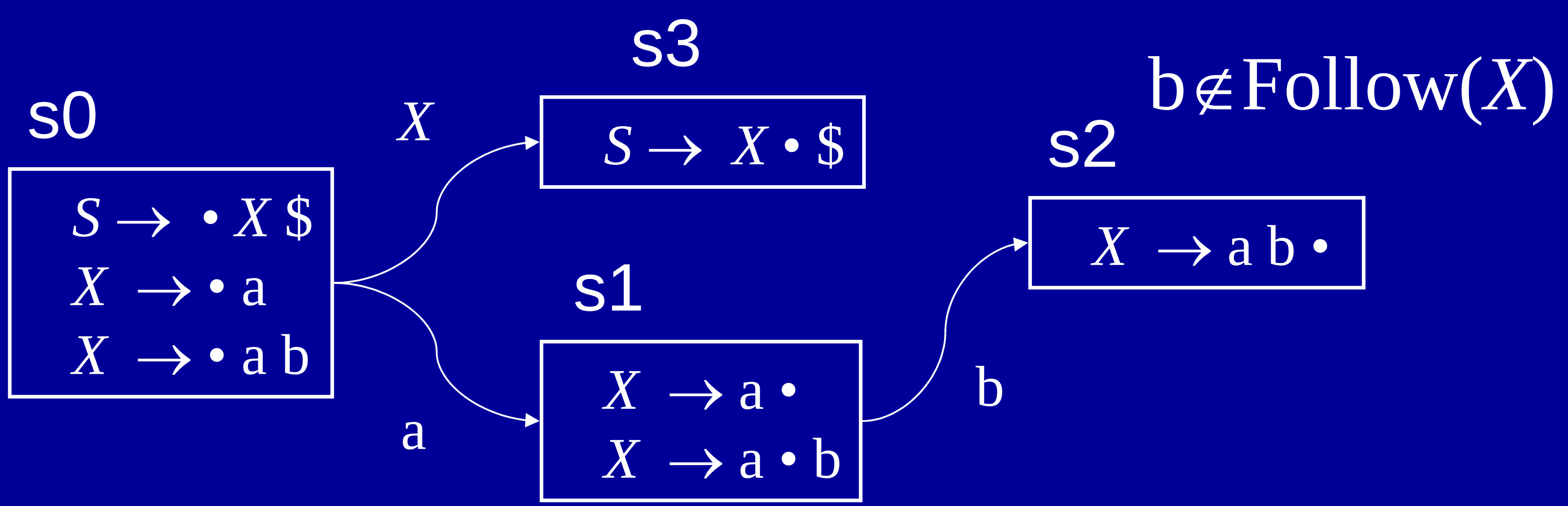
$$X \rightarrow \text{a b}$$

$$\text{Follow}(S) = \{ \$ \}$$

$$\text{Follow}(X) = \{ \$ \}$$

SLR Eliminates Shift/Reduce Conflict

		ACTION		Goto
State	**a**	**b**	**$**	**X**
s0	shift to s1	error	error	goto s3
s1	reduce(2)	shift to s2	reduce(2)	
s2	reduce(3)	reduce(3)	reduce(3)	
s3	error	error	accept	



Basic Idea Behind LR(1)

- Split states in LR(0) DFA based on lookahead
- Reduce based on item and lookahead

LR(1) Items

- Items will keep info on
  - production
  - right-hand-side position (the dot)
  - look ahead symbol
- LR(1) item is of the form [$A \rightarrow \alpha \bullet \beta$ T]
  - $A \rightarrow \alpha\ \beta$ is a production
  - The dot in $A \rightarrow \alpha \bullet \beta$ denotes the position
  - T is a terminal or the end marker ($)

Meaning of LR(1) Items

- Item $[A \to \alpha \bullet \beta \quad T]$ means
  - The parser has parsed an α
  - If it parses a β and the next symbol is T
  - Then parser should reduce by $A \to \alpha\ \beta$

- The grammar

  $S \rightarrow X\$$

  $X \rightarrow (X)$

  $X \rightarrow \varepsilon$

- Terminal symbols
  - '(' ')'
- End of input symbol
  - '$'

LR(1) Items

$[S \rightarrow \bullet X\$ \quad)]$

$[S \rightarrow \bullet X\$ \quad (]$

$[S \rightarrow \bullet X\$ \quad \$]$

$[S \rightarrow X \bullet \$ \quad)]$

$[S \rightarrow X \bullet \$ \quad (]$

$[S \rightarrow X \bullet \$ \quad \$]$

$[X \rightarrow \bullet (X) \quad)]$

$[X \rightarrow \bullet (X) \quad (]$

$[X \rightarrow \bullet (X) \quad \$]$

$[X \rightarrow (\bullet X) \quad)]$

$[X \rightarrow (\bullet X) \quad (]$

$[X \rightarrow (\bullet X) \quad \$]$

$[X \rightarrow (X \bullet \quad)]$

$[X \rightarrow (X \bullet) \quad (]$

$[X \rightarrow (X \bullet) \quad \$]$

$[X \rightarrow (X) \bullet \quad)]$

$[X \rightarrow (X) \bullet \quad (]$

$[X \rightarrow (X) \bullet \quad \$]$

$[X \rightarrow \bullet \quad)]$

$[X \rightarrow \bullet \quad (]$

$[X \rightarrow \bullet \quad \$]$

Creating a LR(1) Parser Engine

- Need to define Closure() and Goto() functions for LR(1) items
- Need to provide an algorithm to create the DFA
- Need to provide an algorithm to create the parse table

Closure algorithm

Closure(I)

repeat

for all items $[A \to \alpha \bullet X \beta \quad c]$ in I

for any production $X \to \gamma$

for any $d \in First(\beta c)$

$I = I \cup \{ [X \to \bullet \gamma \quad d] \}$

until I does not change

Goto algorithm

Goto(I, X)

J = { }

for any item $[A \to \alpha \bullet X \beta \quad c]$ in I

$J = J \cup \{[A \to \alpha X \bullet \beta \quad c]\}$

return Closure(J)

Building the LR(1) DFA

- Start with the item $[\langle S' \rangle \rightarrow \bullet \langle S \rangle \$ \quad I]$
  - I irrelevant because we will never shift $
- Find the closure of the item and make an state
- Pick a state I
  - for each item $[A \rightarrow \alpha \bullet X \beta \quad c]$ in I
    - find Goto(I, X)
    - if Goto(I, X) is not already a state, make one
    - Add an edge X from state I to Goto(I, X) state
- Repeat until no more additions possible

Creating the parse tables

- For each LR(1) DFA state
  - Transition to another state using a terminal symbol is a shift to that state (*shift to sn*)
  - Transition to another state using a non-terminal symbol is a goto that state (*goto sn*)
  - If there is an item $[A \rightarrow \alpha \bullet\ a]$ in the state, action for input symbol a is a reduction via the production $A \rightarrow \alpha$ (*reduce k*)

LALR(1) Parser

- Motivation
  - LR(1) parse engine has a large number of states
  - Simple method to eliminate states
- If two LR(1) states are identical except for the look ahead symbol of the items
  Then Merge the states
- Result is LALR(1) DFA
- Typically has many fewer states than LR(1)
- May also have more reduce/reduce conflicts