

Compiler design

Lecture 6: Bottom-Up Parsing (EaC§3.4)

Timestamp: 2025/01/17 10:06:00

Christophe Dubach

Winter 2025

Top-Down Parser

A Top-Down parser builds a derivation by working from the start symbol to the input sentence.



Bottom-Up Parser

A Bottom-Up parser builds a derivation by working from the input sentence back to the start symbol.



Bottom-Up Parsing

Example: CFG

Goal $::= a A B e$

A $::= A b c$

A $::= b$

B $::= d$

Input: abbcde

Bottom-Up Parsing

abbcde

Bottom-Up Parsing

Example: CFG

Goal $::= a A B e$

A $::= A b c$

A $::= b$

B $::= d$

Input: abbcde

Bottom-Up Parsing

a**b**bcde

aAbcde

Bottom-Up Parsing

Example: CFG

Goal $::= a A B e$

$A ::= A b c$

$A ::= b$

$B ::= d$

Input: abcde

Bottom-Up Parsing

abcde

aAbcde

aAde

Bottom-Up Parsing

Example: CFG

Goal $::= a A B e$

A $::= A b c$

A $::= b$

B $::= d$

Input: abcde

Bottom-Up Parsing

abcde

aAbcde

aAde

aABe

Bottom-Up Parsing

Example: CFG

Goal $::=$ a A B e

A $::=$ A b c

A $::=$ b

B $::=$ d

Input: abcde

Bottom-Up Parsing

abcde

aAbcde

aAde

aABe

Goal

Bottom-Up Parsing

Example: CFG

Goal ::= a A B e

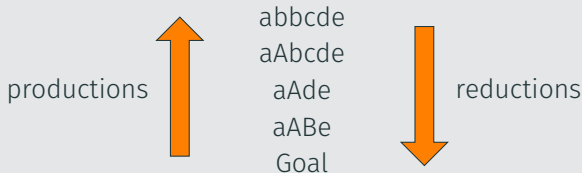
A ::= A b c

A ::= b

B ::= d

Input: abbcde

Bottom-Up Parsing



Note that the production follows a **rightmost** derivation.

Leftmost vs Rightmost derivation

Leftmost vs Rightmost derivation

Example: CFG

Goal ::= a A B e

A ::= A b c | b

B ::= d

Leftmost derivation

Goal

aABe

aAbcBe

abbcBe

abbcede

LL parsers

Rightmost derivation

Goal

aABe

aAde

aAbcde

abbcede

LR parsers

Shift-Reduce Parser

Shift-reduce parser

- It consists of a stack and the input
- It uses four actions:
 1. **shift**: next symbol is shifted onto the stack
 2. **reduce**: pop the symbols $Y_n, \dots, Y_1$ from the stack that form the right member of a production $X ::= Y_n, \dots, Y_1$
 3. **accept**: stop parsing and report success
 4. **error**: error reporting routine

How does the parser know when to shift or when to reduce?

Similarly to a top-down parser, could back-track if wrong decision made or look ahead to decide.

Can build a DFA to decide when we should shift or reduce (will not see the construction process in this course).

The DFA recognizes **handles** on the stack: *i.e.* parts of inputs that can be reduced.

Handle-recognizer DFA

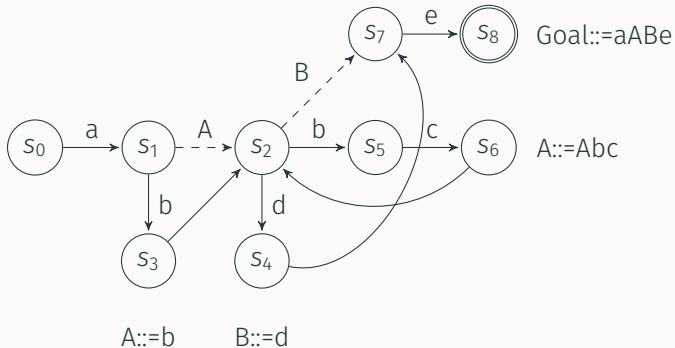
Example CFG

Goal ::= a A B e

A ::= A b c | b

B ::= d

Corresponding handle-recognizer DFA:



In states $\{s_3, s_4, s_6, s_8\}$, we reduce.

Shift-reduce parser

Example: CFG

Goal ::= a A B e

A ::= A b c | b

B ::= d

Operation:

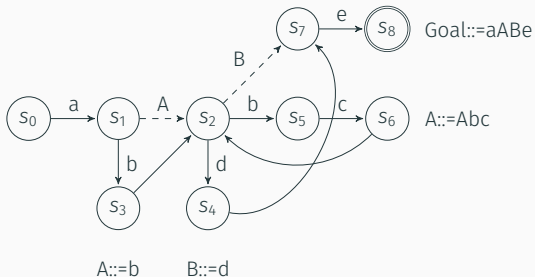
Input

abcde

Stack

State

s0



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation: shift

Input

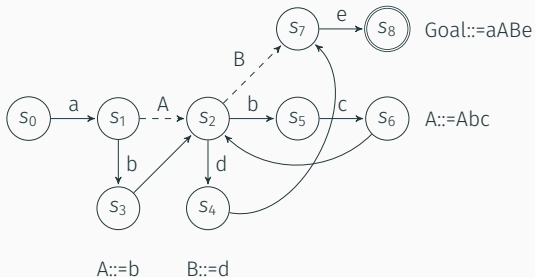
bbcde

Stack

a

State

s1



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation: **shift**

Input

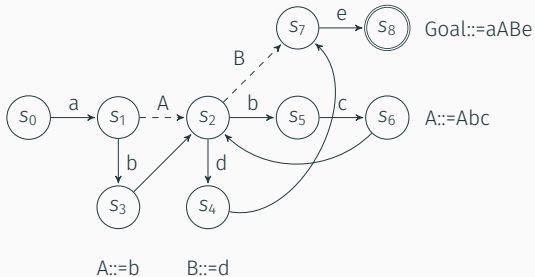
bcde

Stack

ab

State

s3



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation:

Input

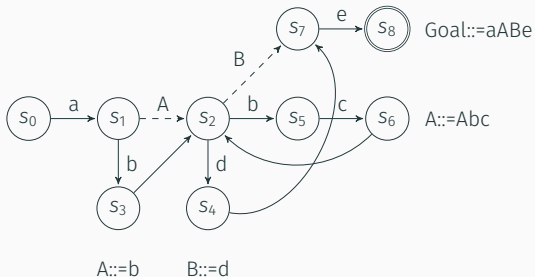
bcde

Stack

ab

State

s3



Shift-reduce parser

Example: CFG

Goal ::= a A B e

A ::= A b c | b

B ::= d

Operation: **reduce**

Input

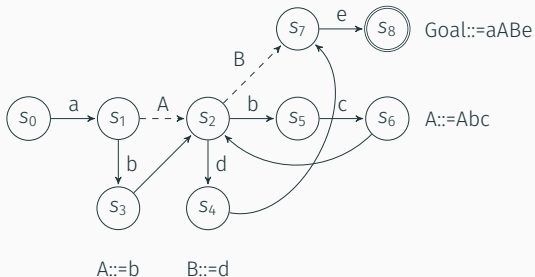
bcde

Stack

aA

State

s2



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation: **shift**

Input

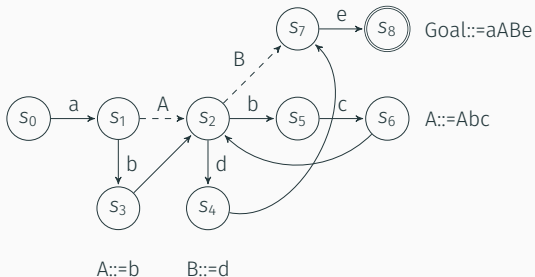
cde

Stack

aAb

State

s5



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation:

Input

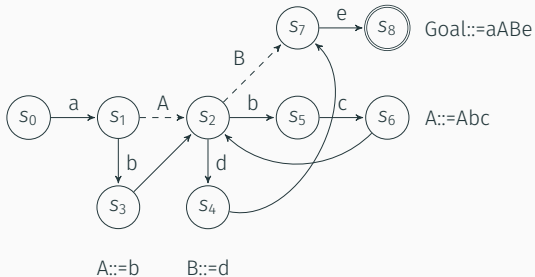
cde

Stack

aAb

State

s5



Shift-reduce parser

Example: CFG

Goal ::= a A B e

A ::= A b c | b

B ::= d

Operation: shift

Input

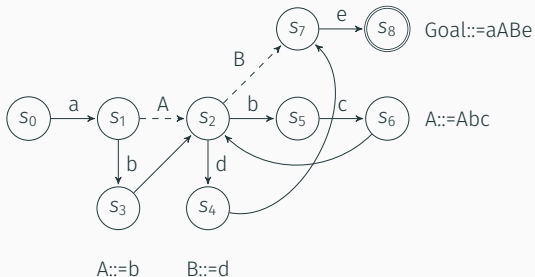
de

Stack

aAbc

State

s6



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation: **reduce**

Input

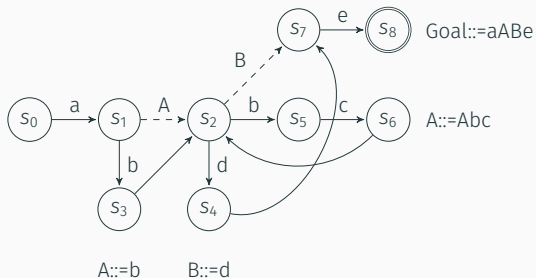
de

Stack

aA

State

s2



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation: **shift**

Input

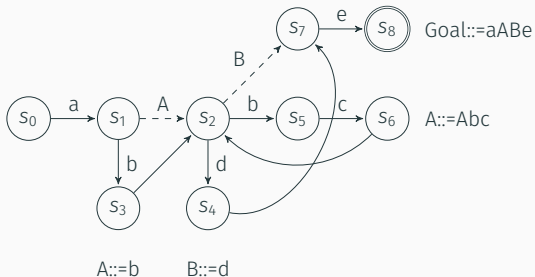
e

Stack

aAd

State

s4



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation: **reduce**

Input

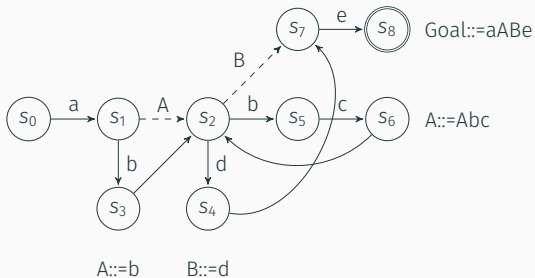
e

Stack

aAB

State

s7



Shift-reduce parser

Example: CFG

Goal $::= a A B e$

A $::= A b c \mid b$

B $::= d$

Operation: shift

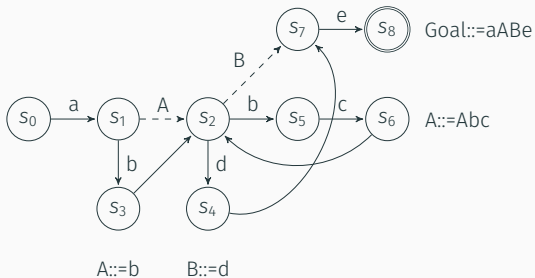
Input

Stack

aABe

State

s8



Shift-reduce parser

Example: CFG

Goal ::= a A B e

A ::= A b c | b

B ::= d

Operation: **reduce**

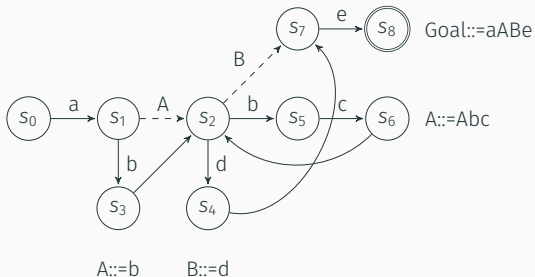
Input

Stack

Goal

State

s8



Top-Down vs Bottom-Up Parsing

Top-Down / LL parsers (*e.g.* recursive descent parser)

- 👍 Easy to write by hand
- 👍 Easy to integrate with the compiler
- 👎 Supports a smaller class of grammars
 - ⇒ cannot handle left recursion in the grammar
- 👎 Recursion might lead to performance issues
 - 👍 Table encoding possible for better performance

Top-Down vs Bottom-Up Parsing

Top-Down / LL parsers (*e.g.* recursive descent parser)

- 👍 Easy to write by hand
- 👍 Easy to integrate with the compiler
- 👎 Supports a smaller class of grammars
 - ⇒ cannot handle left recursion in the grammar
- 👎 Recursion might lead to performance issues
 - 👍 Table encoding possible for better performance

Bottom-Up / LR parsers (*e.g.* shift-reduce parser)

- 👍 Very efficient (no recursion)
- 👍 Supports a larger class of grammar
 - Handles left/right recursion in the grammar
- 👎 Harder to write by hand
 - ⇒ Requires generation tools
- 👎 Hard to integrate in compiler

Real-world examples of parser technology used

Parser generators:

- YACC: bottom up (LR)
- ANTLR: recursive descent (LL)
- JavaCC: recursive descent (LL)

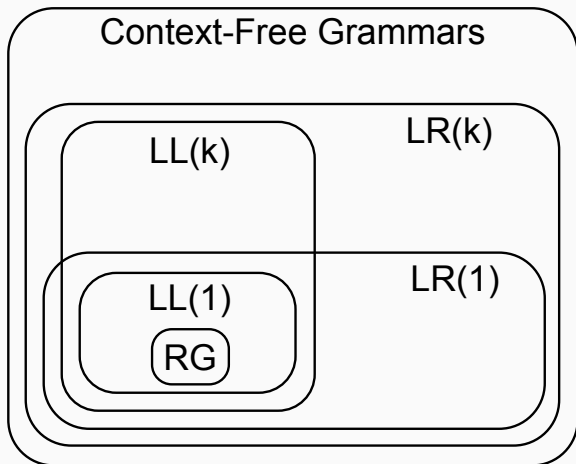
C compilers

- LLVM: hand-written recursive descent parser (LL)
- GCC: started with parser generator (YACC $\Rightarrow$ LR),
now uses hand-written recursive descent (LL)

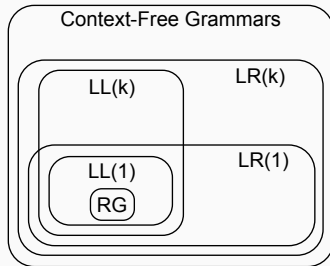
Java compilers

- Eclipse compiler frontend:
auto-generated using Jikes Parser Generator, bottom-up (LR)
- IntelliJ compiler frontend: hand-written recursive descent (LL)
- OpenJDK compiler frontend:
hand-written recursive descent (LL)

<https://github.com/openjdk/jdk/blob/master/src/jdk.compiler/share/classes/com/sun/tools/javac/parser/JavacParser.java>



Language vs. Grammar



Language $\neq$ Grammar

- A language can be defined by more than one grammar
- These grammars might be of different “complexity” (LL(1), LL(k), LR(k))
- $\Rightarrow$ Language complexity $\neq$ grammar complexity

- Parse tree and abstract syntax tree