CS 4300: Compiler Theory

Chapter 5
Syntax-Directed Translation

Dr. Xuejun Liang

Quick Review of Last Two Lectures

- Parser Generator Yacc and Bison
  - Yacc Specification
  - Writing a Grammar in Yacc
  - Dealing With Ambiguous Grammars
  - Resolve Parsing Action Conflicts
  - Combining Lex/Flex with Yacc/Bison
  - Error Recovery in Yacc
- Programming project III
- Review of homework assignments #6 and #7

Outlines (Sections)

1. Syntax-Directed Definitions
2. Evaluation Orders for SDD's
3. Applications of Syntax-Directed Definition
4. Syntax-Directed Translation Schemes
5. Implementing L-Attributed SDD's

1. Syntax-directed Definition

- A syntax-directed definition (SDD) specifies the values of attributes by associating semantic rules with the grammar productions

PRODUCTION	SEMANTIC RULE
$E \rightarrow E_1 + T$	$E.code = E_1.code \parallel T.code \parallel '+'$

- A syntax-directed translation scheme embeds program fragments called semantic actions within production bodies

$$E \rightarrow E_1 + T \quad \{ \text{ print } '+' \ \}$$

- Between the two notations
  - syntax-directed definitions can be more readable, and hence more useful for specifications.
  - However, translation schemes can be more efficient, and hence more useful for implementations

Attributes

- A synthesized attribute at node N is defined only in terms of attribute values at the children of N and at N itself.
- An inherited attribute at node N is defined only in terms of attribute values at N's parent , N itself, and N's siblings
- Attribute values typically represent
  - Numbers (literal constants)
  - Strings (literal constants)
  - Memory locations, such as a frame index of a local variable or function argument
  - A data type for type checking of expressions
  - Scoping information for local declarations
  - Intermediate program representations

Example Syntax-directed Definition

A simple desk calculator

Production	Semantic Rule
$L \to E\ \mathbf{n}$	$L.val = E.val$
$E \to E_1 + T$	$E.\text{val} = E_1.\text{val} + T.\text{val}$
$E \to T$	$E.\text{val} = T.\text{val}$
$T \to T_1 * F$	$T.\text{val} = T_1.\text{val} * F.\text{val}$
$T \to F$	$T.\text{val} = F.\text{val}$
$F \to (E)$	$F.\text{val} = E.\text{val}$
$F \to \textbf{digit}$	$F.\text{val} = \textbf{digit}.\text{lexval}$

Note: all attributes in this example are of the synthesized type

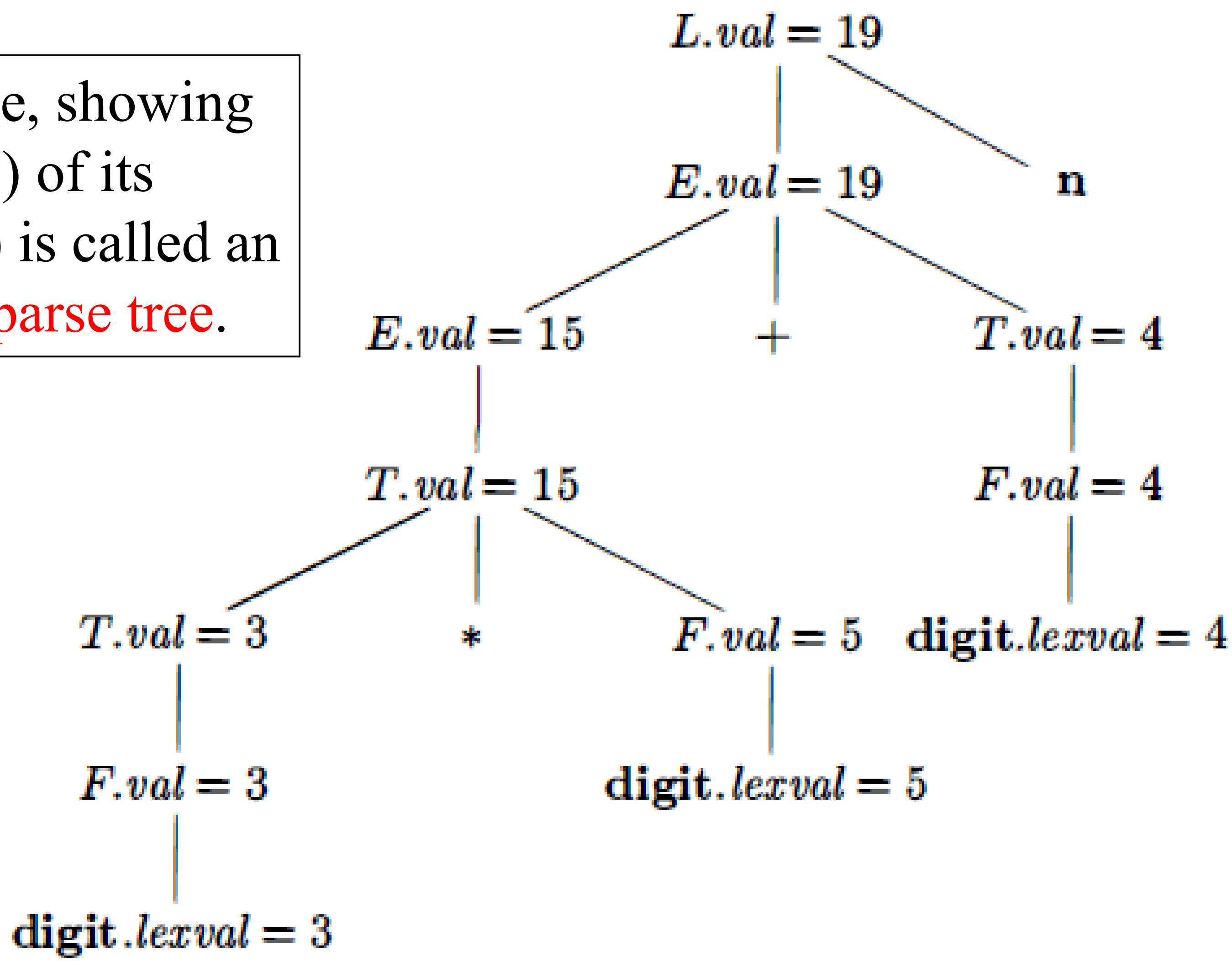
An SDD with only synthesized attributes is called S-attributed.

An SDD without side effects is called an attribute grammar

Annotated Parse Tree for 3 * 5 + 4 **n**

A parse tree, showing the value(s) of its attribute(s) is called an annotated parse tree.

$L \rightarrow E$ **n**
$E \rightarrow E_1 + T$
$E \rightarrow T$
$T \rightarrow T_1 * F$
$T \rightarrow F$
$F \rightarrow (E)$
$F \rightarrow$ **digit**



Annotating a Parse Tree With Depth-First Traversals

With synthesized attributes, we can evaluate attributes in any bottom-up order, such as that of a postorder traversal of the parse tree.

procedure *visit*(*n* : *node*);
begin
    for each child *m* of *n*, from left to right **do**
        visit(*m*);
    evaluate semantic rules at node *n*
end

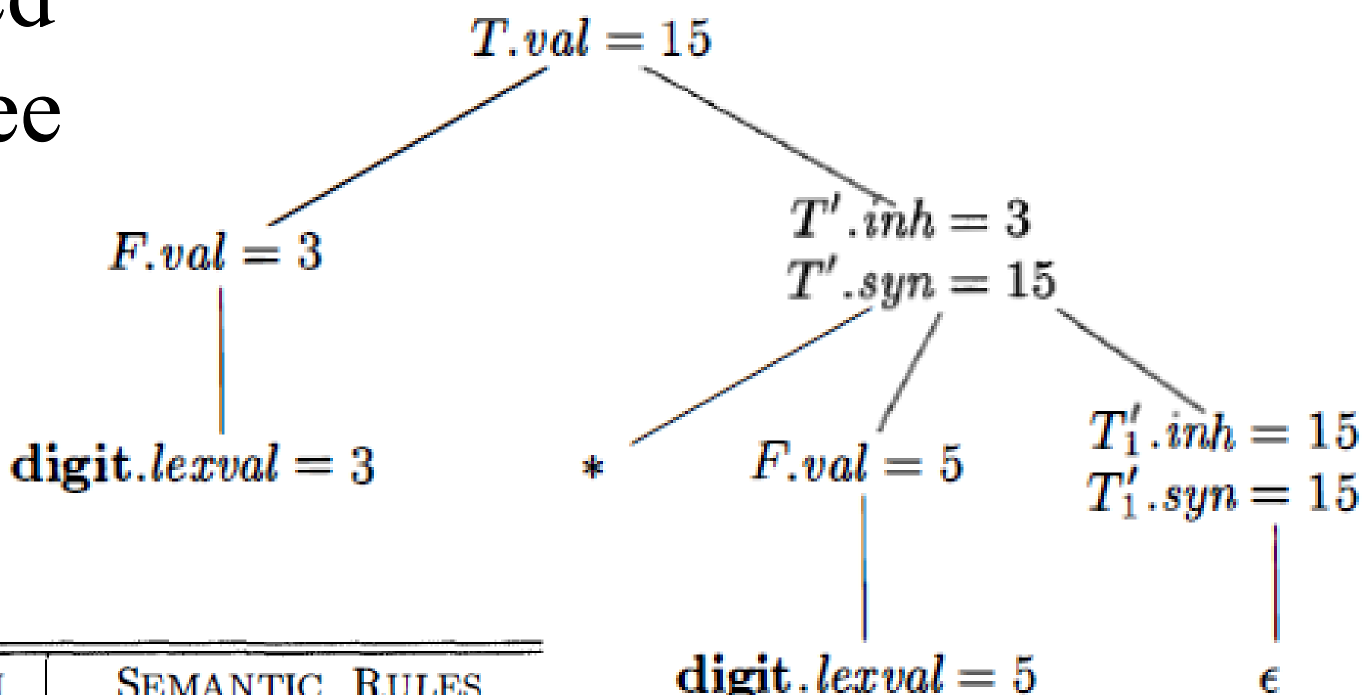
An SDD Based on a Grammar Suitable for Top-down Parsing

	Production	Semantic Rules
1)	$T \rightarrow F\, T'$	$T'.inh = F.val$ $T.val = T'.syn$
2)	$T' \rightarrow *\, F\, T_1'$	$T_1'.inh = T'.inh \times F.val$ $T'.syn = T_1'.syn$
3)	$T' \rightarrow \epsilon$	$T'.syn = T'.inh$
4)	$F \rightarrow \textbf{digit}$	$F.val = \textbf{digit}.lexval$

$T \rightarrow T * F$
$T \rightarrow F$
$F \rightarrow \textbf{digit}$

An inherited attribute for nonterminal T'
is used to pass the operand to the operator

Annotated Parse Tree for 3 * 5



PRODUCTION	SEMANTIC RULES
1) $T \rightarrow F\,T'$	$T'.inh = F.val$ $T.val = T'.syn$
2) $T' \rightarrow *\,F\,T_1'$	$T_1'.inh = T'.inh \times F.val$ $T'.syn = T_1'.syn$
3) $T' \rightarrow \epsilon$	$T'.syn = T'.inh$
4) $F \rightarrow \mathbf{digit}$	$F.val = \mathbf{digit}.lexval$

Example Attribute Grammar with Synthesized & Inherited Attributes

Simple Type Declaration

Production	Semantic Rule
$D \rightarrow T\ L$	L.inh $= T$.type
$T \rightarrow$ **int**	T.type $=$ '*integer*'
$T \rightarrow$ **float**	T.type $=$ '*float*'
$L \rightarrow L_1$ **, id**	L_1.inh $= L$.inh;
	addtype(**id**.entry, L.inh)
$L \rightarrow$ **id**	*addtype*(**id**.entry, L.inh)

Treated as dummy synthesized attribute with the head

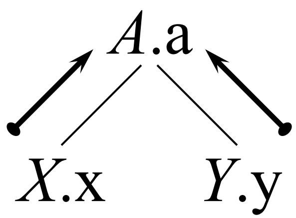
Synthesized: T.type, **id**.entry

Inherited: L.inh

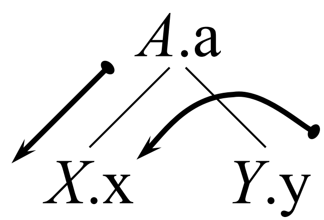
2. Evaluation Orders for SDD 's

A dependency graph depicts the flow of information among the attribute instances in a particular parse tree

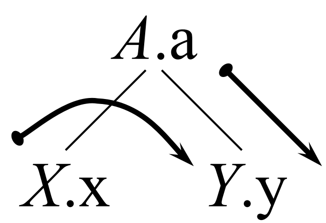
$A \rightarrow X\, Y$



$A.\mathrm{a} = f(X.\mathrm{x},\ Y.\mathrm{y})$



$X.\mathrm{x} = f(A.\mathrm{a},\ Y.\mathrm{y})$



$Y.\mathrm{y} = f(A.\mathrm{a},\ X.\mathrm{x})$

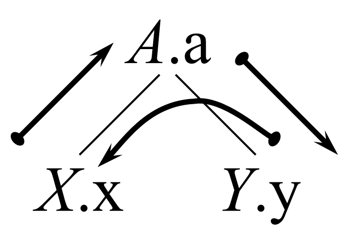
Direction of

value dependence

Evaluation Orders for SDD 's (Cont.)

- Edges in the dependency graph determine the evaluation order for attribute values
  - Dependency graphs cannot be cyclic
- So, dependency graph is a directed acyclic graph (DAG)

$A \rightarrow X\ Y$



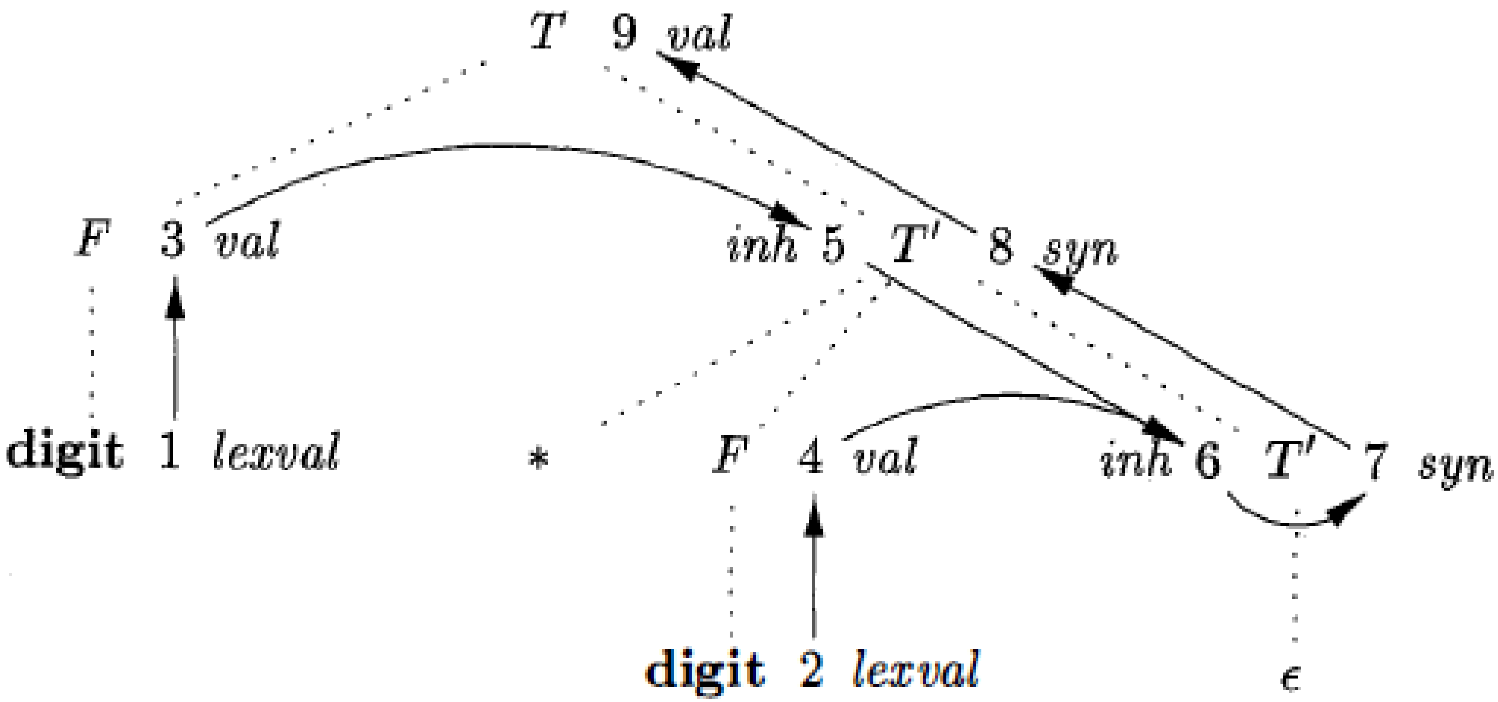
A.a := f(X.x)
X.x := f(Y.y)
Y.y := f(A.a)

Direction of ⟶ value dependence

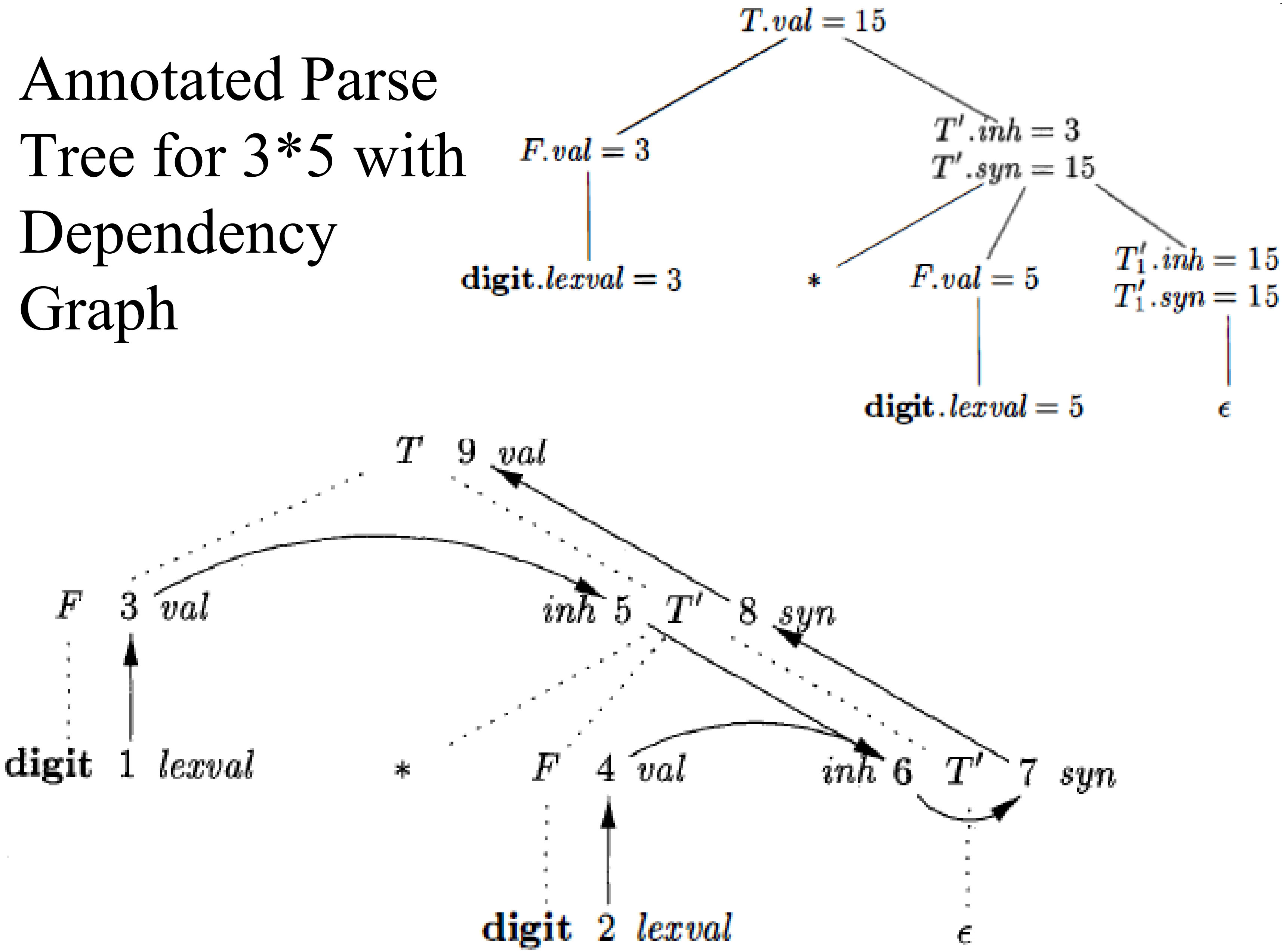
Error: cyclic dependence

Annotated Parse Tree for 3*5 with Dependency Graph

	Production	Semantic Rules
1)	$T \rightarrow F\,T'$	$T'.inh = F.val$ $T.val = T'.syn$
2)	$T' \rightarrow *\,F\,T'_1$	$T'_1.inh = T'.inh \times F.val$ $T'.syn = T'_1.syn$
3)	$T' \rightarrow \epsilon$	$T'.syn = T'.inh$
4)	$F \rightarrow \textbf{digit}$	$F.val = \textbf{digit}.lexval$



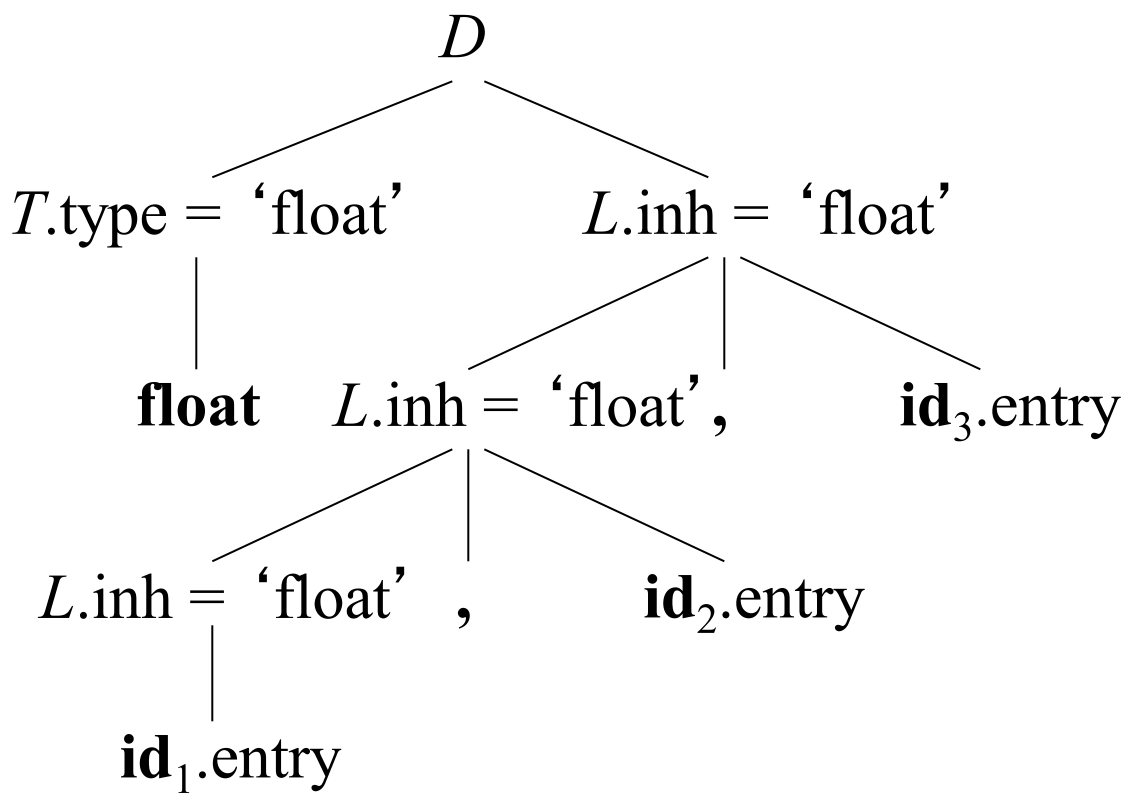
Annotated Parse Tree for 3*5 with Dependency Graph



Annotated Parse Tree

float id_1 , id_2 , id_3

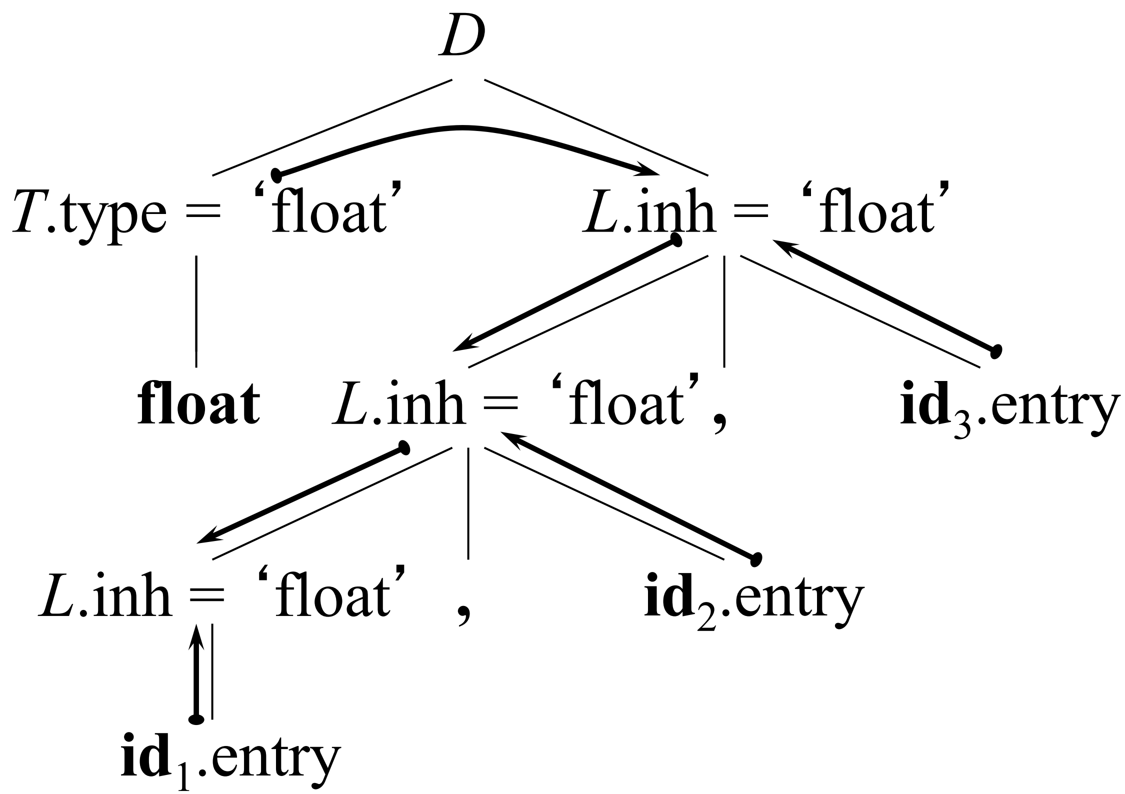
Production	Semantic Rules
$D \rightarrow T\ L$	L.inh = T.type
$T \rightarrow$ **int**	T.type = '*integer*'
$T \rightarrow$ **float**	T.type = '*float*'
$L \rightarrow L_1$ **, id**	L_1.inh = L.inh; *addtype*(**id**.entry, L.inh)
$L \rightarrow$ **id**	*addtype*(**id**.entry, L.inh)



Annotated Parse Tree with Dependency Graph

float id_1 , id_2 , id_3

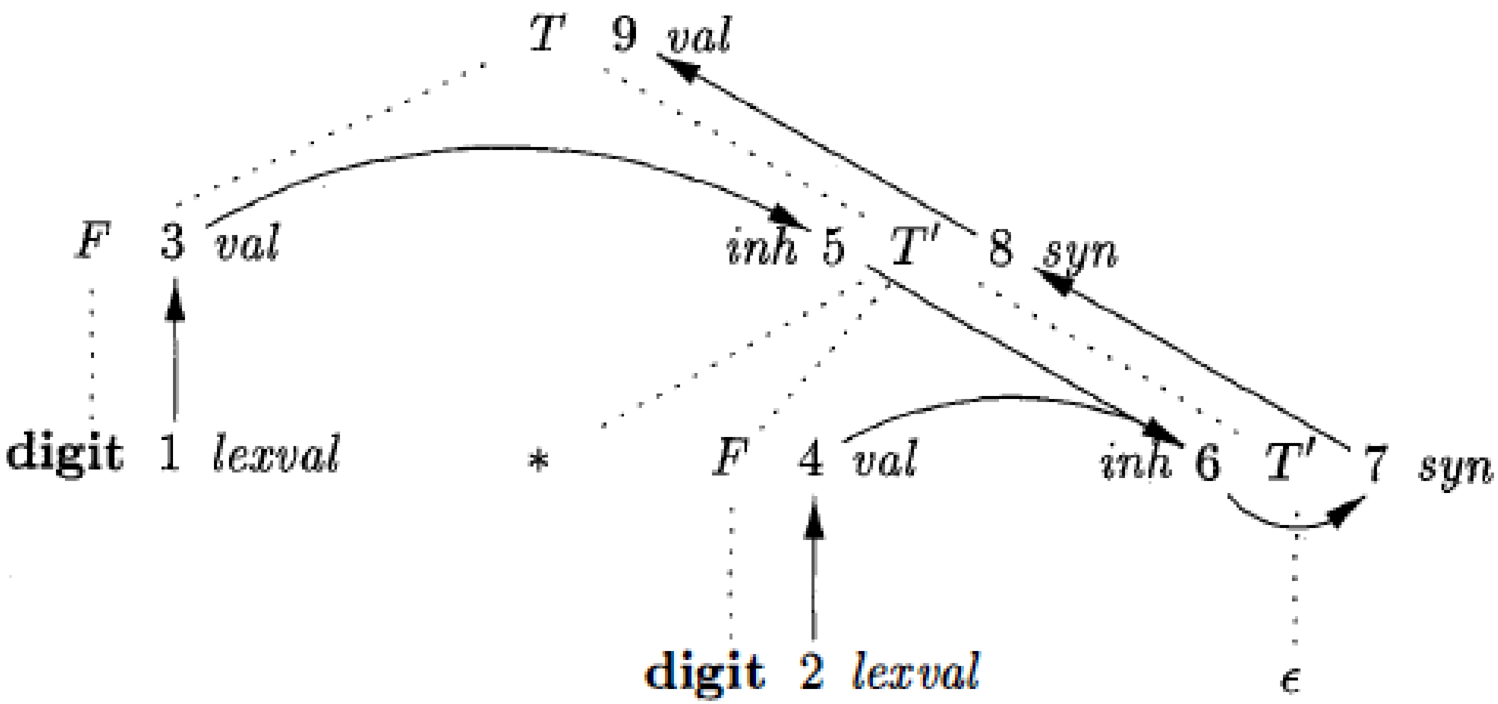
Production	Semantic Rules
$D \rightarrow T\ L$	L.inh = T.type
$T \rightarrow$ **int**	T.type = '*integer*'
$T \rightarrow$ **float**	T.type = '*float*'
$L \rightarrow L_1$ **, id**	L_1.inh = L.inh; *addtype*(**id**.entry, L.inh)
$L \rightarrow$ **id**	*addtype*(**id**.entry, L.inh)



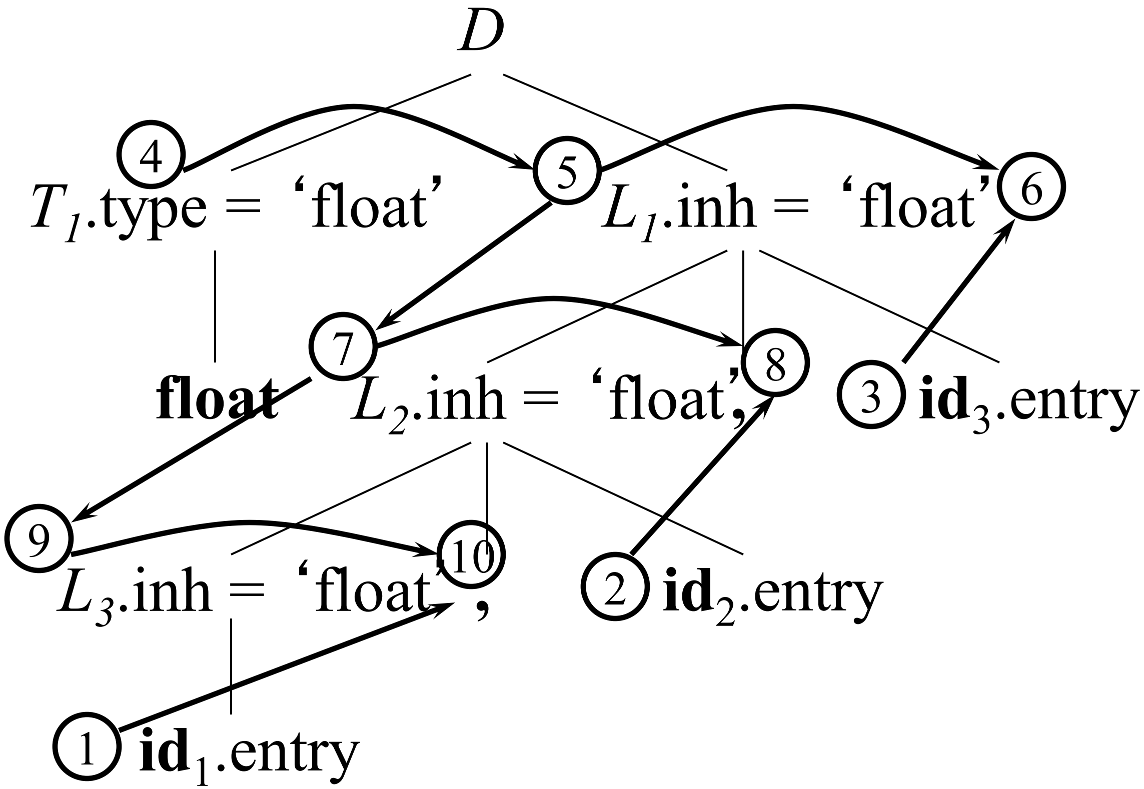
Evaluation Order

- A **topological sort** of a directed acyclic graph (DAG) is any ordering $m_1, m_2, \ldots, m_n$ of the nodes of the graph, such that if $m_i \rightarrow m_j$ is an edge, then m_i appears before m_j
- Any topological sort of a dependency graph gives a valid evaluation order of the semantic rules
- Example: Topological orders of DAG on slide 15
  - 1, 2, 3, 4, 5, 6, 7, 8, 9.
  - 1, 3, 5, 2, 4, 6, 7, 8, 9.

- Example: Topological orders of the following DAG
  - 1, 2, 3, 4, 5, 6, 7, 8, 9.
  - 1, 3, 5, 2, 4, 6, 7, 8, 9.



Example Parse Tree with Topologically Sorted Actions



Topological sort:

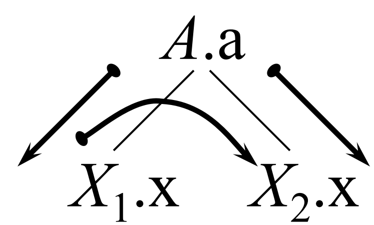
1. Get **id**$_1$.entry
2. Get **id**$_2$.entry
3. Get **id**$_3$.entry
4. T_1.type= 'float'
5. L_1.inh=T_1.type
6. *addtype*(**id**$_3$.entry, L_1.inh)
7. L_2.inh=L_1.inh
8. *addtype*(**id**$_2$.entry, L_2.inh)
9. L_3.inh=L_2.inh
10. *addtype*(**id**$_1$.entry, L_3.inh)

float id$_1$, id$_2$, id$_3$

L-Attributed Definitions

- A syntax-directed definition is **L-attributed** if each inherited attribute of X_j on the right side of production $A \rightarrow X_1\, X_2 \ldots X_n$ depends only on
  1. the attributes of the symbols $X_1, X_2, \ldots, X_{j\text{-}1}$
  2. the inherited attributes of A

Shown: dependences of inherited attributes



- L-attributed definitions allow for a natural order of evaluating attributes: **depth-first and left to right**
- Note: every S-attributed syntax-directed definition is also L-attributed

3. Applications of SDD
Construction of Syntax Trees

S-attributed Definition for Simple Expressions

	Production	Semantic Rules
1)	$E \rightarrow E_1 + T$	$E.node = \textbf{new}\ Node('+', E_1.node, T.node)$
2)	$E \rightarrow E_1 - T$	$E.node = \textbf{new}\ Node('-', E_1.node, T.node)$
3)	$E \rightarrow T$	$E.node = T.node$
4)	$T \rightarrow (\ E\ )$	$T.node = E.node$
5)	$T \rightarrow \textbf{id}$	$T.node = \textbf{new}\ Leaf(\textbf{id}, \textbf{id}.entry)$
6)	$T \rightarrow \textbf{num}$	$T.node = \textbf{new}\ Leaf(\textbf{num}, \textbf{num}.val)$

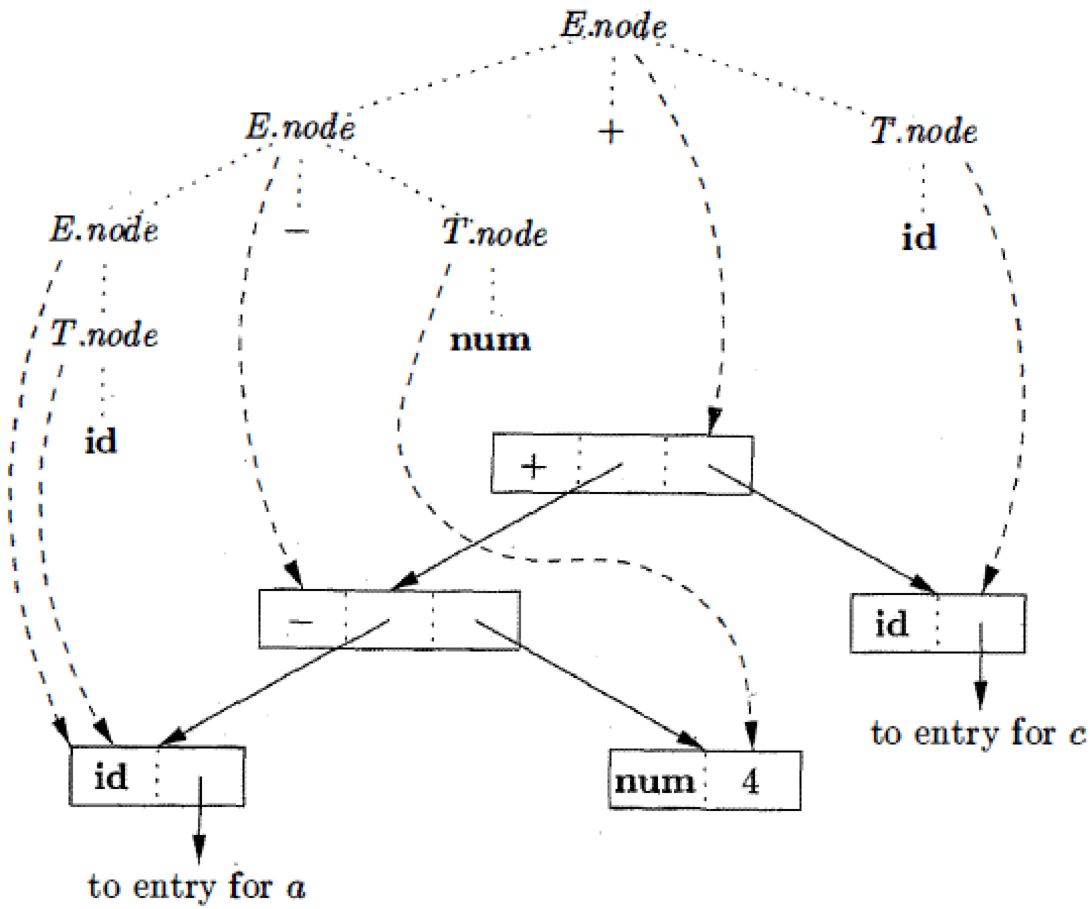
Note: This is a S-attributed definition, then can be done during bottom-up parsing

Example: Syntax Tree for a - 4 + c

Steps in the construction of the syntax tree

1) $p_1 = \textbf{new } \mathit{Leaf}(\textbf{id}, \mathit{entry}\text{-}a);$
2) $p_2 = \textbf{new } \mathit{Leaf}(\textbf{num}, 4);$
3) $p_3 = \textbf{new } \mathit{Node}('-', p_1, p_2);$
4) $p_4 = \textbf{new } \mathit{Leaf}(\textbf{id}, \mathit{entry}\text{-}c);$
5) $p_5 = \textbf{new } \mathit{Node}('+', p_3, p_4);$

	PRODUCTION	SEMANTIC RULES
1)	$E \rightarrow E_1 + T$	$E.node = \textbf{new } \mathit{Node}('+', E_1.node, T.node)$
2)	$E \rightarrow E_1 - T$	$E.node = \textbf{new } \mathit{Node}('-', E_1.node, T.node)$
3)	$E \rightarrow T$	$E.node = T.node$
4)	$T \rightarrow (E)$	$T.node = E.node$
5)	$T \rightarrow \textbf{id}$	$T.node = \textbf{new } \mathit{Leaf}(\textbf{id}, \textbf{id}.entry)$
6)	$T \rightarrow \textbf{num}$	$T.node = \textbf{new } \mathit{Leaf}(\textbf{num}, \textbf{num}.val)$



Constructing Syntax Tree During Top-Down Parsing

L-attributed Definition for Simple Expression

	Production	Semantic Rules
1)	$E \rightarrow T\ E'$	$E.node = E'.syn$ $E'.inh = T.node$
2)	$E' \rightarrow +\ T\ E'_1$	$E'_1.inh = \textbf{new}\ Node('+', E'.inh, T.node)$ $E'.syn = E'_1.syn$
3)	$E' \rightarrow -\ T\ E'_1$	$E'_1.inh = \textbf{new}\ Node('-', E'.inh, T.node)$ $E'.syn = E'_1.syn$
4)	$E' \rightarrow \epsilon$	$E'.syn = E'.inh$
5)	$T \rightarrow (\ E\ )$	$T.node = E.node$
6)	$T \rightarrow \textbf{id}$	$T.node = \textbf{new}\ Leaf(\textbf{id}, \textbf{id}.entry)$
7)	$T \rightarrow \textbf{num}$	$T.node = \textbf{new}\ Leaf(\textbf{num}, \textbf{num}.val)$