

CS 314 Principles of Programming Languages

Lecture 4: Syntax Analysis (Parsing)

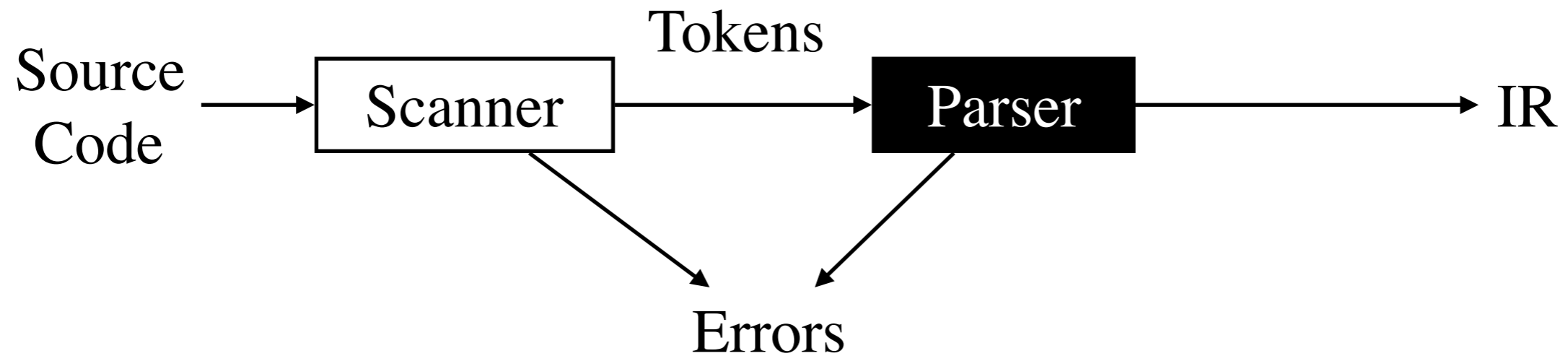
Zheng (Eddy) Zhang



Rutgers University

January 29, 2018

Review: Front End



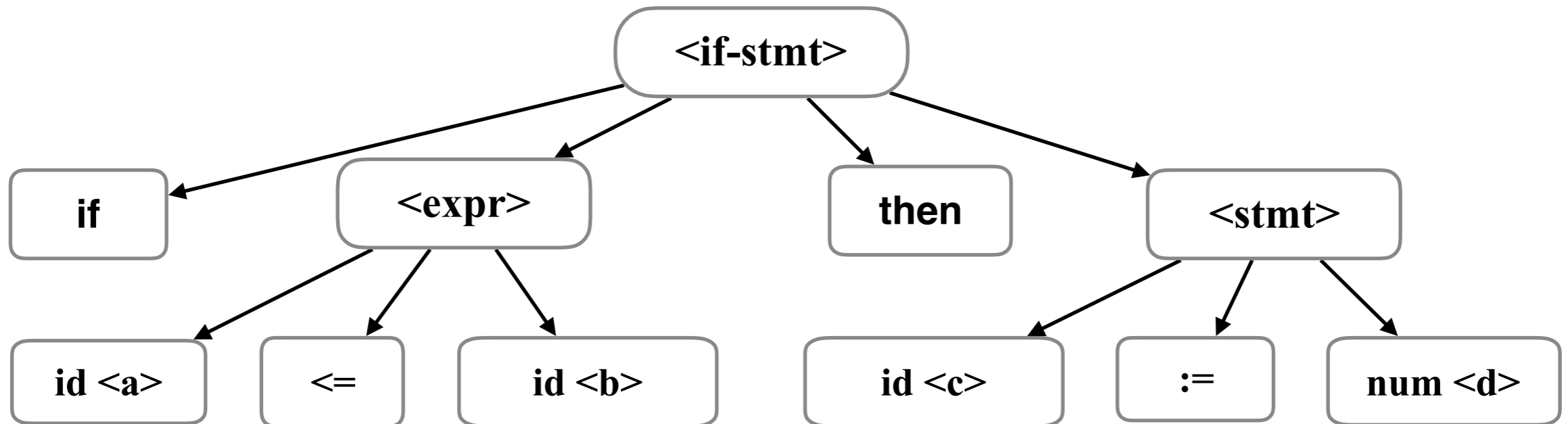
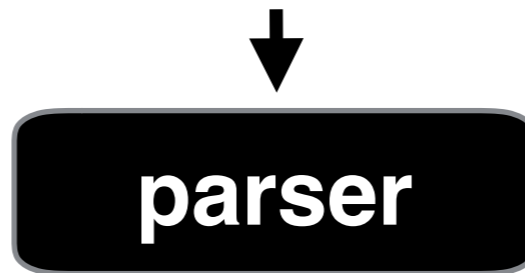
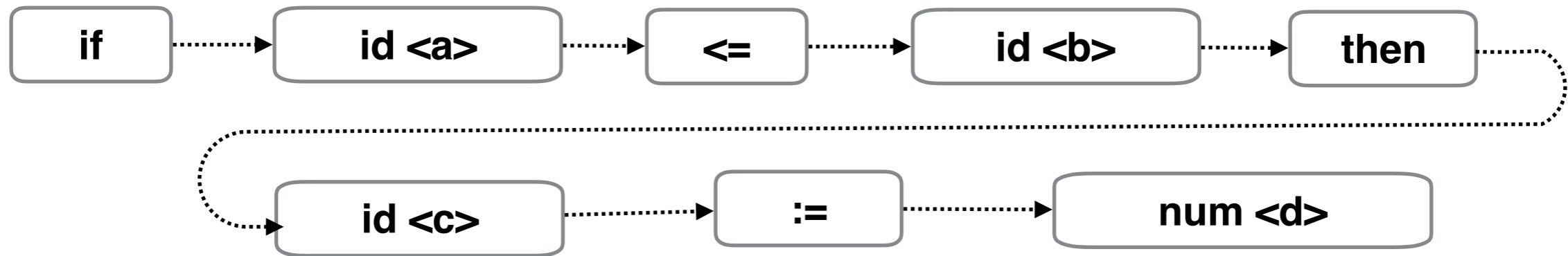
Front End Responsibilities:

- Recognize legal programs
- Report errors
- Produce IR
- Preliminary storage map
- Shape the code for the back end

Much of front end construction can be automated

Review: Parsing Example

Token Sequence



Parse Tree

Review: Context Free Grammars (CFGs)

- A formalism to for describing languages
- A CFG $G = \langle T, N, P, S \rangle$:
 1. A set T of terminal symbols (tokens).
 2. A set N of nonterminal symbols.
 3. A set P production (rewrite) rules.
 4. A special start symbol S .
- The language $L(G)$ is the set of sentences of terminal symbols in T^* that can be derived from the start symbol S :

$$L(G) = \{w \in T^* \mid S \Rightarrow^* w\}$$

An Example of a Partial Context Free Grammar

...

$\langle \text{if-stmt} \rangle ::= \text{if } \langle \text{expr} \rangle \text{ then } \langle \text{stmt} \rangle$
 $\langle \text{expr} \rangle ::= \text{id } \leq \text{id}$
 $\langle \text{stmt} \rangle ::= \text{id } := \text{num}$

...

Context free grammar

Rule 1 $\$1 \Rightarrow 1\&$
Rule 2 $\$0 \Rightarrow 0\$$
Rule 3 $\&1 \Rightarrow 1\$$
Rule 4 $\&0 \Rightarrow 0\&$
Rule 5 $\$\# \Rightarrow \rightarrow A$
Rule 6 $\&\# \Rightarrow \rightarrow B$

Not a context free grammar

CFGs are rewrite systems with restrictions on the form of rewrite (production) rules that can be used. The left hand side of a production rule can only be **one non-terminal symbol**.

Elements of BNF Syntax

Terminal Symbol:

Symbol-in-Boldface

Non-Terminal Symbol:

Symbol-in-Angle-Brackets

Production Rule:

Non-Terminal ::= Sequence of Symbols

or

Non-Terminal ::= Sequence | Sequence | ...

Example:

...

$\langle \text{if-stmt} \rangle ::= \mathbf{if} \langle \text{expr} \rangle \mathbf{then} \langle \text{stmt} \rangle$

$\langle \text{expr} \rangle ::= \mathbf{id} \langle \mathbf{=} \mathbf{id}$

$\langle \text{stmt} \rangle ::= \mathbf{id} \mathbf{:= num}$

How a BNF Grammar Describes a Language

A sentence is a sequence of terminal symbols (tokens).

The language $L(G)$ of a BNF grammar G is the set of sentences generated using the grammar:

- Begin with start symbol.
- Iteratively replace non-terminals with terminals according to the rules (rewrite system).

This replacing process is called a derivation (\Rightarrow).

Zero or multiple derivation steps are written as \Rightarrow^ .*

Formally, $L(G) = \{w \in T^* \mid S \Rightarrow^* w\}$

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, in another word, can $X2 := 0$ be derived in G ?

Start Symbol : $\langle \text{stmt} \rangle$



⋮



$X2 := 0$

Grammar G :

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

In which order to apply the rules?

Typically, there are three options:

leftmost (\Rightarrow_L), rightmost (\Rightarrow_R), any (\Rightarrow)

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	1
$X \langle \text{digit} \rangle := 0 \Rightarrow_L$	2
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	1
$X \langle \text{digit} \rangle := 0 \Rightarrow_L$	2
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	

$X2 := 0$	
-----------	--

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	

$X2 := 0$	
-----------	--

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	

$X2 := 0$	
-----------	--

Derivation in a Grammar (\mathcal{G})

Is $X2 := 0 \in L(\mathcal{G})$, i.e., can $X2 := 0$ be derived in \mathcal{G} ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_R	

$X2 := 0$	
-----------	--

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_R	

$X2 := 0$	
-----------	--

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_R	2
$\langle \text{identifier} \rangle 2 := 0$	\Rightarrow_R	

$X2 := 0$	
-----------	--

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_R	2
$\langle \text{identifier} \rangle 2 := 0$	\Rightarrow_R	

$X2 := 0$	
-----------	--

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_R	2
$\langle \text{identifier} \rangle 2 := 0$	\Rightarrow_R	3
$\langle \text{letter} \rangle 2 := 0$	\Rightarrow_R	
$X2 := 0$		

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_L	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_L	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_L	1
$X \langle \text{digit} \rangle := 0$	\Rightarrow_L	2
$X2 := 0$		

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation		rule
$\langle \text{stmt} \rangle$	\Rightarrow_R	6
$\langle \text{identifier} \rangle := 0$	\Rightarrow_R	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	\Rightarrow_R	2
$\langle \text{identifier} \rangle 2 := 0$	\Rightarrow_R	3
$\langle \text{letter} \rangle 2 := 0$	\Rightarrow_R	
$X2 := 0$		

Derivation in a Grammar (G)

Is $X2 := 0 \in L(G)$, i.e., can $X2 := 0$ be derived in G ?

leftmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_L$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_L$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	3
$\langle \text{letter} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_L$	1
$X \langle \text{digit} \rangle := 0 \Rightarrow_L$	2
$X2 := 0$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

rightmost derivation	rule
$\langle \text{stmt} \rangle \Rightarrow_R$	6
$\langle \text{identifier} \rangle := 0 \Rightarrow_R$	5
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0 \Rightarrow_R$	2
$\langle \text{identifier} \rangle 2 := 0 \Rightarrow_R$	3
$\langle \text{letter} \rangle 2 := 0 \Rightarrow_R$	1
$X2 := 0$	

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
	1
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
	3
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
	2
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$
- 6.

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$
- 6.

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
	5
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$
- 6.

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
$\langle \text{identifier} \rangle := 0$	5
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
$\langle \text{identifier} \rangle := 0$	5
$\langle \text{stmt} \rangle$	

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
$\langle \text{identifier} \rangle := 0$	5
$\langle \text{stmt} \rangle$	4

1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

Parsing of a Language $L(G)$

Can we recognize $X2 := 0$ as being in $L(G)$?

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
$\langle \text{identifier} \rangle := 0$	5
$\langle \text{stmt} \rangle$	4

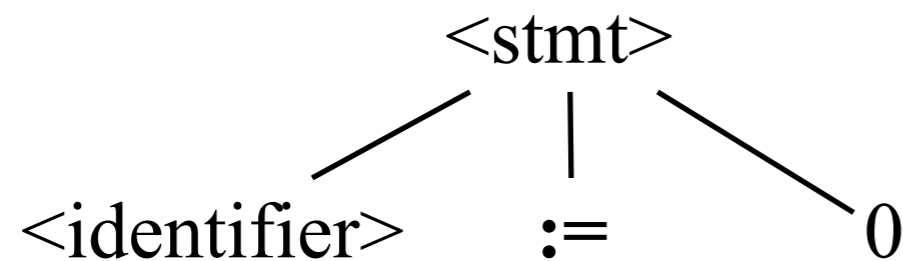
1. $\langle \text{letter} \rangle ::= A | B | C | \dots | Z$
2. $\langle \text{digit} \rangle ::= 0 | 1 | 2 | \dots | 9$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle |$
 $\langle \text{identifier} \rangle \langle \text{letter} \rangle |$
4. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
5. $\langle \text{identifier} \rangle := 0$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

We will talk about LL(1) grammars and an example parser for a small language (tinyL) that is implemented using mutually recursive procedures (recursive descent parser) in next class.

Parse Trees (in \mathcal{G})

*Each internal (non-leaf) node is a nonterminal;
its children are the RHS of a production for that NT.*

The parse tree demonstrates that the grammar generates the terminal string on the frontier.



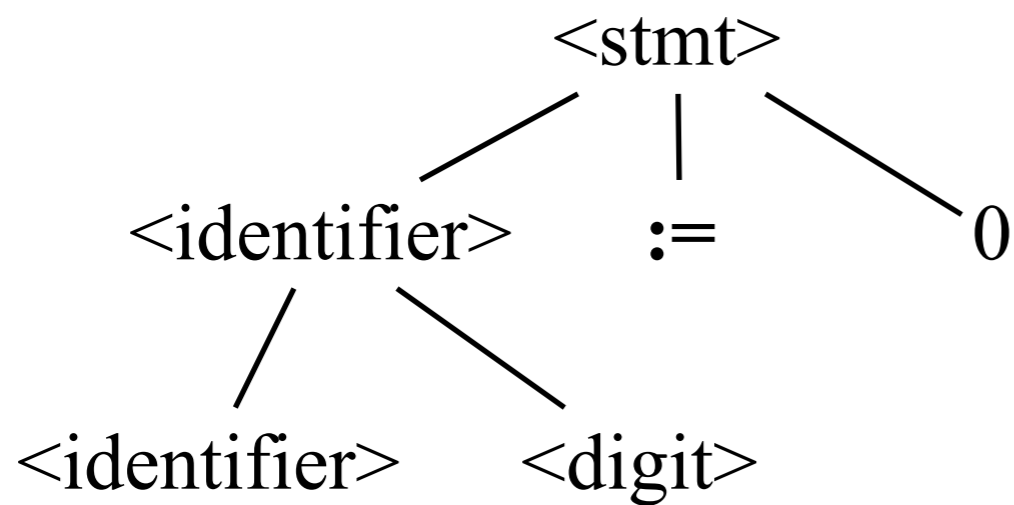
X2 := 0	Rule
<letter> 2 := 0	1
<identifier> 2 := 0	3
<identifier> <digit> := 0	2
<identifier> := 0	5
<stmt>	4

A parse tree of X2 := 0 in (\mathcal{G}) :

Parse Trees (in \mathcal{G})

*Each internal (non-leaf) node is a nonterminal;
its children are the RHS of a production for that NT.*

The parse tree demonstrates that the grammar generates the terminal string on the frontier.



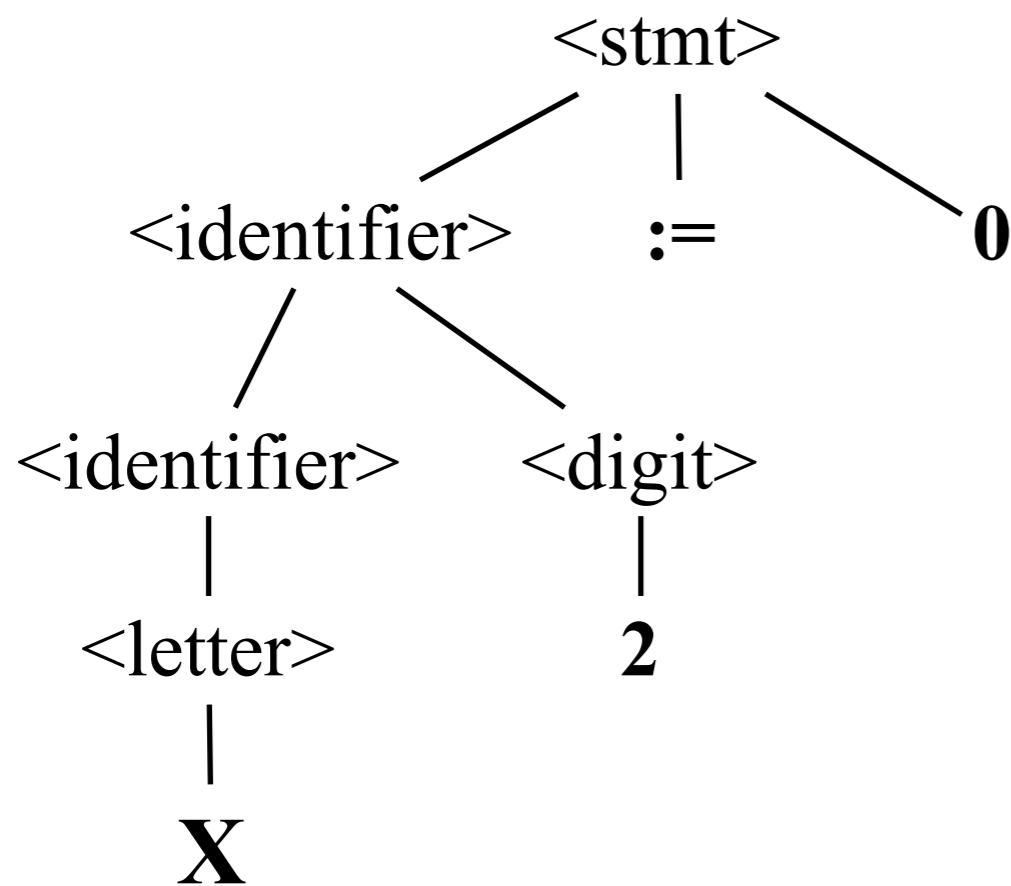
X2 := 0	Rule
<letter> 2 := 0	1
<identifier> 2 := 0	3
<identifier> <digit> := 0	2
<identifier> := 0	5
<stmt>	4

A parse tree of X2 := 0 in (\mathcal{G}) :

Parse Trees (in \mathcal{G})

*Each internal node is a nonterminal;
its children are the RHS of a production for that NT.*

The parse tree demonstrates that the grammar generates the terminal string on the frontier.



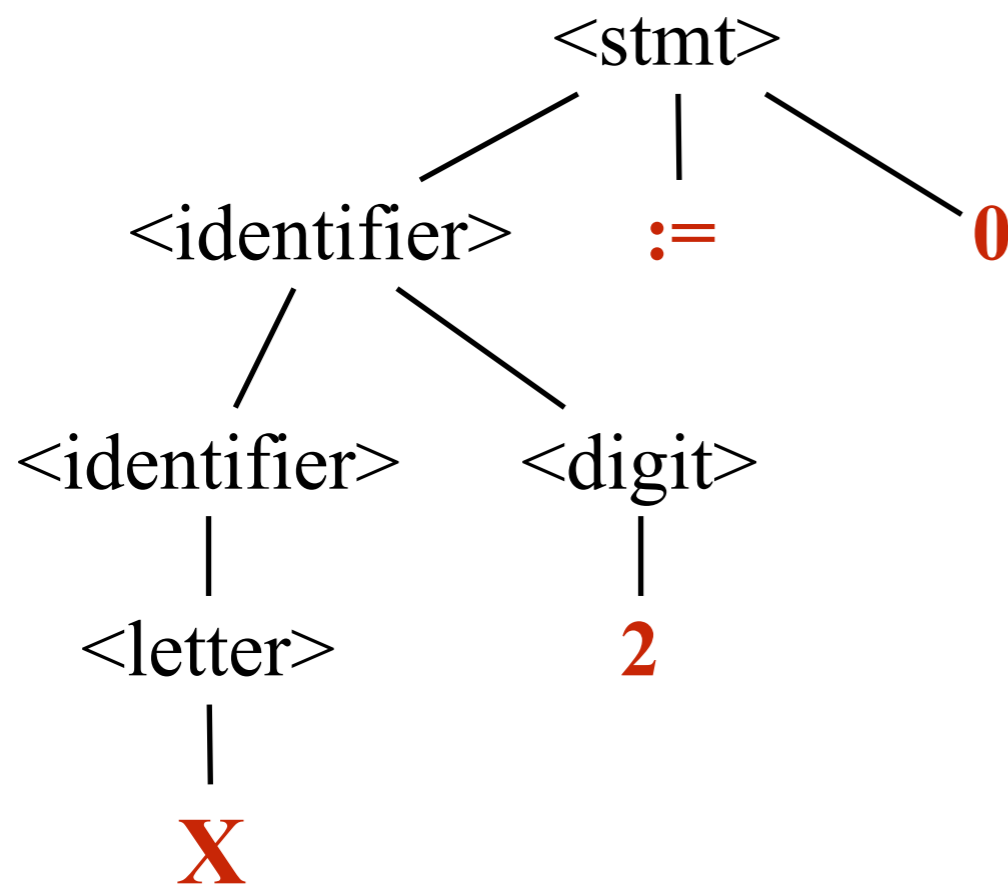
A parse tree of $X2 := 0$ in (\mathcal{G}) :

$X2 := 0$	Rule
$\langle \text{letter} \rangle 2 := 0$	1
$\langle \text{identifier} \rangle 2 := 0$	3
$\langle \text{identifier} \rangle \langle \text{digit} \rangle := 0$	2
$\langle \text{identifier} \rangle := 0$	5
$\langle \text{stmt} \rangle$	4

Parse Trees (in \mathcal{G})

*Each internal node is a nonterminal;
its children are the RHS of a production for that NT.*

The parse tree demonstrates that the grammar generates the terminal string on the frontier.



X2 := 0	Rule
<letter> 2 := 0	1
<identifier> 2 := 0	3
<identifier> <digit> := 0	2
<identifier> := 0	5
<stmt>	4

A parse tree of X2 := 0 in (\mathcal{G}) :

A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\qquad \qquad \qquad \langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

How about the parse tree of $X2 := 0$ in G' and G ?

A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

$\langle \text{stmt} \rangle$

$X2 := 0$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

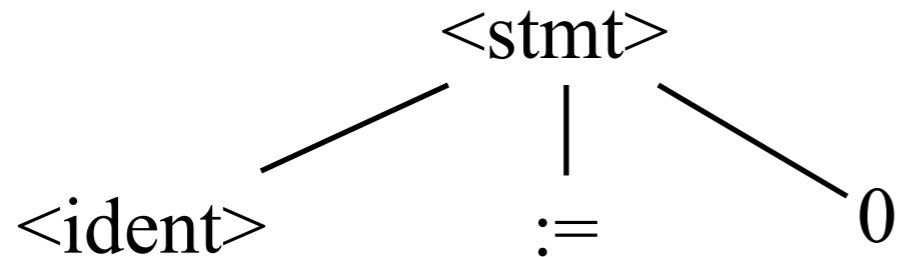
A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



X2 := 0

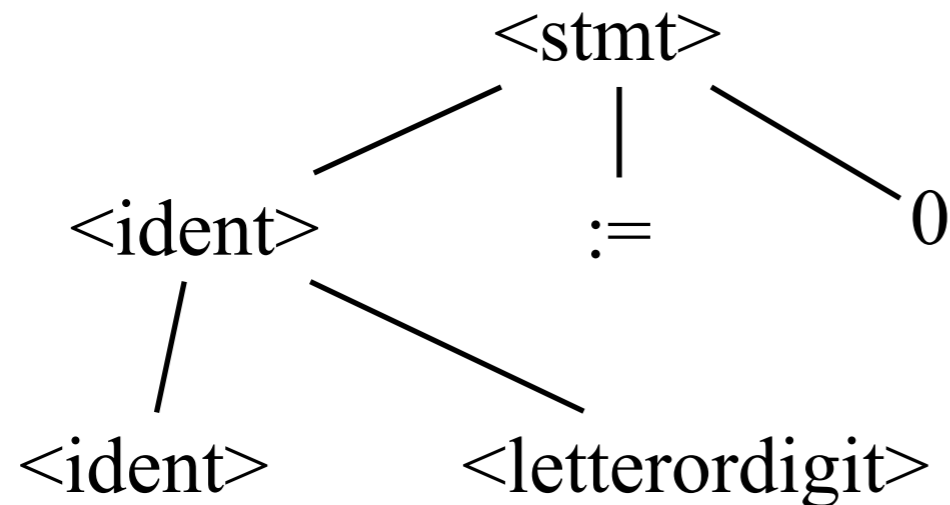
A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$

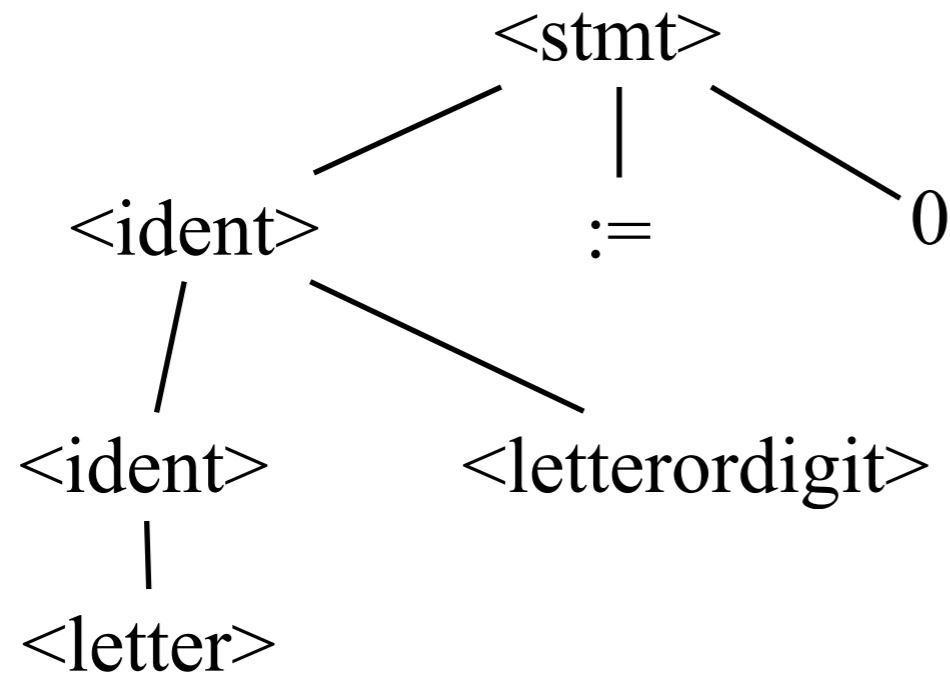
A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$

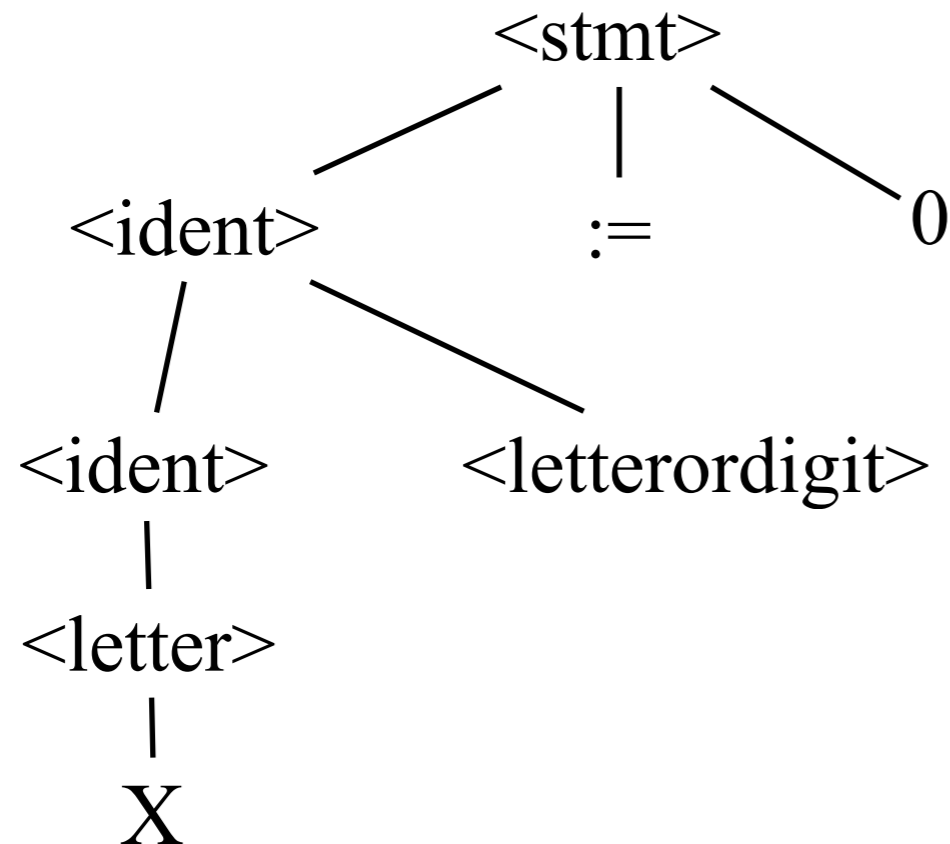
A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$

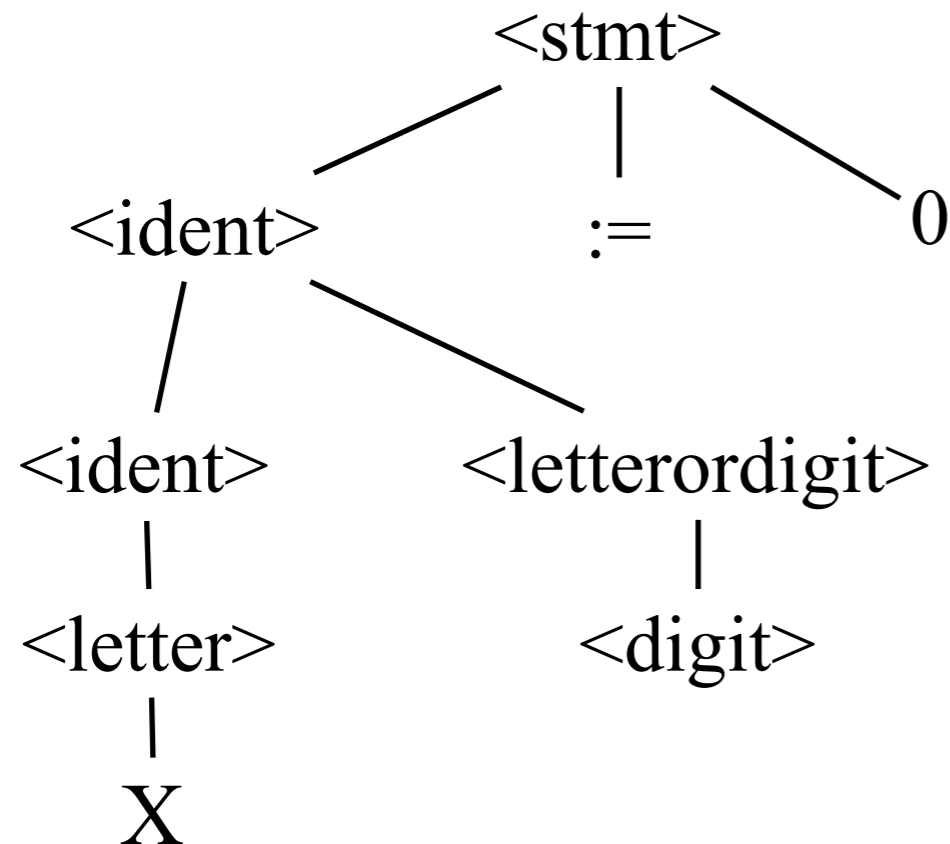
A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X := 0$

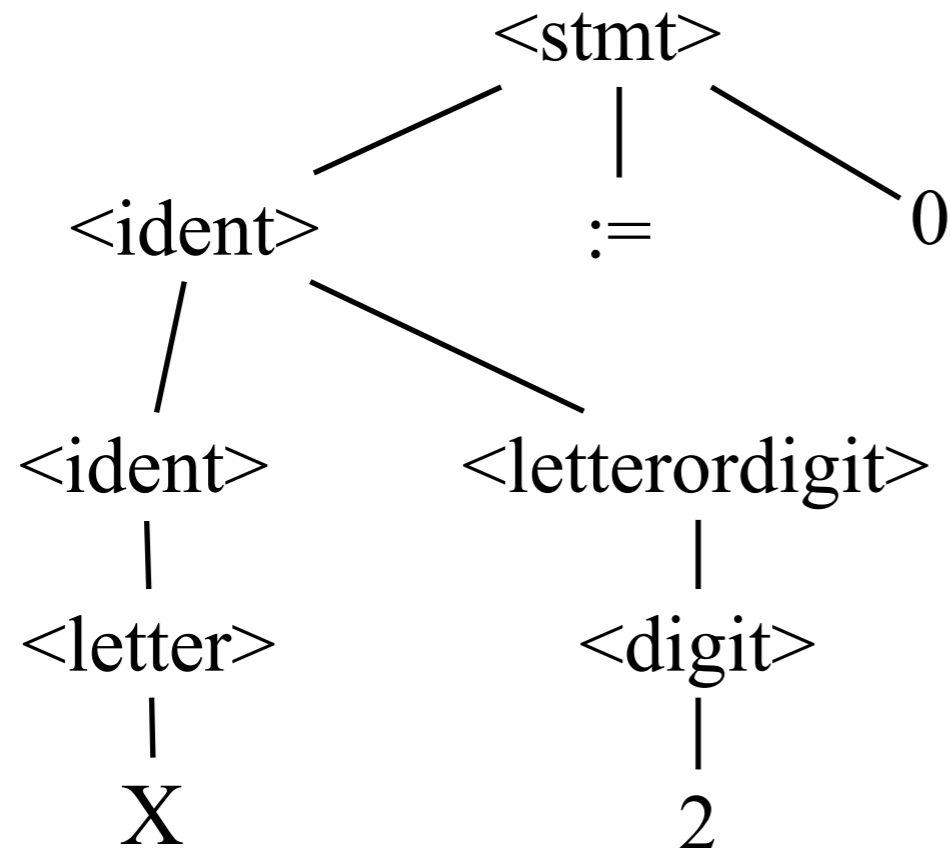
A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$

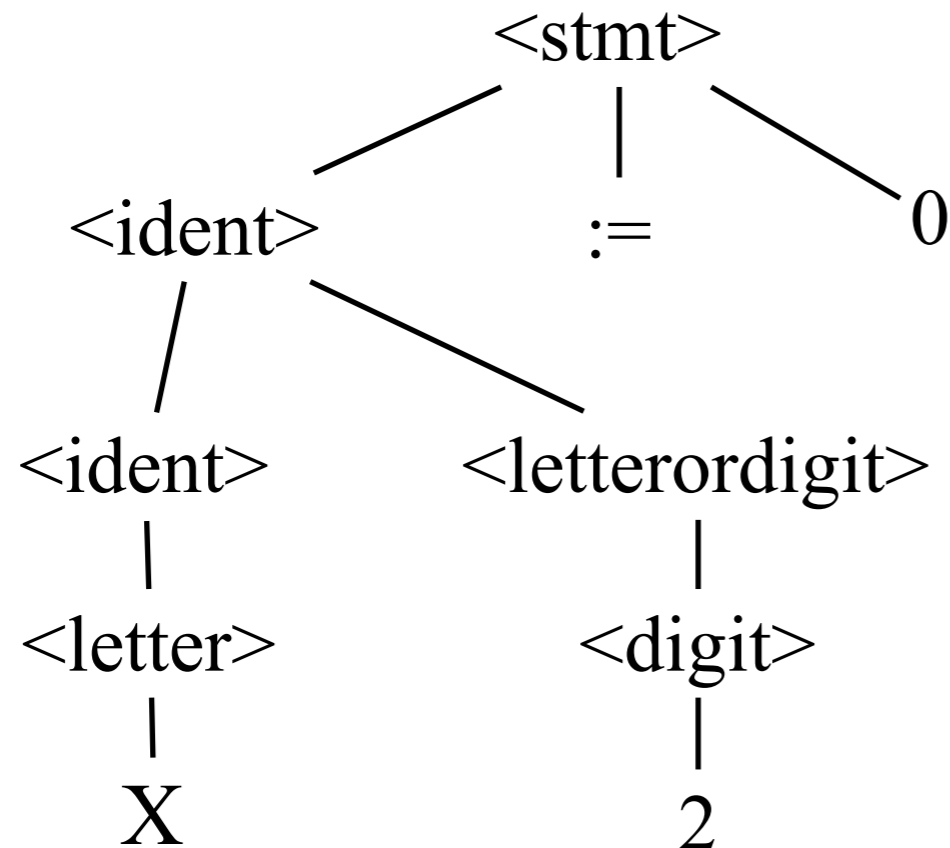
A Language May Have Many Grammars

Consider G' :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



X2 := 0

$\langle \text{stmt} \rangle$

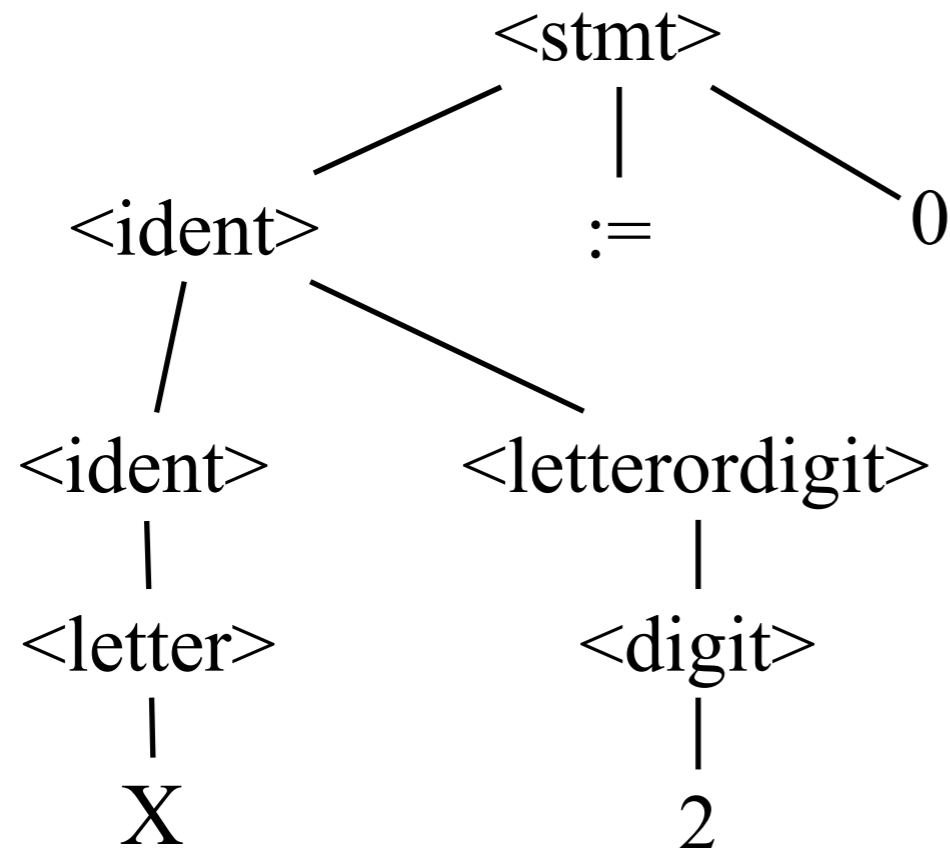
A Language May Have Many Grammars

Consider G' :

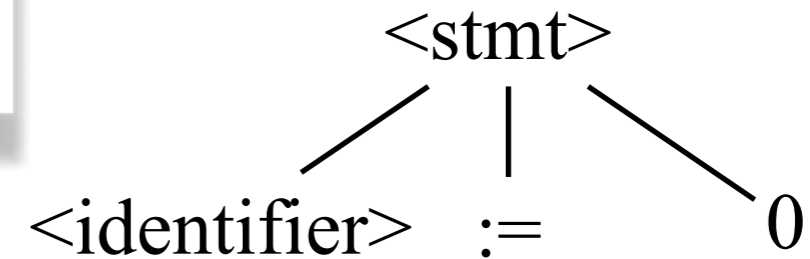
1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\quad \quad \quad \langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$



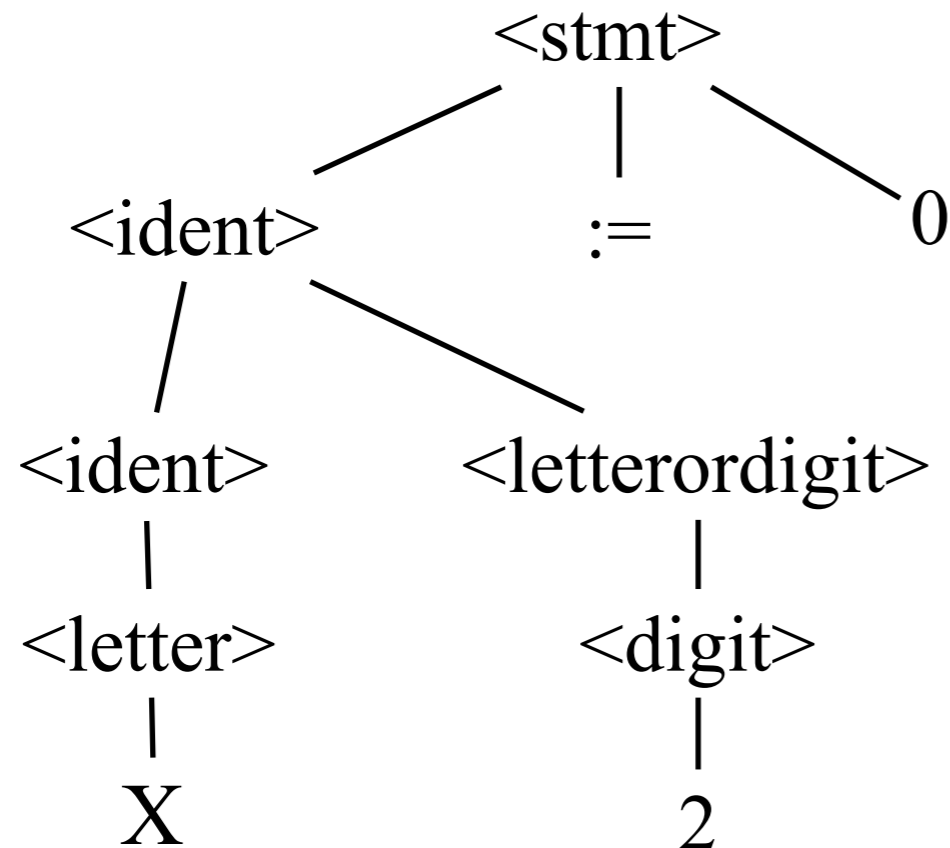
A Language May Have Many Grammars

Consider G' :

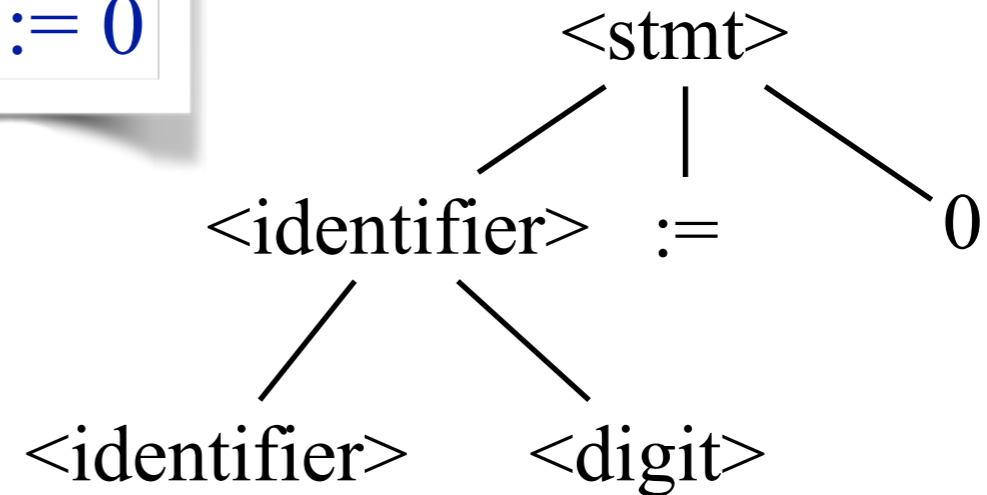
1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$



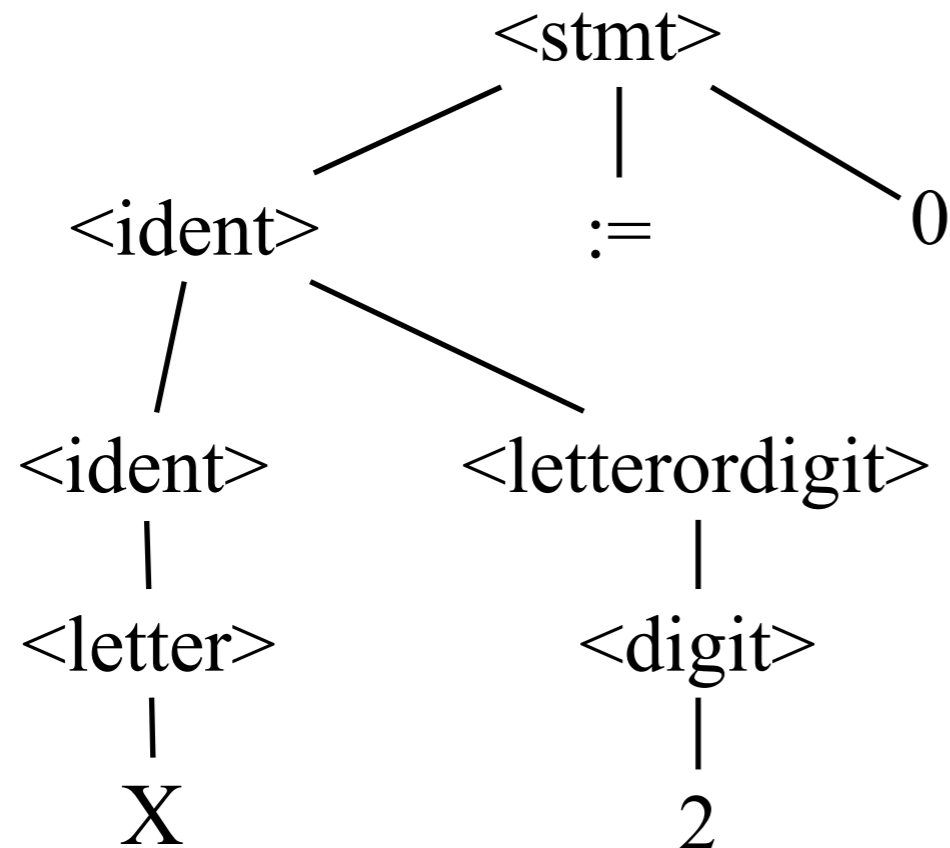
A Language May Have Many Grammars

Consider G' :

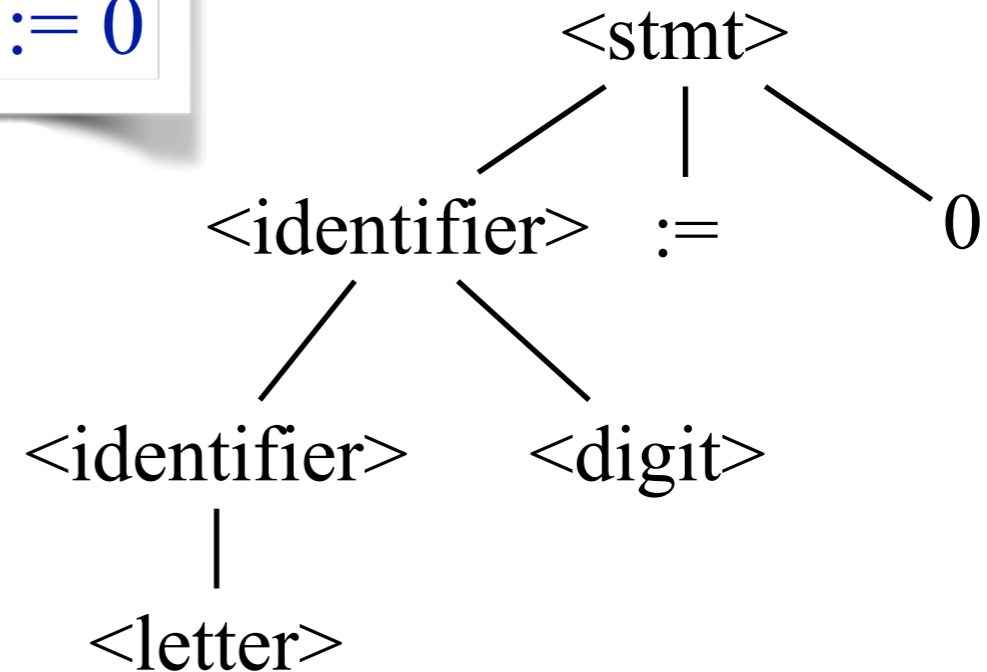
1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$



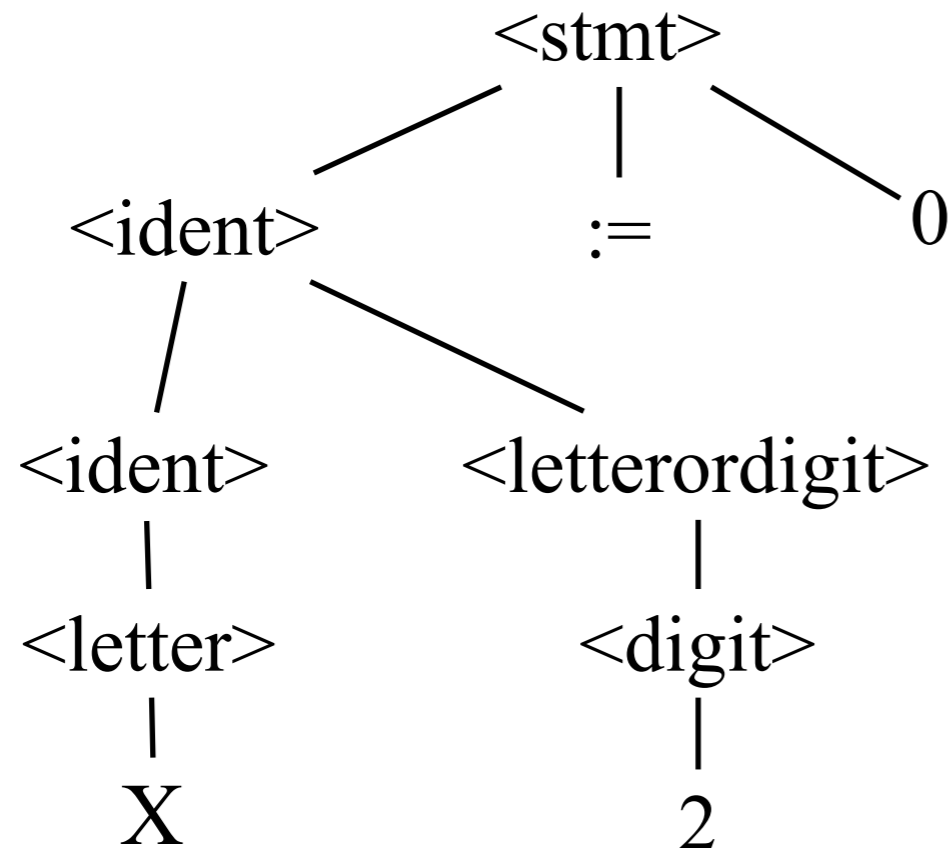
A Language May Have Many Grammars

Consider G' :

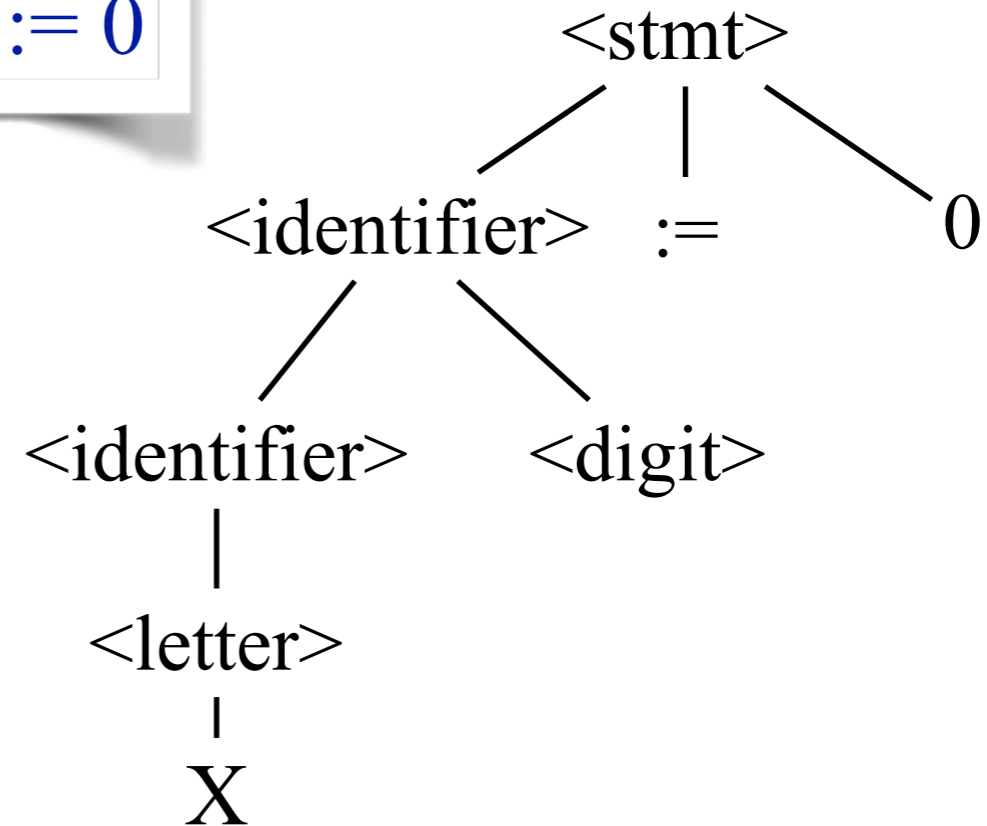
1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$



$X2 := 0$



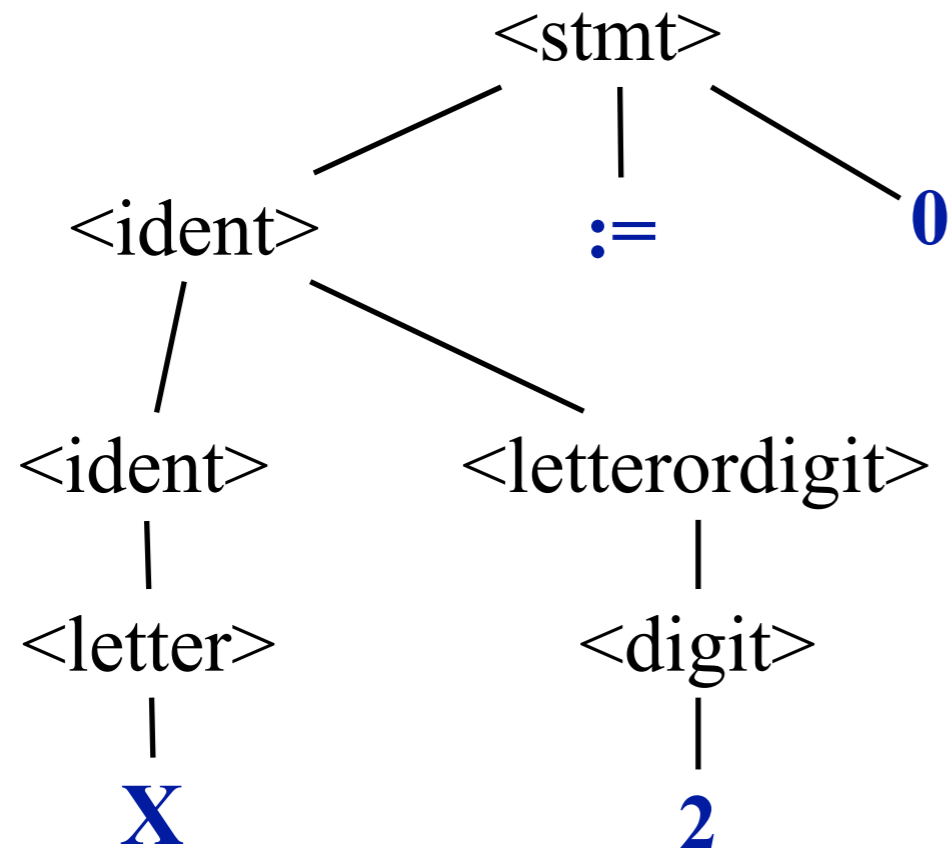
A Language May Have Many Grammars

Consider G' :

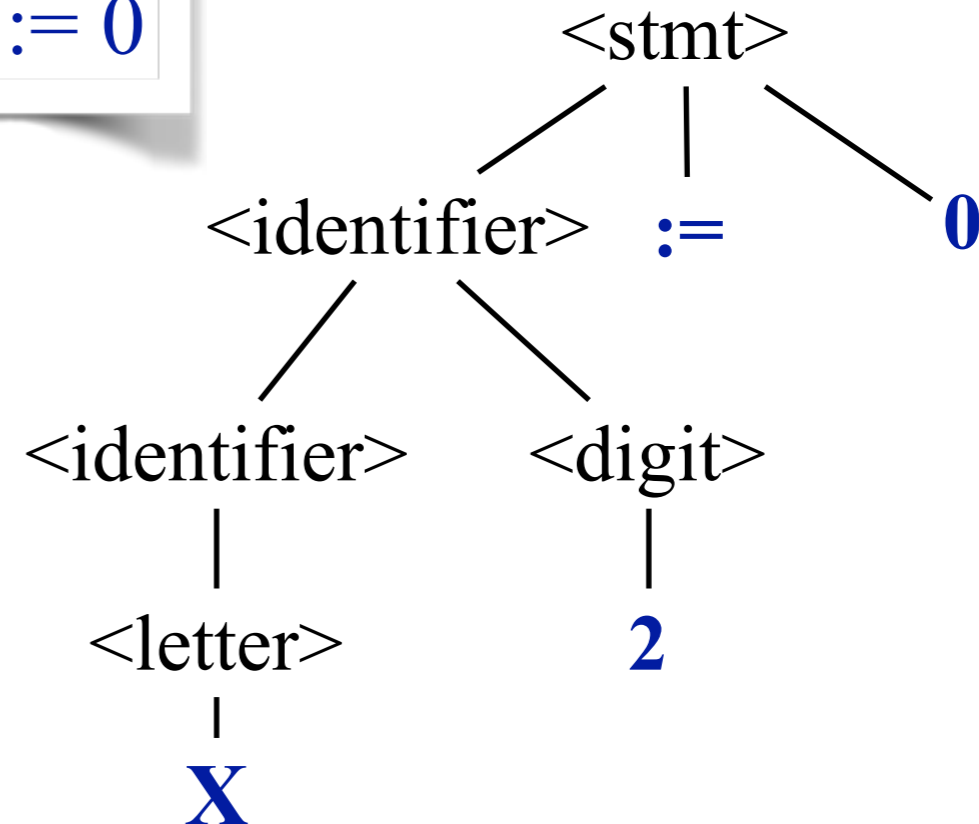
1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{ident} \rangle \langle \text{letterordigit} \rangle$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := \mathbf{0}$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := \mathbf{0}$



$X2 := 0$



A Language May Have Many Grammars

Consider G' :

The Original Grammar G :

1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$	1. $\langle \text{letter} \rangle ::= \mathbf{A} \mid \mathbf{B} \mid \mathbf{C} \mid \dots \mid \mathbf{Z}$
2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$	2. $\langle \text{digit} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \dots \mid \mathbf{9}$
3. $\langle \text{ident} \rangle ::= \langle \text{letter} \rangle \mid$	3. $\langle \text{identifier} \rangle ::= \langle \text{letter} \rangle \mid$
4. $\langle \text{identifier} \rangle \langle \text{letterordigit} \rangle$	4. $\langle \text{identifier} \rangle \langle \text{letter} \rangle \mid$
5. $\langle \text{stmt} \rangle ::= \langle \text{ident} \rangle := 0$	5. $\langle \text{identifier} \rangle \langle \text{digit} \rangle$
6. $\langle \text{letterordigit} \rangle ::= \langle \text{letter} \rangle \mid \langle \text{digit} \rangle$	6. $\langle \text{stmt} \rangle ::= \langle \text{identifier} \rangle := 0$

How about the parse tree of $X2 := 0$ in G and G' ?

G and G' generate the same language, but yield different parse trees.

Grammars and Programming Languages

Many grammars may correspond to one programming language.

Good grammars:

- Captures the logic structure of the language
⇒ structure carries some semantic information
(example: expression grammar)
- Use meaningful names
- Are easy to read
- Are unambiguous
- ...

What's problem with ambiguity?

Ambiguous Grammars

“Time flies like an arrow; fruit flies like a banana.”



A grammar G is ambiguous iff there exists a $w \in L(G)$ such that there are:

- two distinct parse trees for w , or
- two distinct leftmost derivations for w , or
- two distinct rightmost derivations for w .

We want a unique semantics of our programs, which typically requires a unique syntactic structure.

Simple Statement Grammar

How are nested if statements parsed with this grammar?

if x == 0 then if y == 0 then z := 1 else z := 2

<start> ::= <stmt>

<stmt> ::= <if-stmt> | <assgn>

**<if-stmt> ::= if <expr> then <stmt> |
 if <expr> then <stmt> else <stmt>**

<assgn> ::= <id> := <d>

<expr> ::= <id> == 0

<d> ::= 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

<id> ::= a | b | c | ... | z |

Simple Statement Grammar

How are nested if statements parsed with this grammar?

if x == 0 **then**[**if** y == 0 **then** z := 1 **else** z := 2]

<start> ::= <stmt>

<stmt> ::= <if-stmt> | <assgn>

<if-stmt> ::= **if** <expr> **then** <stmt> |
 if <expr> **then** <stmt> **else** <stmt>

<assgn> ::= <id> := <d>

<expr> ::= <id> == 0

<d> ::= **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9**

<id> ::= **a** | **b** | **c** | ... | **z** |

Simple Statement Grammar

How are nested if statements parsed with this grammar?

if x == 0 **then**[**if** y == 0 **then** z := 1] **else** z := 2

<start> ::= <stmt>

<stmt> ::= <if-stmt> | <assgn>

<if-stmt> ::= **if** <expr> **then** <stmt> |
 if <expr> **then** <stmt> **else** <stmt>

<assgn> ::= <id> := <d>

<expr> ::= <id> == 0

<d> ::= **0** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9**

<id> ::= **a** | **b** | **c** | ... | **z** |

Dangling Else Ambiguity

if $x = 0$ then if $y = 0$ then $z := 1$ else $z := 2$

How to deal with ambiguity?

- Change the language to include delimiters
Example: Adding new terminal symbols.

It can fix the *dangling else, expression* grammars.

Changing the Language to Include Delimiters

Algol 68 if statement:

$\langle \text{if-stmt} \rangle ::= \text{if } \langle \text{expr} \rangle \text{ then } \langle \text{stmt} \rangle \text{ fi} \mid$
 $\text{if } \langle \text{expr} \rangle \text{ then } \langle \text{stmt} \rangle \text{ else } \langle \text{stmt} \rangle \text{ fi}$

How would you use this syntax to express the meaning of the two different parse trees for:

if x = 0 then if y = 0 then z := 1 else z := 2



This: **if x = 0 then if y = 0 then z := 1 fi else z := 2 fi**

Or that: **if x = 0 then if y = 0 then z := 1 else z := 2 fi fi**

Dangling Else Ambiguity

if $x = 0$ then if $y = 0$ then $z := 1$ else $z := 2$

How to deal with ambiguity?

- Change the language to include delimiters
Example: Adding new terminal symbols.

It can fix the *dangling else, expression* grammars.

Dangling Else Ambiguity

if $x = 0$ **then** **if** $y = 0$ **then** $z := 1$ **else** $z := 2$

How to deal with ambiguity?

- Change the language to include delimiters
Example: Adding new terminal symbols.

It can fix the *dangling else, expression* grammars.

- Change the grammar
Example: Impose **associativity** and **precedence** in an *arithmetic expression* grammar.

Arithmetic Expression Grammar

Parse “8 - 3 * 2”:

$$\begin{aligned} \langle \text{start} \rangle &::= \langle \text{expr} \rangle \\ \langle \text{expr} \rangle &::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid \\ &\quad \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid \\ &\quad \langle \text{expr} \rangle * \langle \text{expr} \rangle \mid \\ &\quad \langle \text{expr} \rangle / \langle \text{expr} \rangle \mid \\ &\quad \langle \text{expr} \rangle ^ \langle \text{expr} \rangle \mid \\ &\quad \langle \text{d} \rangle \mid \langle \text{l} \rangle \\ \langle \text{d} \rangle &::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9} \\ \langle \text{l} \rangle &::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z} \end{aligned}$$

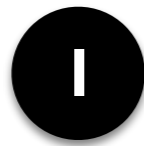
Possible Parse Trees

Parse “8 - 3 * 2”:

< expr >

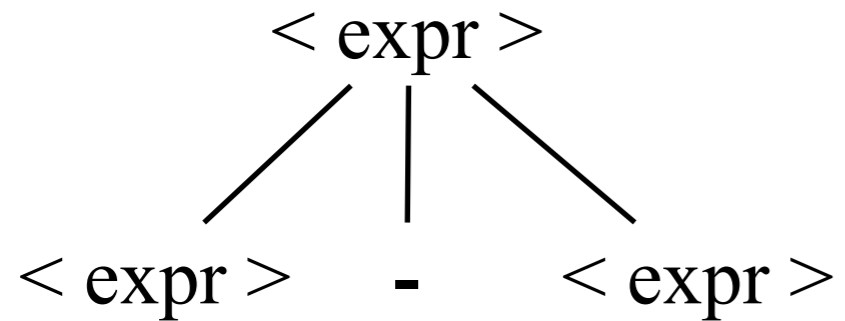
```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
              < expr > - < expr > |
              < expr > * < expr > |
              < expr > / < expr > |
              < expr > ^ < expr > |
              < d > | < l >
< d >      ::= 0 | 1 | 2 | 3 | ... | 9
< l >      ::= a | b | c | ... | z
```

Parse Tree



Possible Parse Trees

Parse “8 - 3 * 2”:



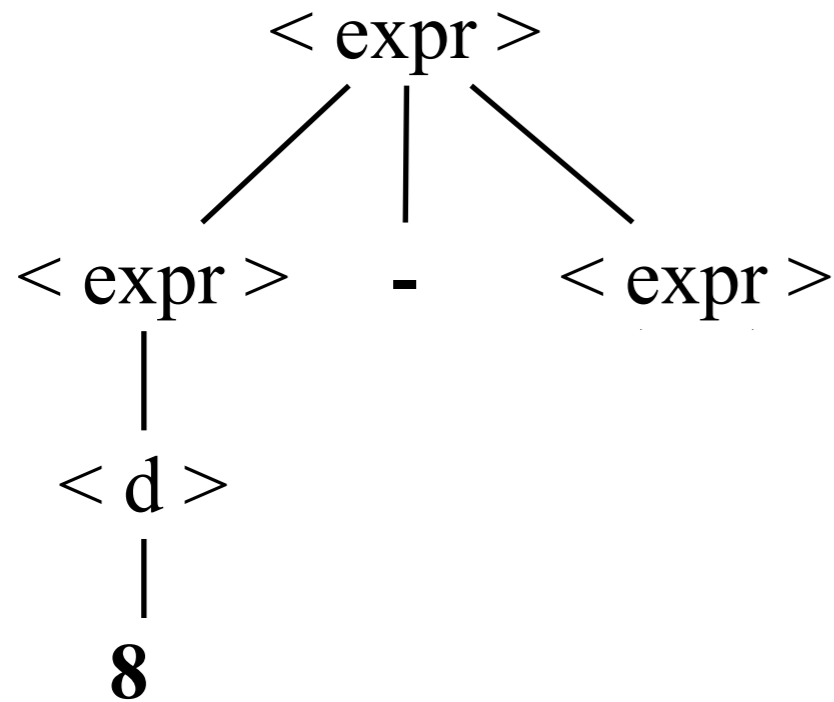
```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
              < expr > - < expr > |
              < expr > * < expr > |
              < expr > / < expr > |
              < expr > ^ < expr > |
              < d > | < l >
< d >      ::= 0 | 1 | 2 | 3 | ... | 9
< l >      ::= a | b | c | ... | z
```

Parse Tree



Possible Parse Trees

Parse “8 - 3 * 2”:



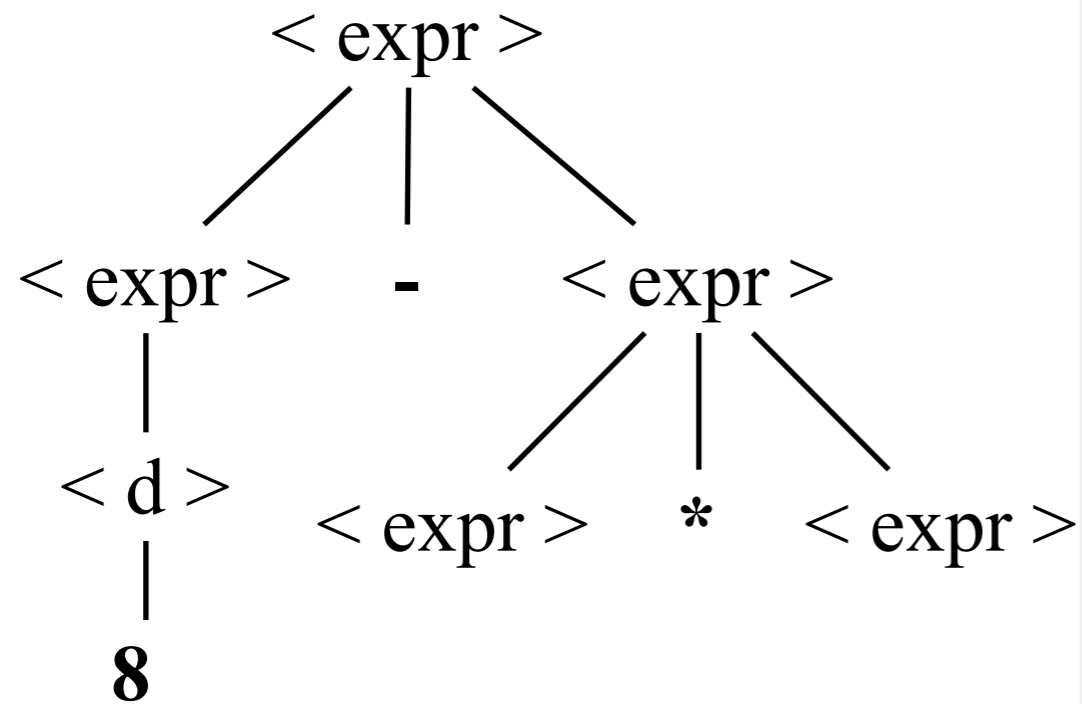
$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$
 $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle * \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle / \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle ^ \langle \text{expr} \rangle \mid$
 $\quad \langle d \rangle \mid \langle l \rangle$
 $\langle d \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$
 $\langle l \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Parse Tree



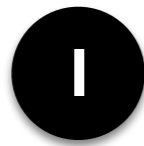
Possible Parse Trees

Parse “8 - 3 * 2”:



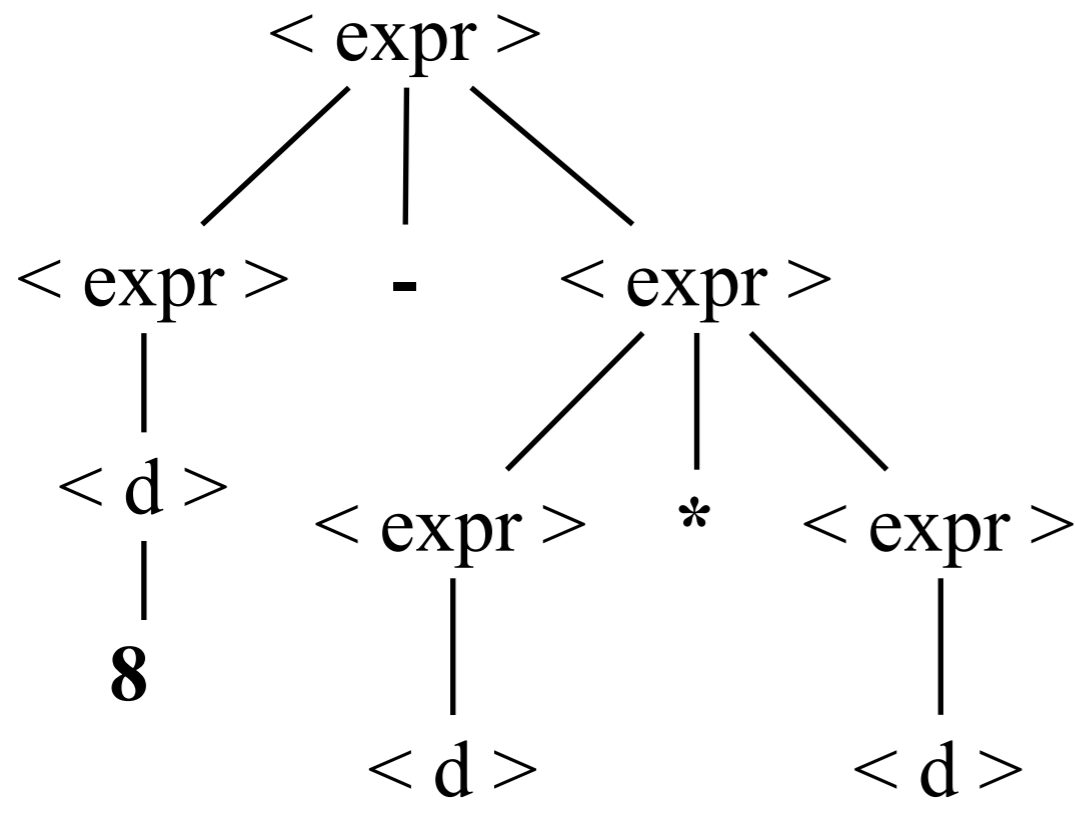
```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
              < expr > - < expr > |
              < expr > * < expr > |
              < expr > / < expr > |
              < expr > ^ < expr > |
              < d > | < l >
< d >      ::= 0 | 1 | 2 | 3 | ... | 9
< l >      ::= a | b | c | ... | z
```

Parse Tree



Possible Parse Trees

Parse “8 - 3 * 2”:



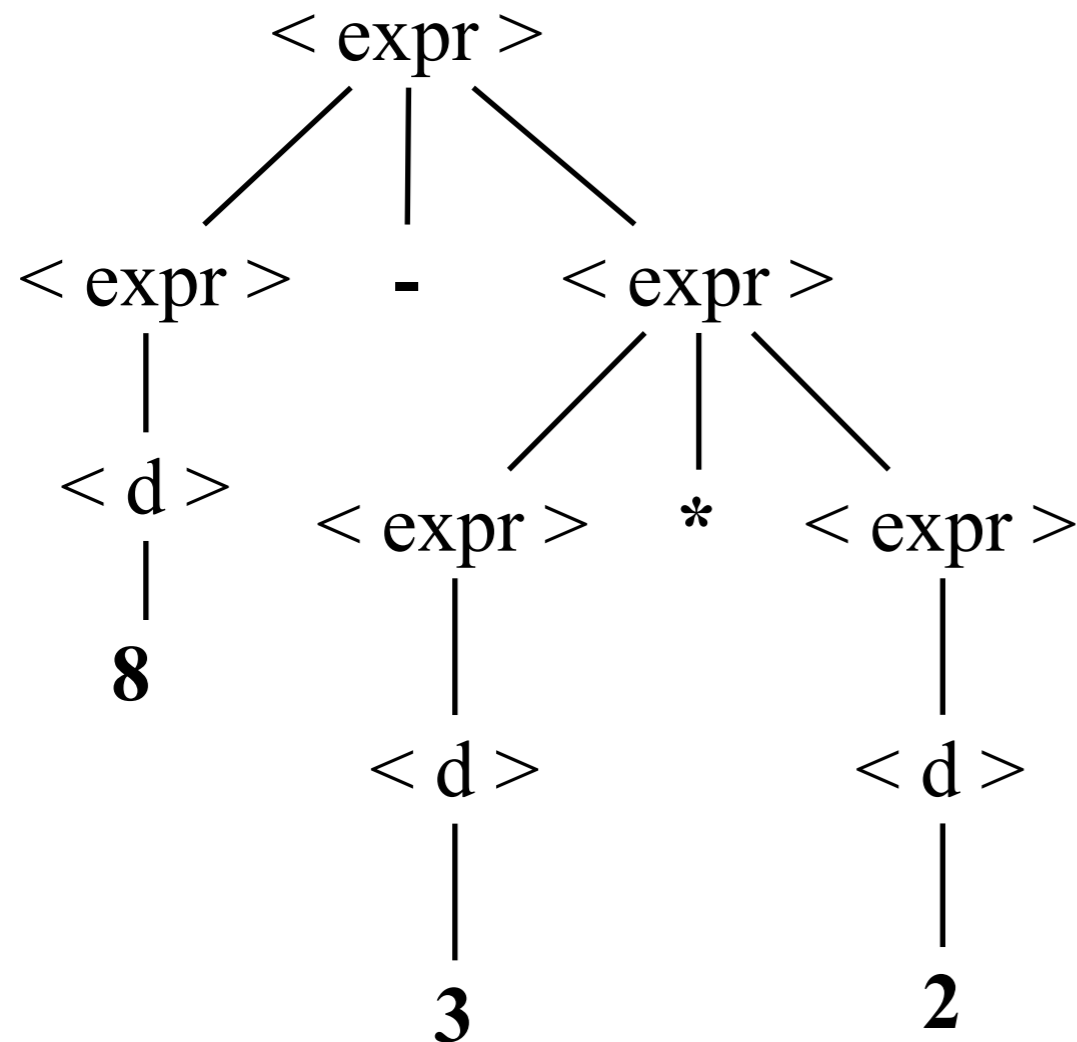
```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
               < expr > - < expr > |
               < expr > * < expr > |
               < expr > / < expr > |
               < expr > ^ < expr > |
               < d > | < l >
< d >      ::= 0 | 1 | 2 | 3 | ... | 9
< l >      ::= a | b | c | ... | z
```

Parse Tree



Possible Parse Trees

Parse “8 - 3 * 2”:



Parse Tree



```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
              < expr > - < expr > |
              < expr > * < expr > |
              < expr > / < expr > |
              < expr > ^ < expr > |
              < d > | < 1 >
< d >      ::= 0 | 1 | 2 | 3 | ... | 9
< 1 >      ::= a | b | c | ... | z
```

Possible Parse Trees

Parse “8 - 3 * 2”:

< expr >

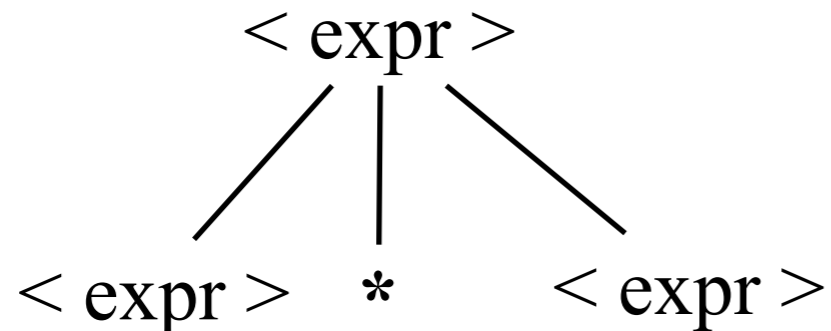
```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
             < expr > - < expr > |
             < expr > * < expr > |
             < expr > / < expr > |
             < expr > ^ < expr > |
             < d > | < l >
< d >      ::= 0 | 1 | 2 | 3 | ... | 9
< l >      ::= a | b | c | ... | z
```

Parse Tree

II

Possible Parse Trees

Parse “8 - 3 * 2”:



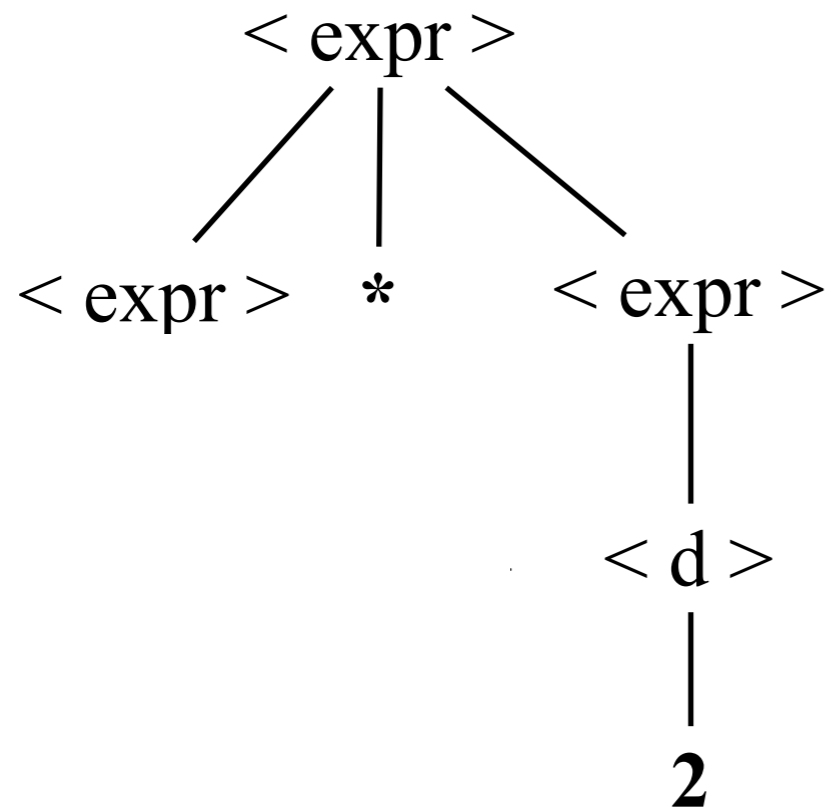
$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$
 $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle * \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle / \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle ^ \langle \text{expr} \rangle \mid$
 $\quad \langle \text{d} \rangle \mid \langle \text{l} \rangle$
 $\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$
 $\langle \text{l} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Parse Tree

II

Possible Parse Trees

Parse “8 - 3 * 2”:



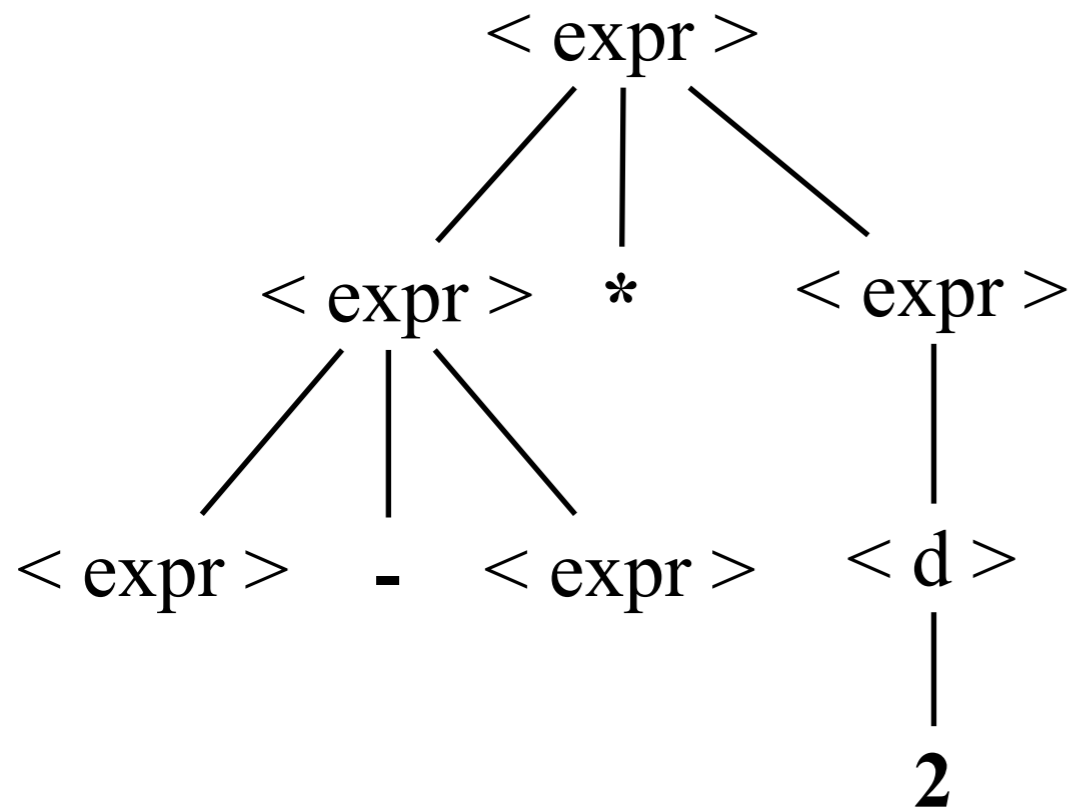
$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$
 $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle * \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle / \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle ^ \langle \text{expr} \rangle \mid$
 $\quad \langle \text{d} \rangle \mid \langle \text{l} \rangle$
 $\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$
 $\langle \text{l} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Parse Tree

II

Possible Parse Trees

Parse “8 - 3 * 2”:



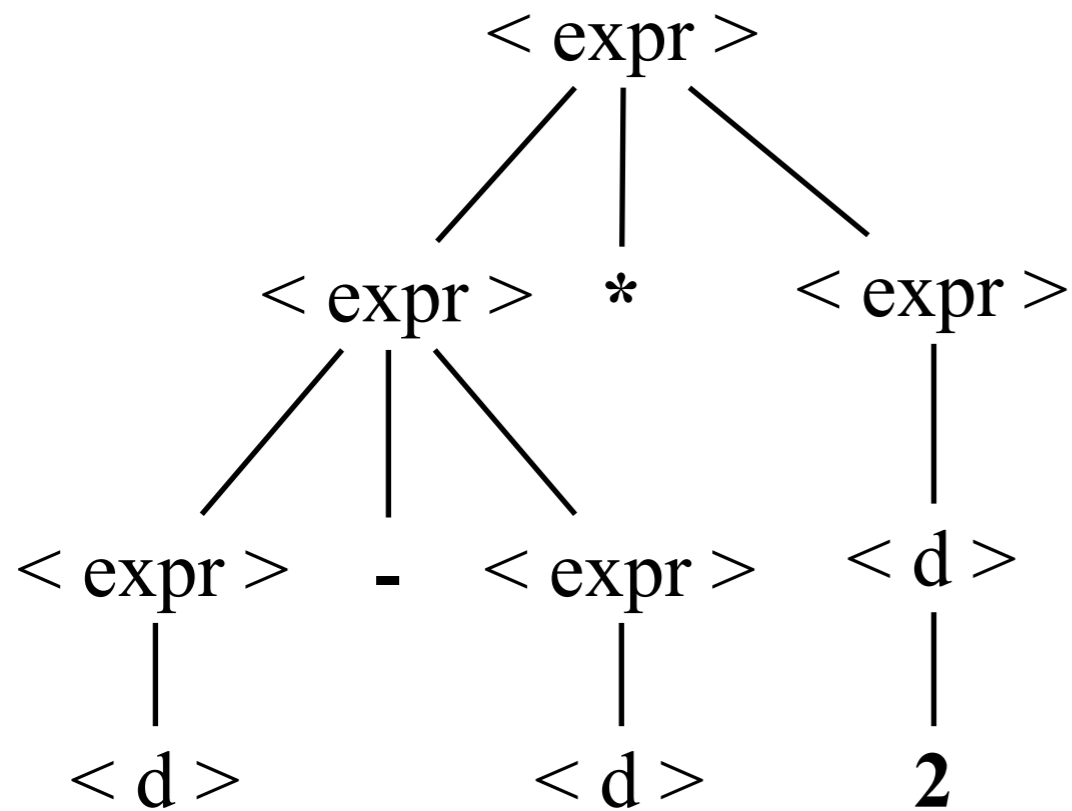
$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$
 $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$
 $\langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$
 $\langle \text{expr} \rangle * \langle \text{expr} \rangle \mid$
 $\langle \text{expr} \rangle / \langle \text{expr} \rangle \mid$
 $\langle \text{expr} \rangle ^ \langle \text{expr} \rangle \mid$
 $\langle \text{d} \rangle \mid \langle 1 \rangle$
 $\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$
 $\langle 1 \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Parse Tree

II

Possible Parse Trees

Parse “8 - 3 * 2”:



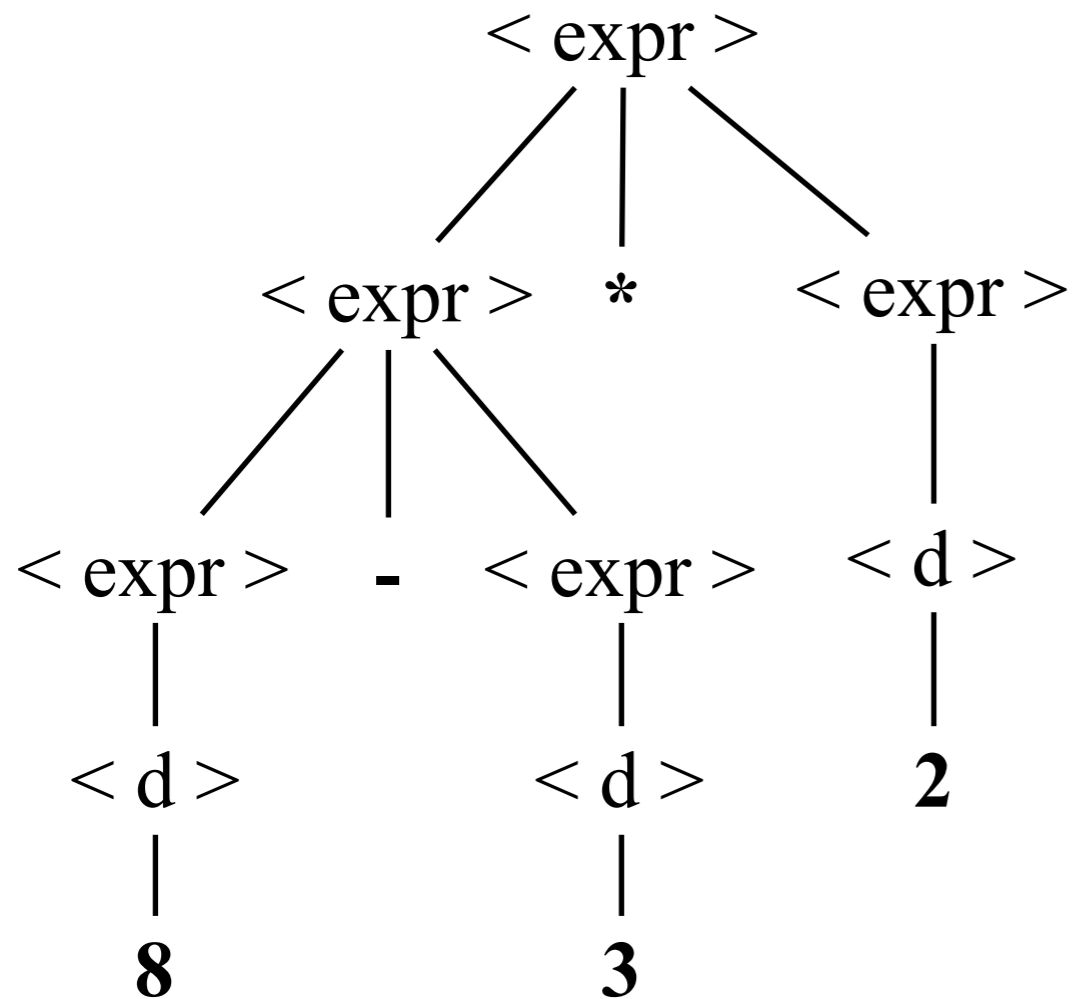
```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
              < expr > - < expr > |
              < expr > * < expr > |
              < expr > / < expr > |
              < expr > ^ < expr > |
              < d > | < 1 >
< d >      ::= 0 | 1 | 2 | 3 | ... | 9
< 1 >      ::= a | b | c | ... | z
```

Parse Tree

II

Possible Parse Trees

Parse “8 - 3 * 2”:



$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$
 $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle * \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle / \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle ^ \langle \text{expr} \rangle \mid$
 $\quad \langle \text{d} \rangle \mid \langle \text{l} \rangle$
 $\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$
 $\langle \text{l} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Parse Tree

II

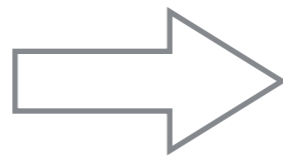
Changing the Language to Include Delimiters

$$\begin{aligned} \langle \text{expr} \rangle ::= & \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid \\ & \langle \text{expr} \rangle * \langle \text{expr} \rangle \mid \\ & \langle l \rangle \mid \langle d \rangle \end{aligned}$$

Changing the Language to Include Delimiters

$$\langle \text{expr} \rangle ::= (\langle \text{expr} \rangle) - (\langle \text{expr} \rangle) \mid$$
$$(\langle \text{expr} \rangle) * (\langle \text{expr} \rangle) \mid$$
$$\langle l \rangle \mid \langle d \rangle$$

“8 - 3 * 2”



(8) - ((3) * (2))

OR

((8) - (3)) * (2)

Pretty ugly, isn't it?

Is there any other way to disambiguate our expression grammar?

Changing the Grammar to Impose Precedence

$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$
 $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$
 $\quad \langle \text{term} \rangle$
 $\langle \text{term} \rangle ::= \langle \text{term} \rangle * \langle \text{term} \rangle \mid$
 $\quad \langle \text{term} \rangle / \langle \text{term} \rangle \mid$
 $\quad \langle d \rangle \mid \langle 1 \rangle$
 $\langle d \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$
 $\langle 1 \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Modified Grammar G'

$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$
 $\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle * \langle \text{expr} \rangle \mid$
 $\quad \langle \text{expr} \rangle / \langle \text{expr} \rangle \mid$
 $\quad \langle d \rangle \mid \langle 1 \rangle$
 $\langle d \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$
 $\langle 1 \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Original Grammar G

Grouping in Parse Tree Now Reflects Precedence

Parse “8 - 3 * 2”:

< expr >

< start > ::= < expr >

< expr > ::= < expr > + < expr > |

< expr > - < expr > |

< term >

< term > ::= < term > * < term > |

< term > / < term > |

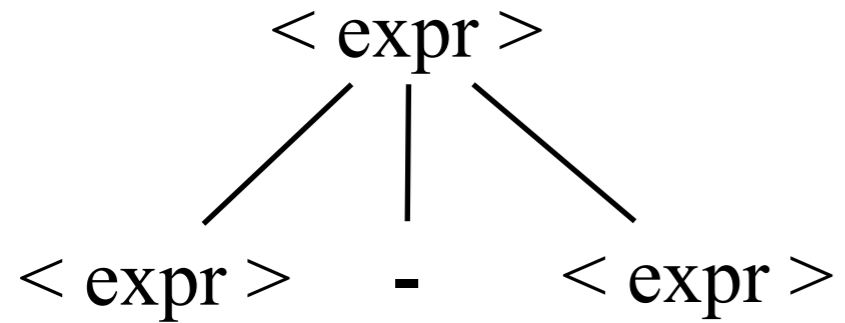
< d > | < l >

< d > ::= **0** | **1** | **2** | **3** | ... | **9**

< l > ::= **a** | **b** | **c** | ... | **z**

Grouping in Parse Tree Now Reflects Precedence

Parse “8 - 3 * 2”:



$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$

$\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$

$\langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$

$\langle \text{term} \rangle$

$\langle \text{term} \rangle ::= \langle \text{term} \rangle * \langle \text{term} \rangle \mid$

$\langle \text{term} \rangle / \langle \text{term} \rangle \mid$

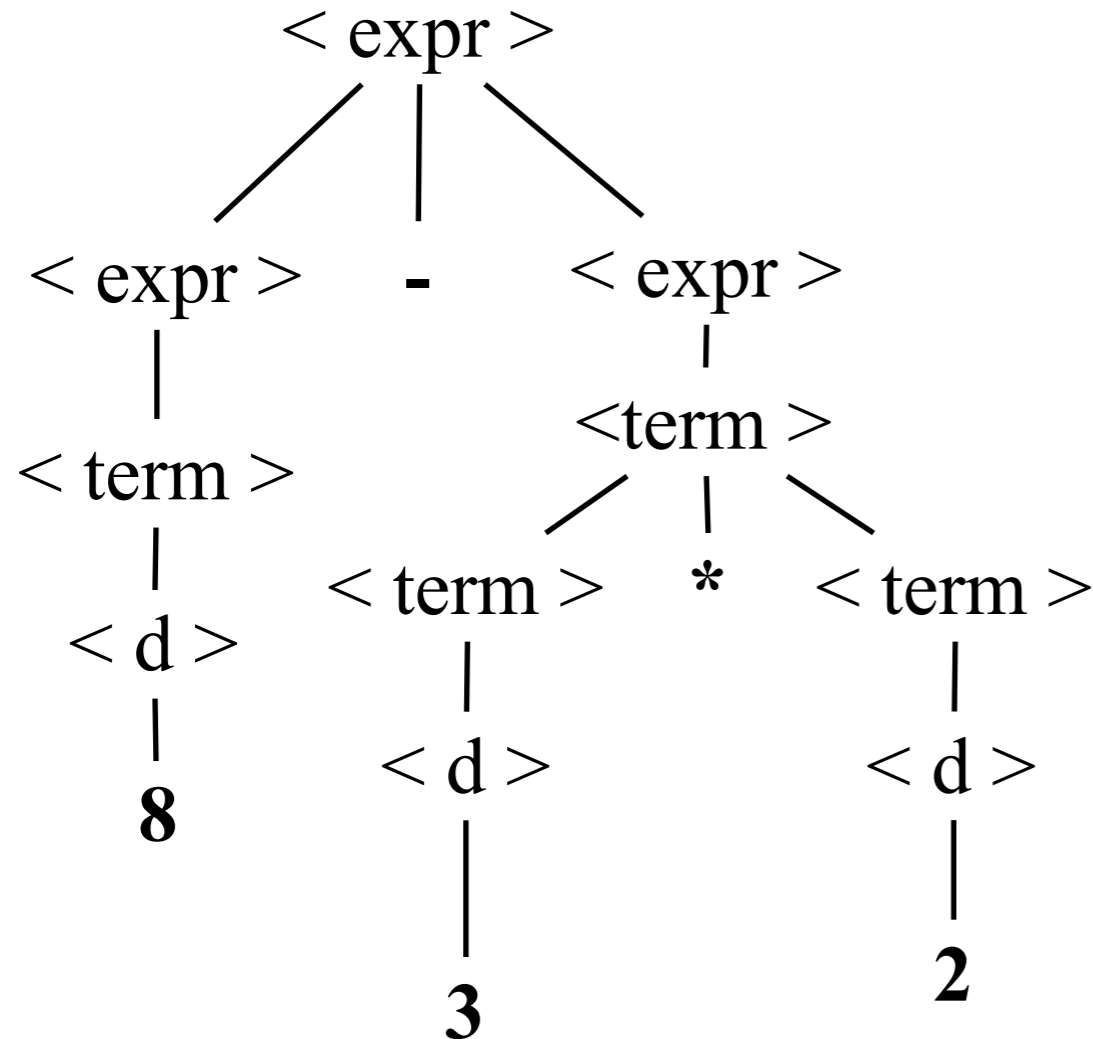
$\langle \text{d} \rangle \mid \langle \text{l} \rangle$

$\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$

$\langle \text{l} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Grouping in Parse Tree Now Reflects Precedence

Parse “8 - 3 * 2”:



< start > ::= < expr >

< expr > ::= < expr > + < expr > |

< expr > - < expr > |

< term >

< term > ::= < term > * < term > |

< term > / < term > |

< d > | < 1 >

< d > ::= **0** | **1** | **2** | **3** | ... | **9**

< 1 > ::= **a** | **b** | **c** | ... | **z**

Only One Possible Parse Tree

Precedence

- *Low Precedence:*
Addition + and Subtraction -
- *Medium Precedence:*
Multiplication * and Division /
- *Highest Precedence:*
Exponentiation ^

$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$

$\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid$

$\langle \text{expr} \rangle - \langle \text{expr} \rangle \mid$

$\langle \text{term} \rangle$

$\langle \text{term} \rangle ::= \langle \text{term} \rangle * \langle \text{term} \rangle \mid$

$\langle \text{term} \rangle / \langle \text{term} \rangle \mid$

$\langle \text{d} \rangle \mid \langle \text{l} \rangle$

$\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$

$\langle \text{l} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Still Have Ambiguity...

3 - 2 - 1 still a problem:

$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$

$\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid \langle \text{term} \rangle$

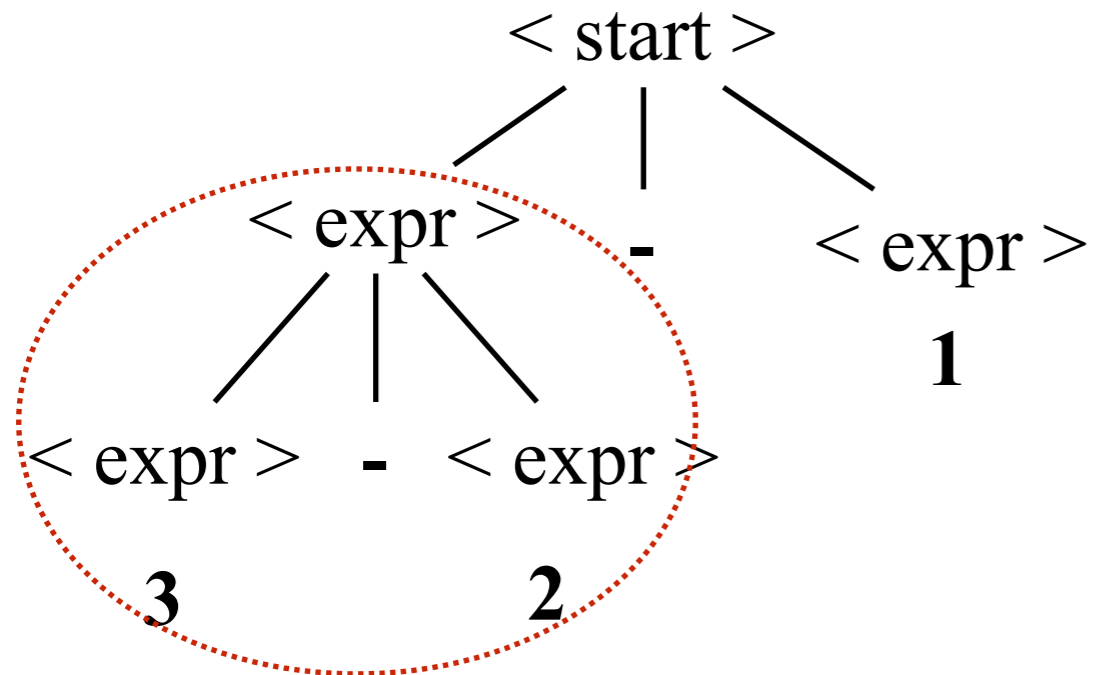
$\langle \text{term} \rangle ::= \langle \text{term} \rangle * \langle \text{term} \rangle \mid \langle \text{term} \rangle / \langle \text{term} \rangle \mid \langle \text{d} \rangle \mid \langle \text{l} \rangle$

$\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$

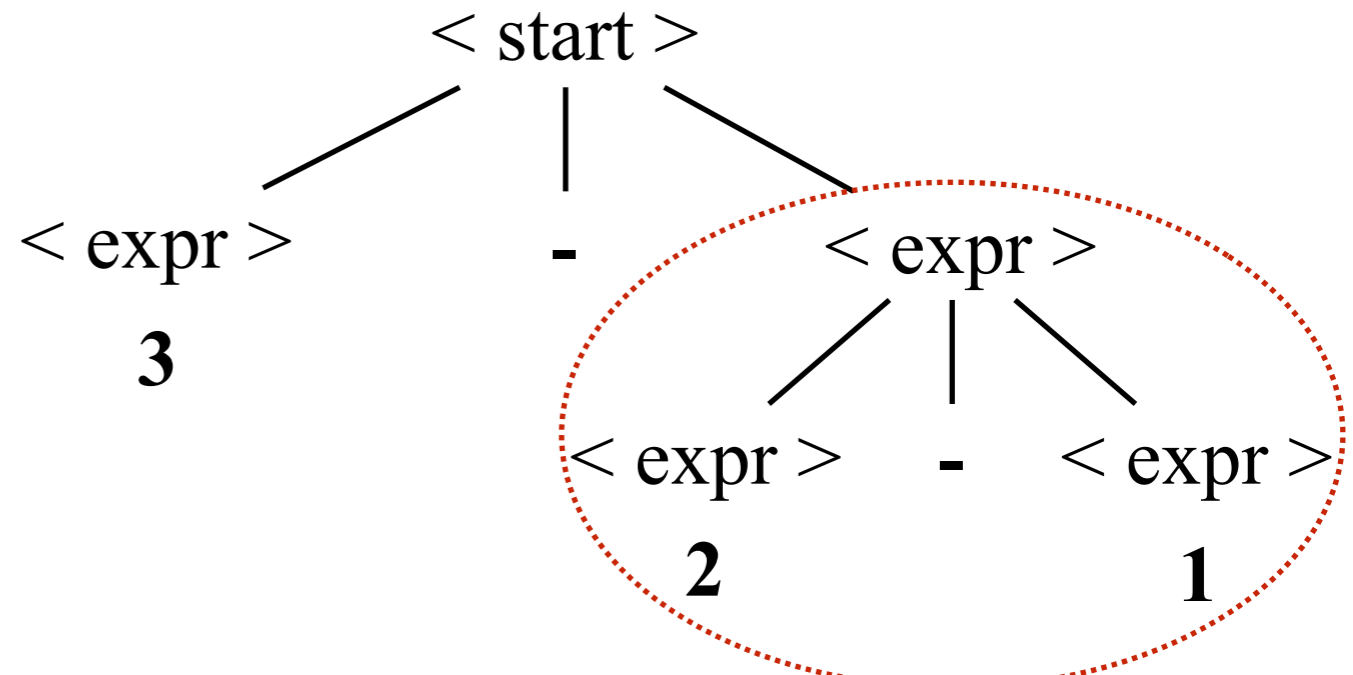
$\langle \text{l} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

Still Have Ambiguity...

3 - 2 - 1 still a problem:



Eval: $(3 - 2) - 1 = 0$



Eval: $3 - (2 - 1) = 2$

< start > ::= < expr >

< expr > ::= < expr > + < expr > | **< expr > - < expr >** | < term >

< term > ::= < term > * < term > | < term > / < term > | < d > | < 1 >

< d > ::= **0** | **1** | **2** | **3** | ... | **9**

< 1 > ::= **a** | **b** | **c** | ... | **z**

Still Have Ambiguity...

3 - 2 - 1 still a problem:

$\langle \text{start} \rangle ::= \langle \text{expr} \rangle$

$\langle \text{expr} \rangle ::= \langle \text{expr} \rangle + \langle \text{expr} \rangle \mid \langle \text{expr} \rangle - \langle \text{expr} \rangle \mid \langle \text{term} \rangle$

$\langle \text{term} \rangle ::= \langle \text{term} \rangle * \langle \text{term} \rangle \mid \langle \text{term} \rangle / \langle \text{term} \rangle \mid \langle \text{d} \rangle \mid \langle \text{l} \rangle$

$\langle \text{d} \rangle ::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9}$

$\langle \text{l} \rangle ::= \mathbf{a} \mid \mathbf{b} \mid \mathbf{c} \mid \dots \mid \mathbf{z}$

- Grouping of operators of same precedence not disambiguated.
- Non-commutative operators: only one parse tree correct.

Imposing Associativity

Same grammar with left / right recursion for - :

Our choices:

$$\begin{aligned} \langle \text{expr} \rangle &::= \langle d \rangle - \langle \text{expr} \rangle \mid \\ &\quad \langle d \rangle \\ \langle d \rangle &::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9} \end{aligned}$$

Or:

$$\begin{aligned} \langle \text{expr} \rangle &::= \langle \text{expr} \rangle - \langle d \rangle \mid \\ &\quad \langle d \rangle \\ \langle d \rangle &::= \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \dots \mid \mathbf{9} \end{aligned}$$

Which one do we want for
3 - 2 - 1 ?

Associativity

- Deals with operators of same precedence
- Implicit grouping or parenthesizing
- Left to right: $*$, $/$, $+$, $-$
- Right to left: $^$

Complete, Unambiguous Arithmetic Expression Grammar

```
< start > ::= < expr >
< expr > ::= < expr > + < expr > |
           < expr > - < expr > |
           < expr > * < expr > |
           < expr > / < expr > |
           < expr > ^ < expr > |
           < d > | < l >
< d >    ::= 0 | 1 | 2 | 3 | ... | 9
< l >    ::= a | b | c | ... | z
```

Original Ambiguous Grammar \mathcal{G}

```
< start > ::= < expr >
< expr > ::= < expr > + < term > |
           < expr > - < term > |
           < term >
< term > ::= < term > * < factor > |
           < term > / < factor > |
           < factor >
< factor > ::= < g > ^ < factor > |
            < g >
< g >     ::= ( < expr > ) | < d > | < l >
< d >     ::= 0 | 1 | 2 | ... | 9
< l >     ::= a | b | c | ... | z
```

Unambiguous Grammar \mathcal{G}

Dealing with Ambiguity

Some facts:

- Can't always remove an ambiguity from a grammar by restructuring productions.
- An inherently ambiguous language does not possess an unambiguous grammar.
- Detecting ambiguity in context-free grammars is an undecidable problem. There is no polynomial-time algorithm that can examine an arbitrary context - free grammar and tell if it is ambiguous.

Next Lecture

Things to do:

- Read Scott, Chapter 2.3 - 2.5 (skip 2.3.3 bottom - up Parsing)