4. Syntax-Directed Translation

Eva Rose Kristoffer Rose

NYU Courant Institute
Compiler Construction (CSCI-GA.2130-001)
http://cs.nyu.edu/courses/fall14/CSCI-GA.2130-001/lecture-4.pdf

September 25, 2014





- 2 SDT: Syntax-Directed Translation
- SDD: Syntax-Directed Definition
- 4 HACS
 - Review
 - S-attributed SDDs
 - Recursive Translation Schemes

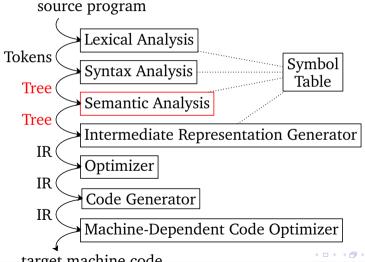




- Introduction
- SDT: Syntax-Directed Translation
- SDD: Syntax-Directed Definition
- HACS
 - Review
 - S-attributed SDDs
 - Recursive Translation Schemes











The Trees

Introduction

TREE	INTERIOR NODES	Grammar
parse tree	nonterminals	concrete syntax
abstract syntax tree	programming constructs	abstract syntax

Example (abstract syntax)



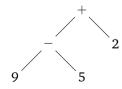
$$E \rightarrow E + E \mid E - E \mid$$
 digit





TREE	INTERIOR NODES	Grammar
parse tree	nonterminals	concrete syntax
abstract syntax tree	programming constructs	abstract syntax

Example (abstract syntax)



$$E \rightarrow E + E \mid E - E \mid$$
digit





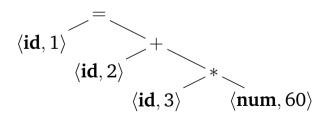
Back to the introductory example...





From Abstract Syntax Tree (AST)

parsed into abstract syntax tree:

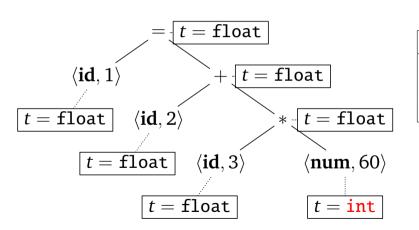


id	lexeme
1	position
2	initial
3	rate





... to Annotated AST....

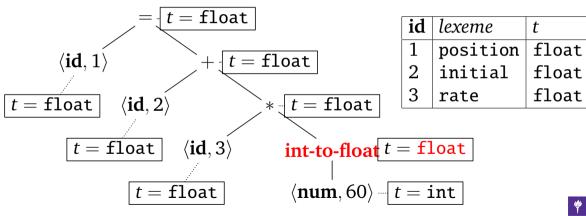


id	lexeme	t
1	position	float
2	initial	float
3	rate	float





... to "Fixed" AST





- Introduction
- SDT: Syntax-Directed Translation
- SDD: Syntax-Directed Definition
- 4 HACS
 - Review
 - S-attributed SDDs
 - Recursive Translation Schemes





Syntax-Directed Translation

We have built a parse (or AST) tree, now what? How will this tree and production rules help the translation?

Intuitively, we need to associate *something* with each production and each tree node:

- something that evaluates the meaning at each node,
- something that emits the meaning as program fragments.





Syntax-Directed Translation

We have built a parse (or AST) tree, now what? How will this tree and production rules help the translation?

Intuitively, we need to associate *something* with each production and each tree node:

- something that evaluates the meaning at each node,
- something that emits the meaning as program fragments.





Syntax-Directed Definition (SDD)

Attributes Each grammar symbol (terminal or nonterminal) has an attribute ("meaning") associated.

Semantic Rules Each production has semantic rules associated for computing the attributes.





Example: SDD for Infix to Postfix Notation

Consider the grammar:

$$expr \rightarrow expr + term$$
 $\mid expr - term$
 $\mid term$
 $term \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9$

expr has an attribute expr.t, term has an attribute term.t.





Example: SDD Infix to Postfix Notation

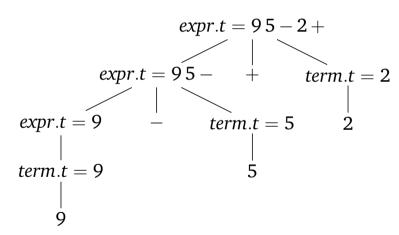
PRODUCTION	SEMANTIC RULE
	$ expr.t = expr_1.t term_2.t '+'$
$expr o expr_1 - term_2$	$\mid expr.t = expr_1.t \mid \mid term_2.t \mid \mid '-'$
$expr o term_1$	$expr.t = term_1.t$
term o 0	term.t = 0'
term ightarrow 1	term.t = '1'
• • •	
$term \rightarrow 9$	term.t =' 9'

|-| means concatenation, |-1|, |-2| disambiguates





Attribute Values in Parse Trees





Translation Schemes

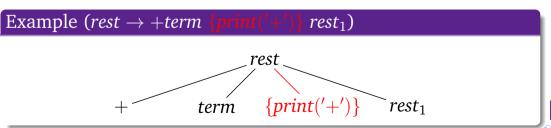
- Equivalent mechanism.
- Attaching program fragments to productions.
- ► The program fragments are called semantic actions/side-effects.
- ▶ They "emit" the program fragments during "tree traversal".





Translation Schemes

- Equivalent mechanism.
- Attaching program fragments to productions.
- ► The program fragments are called semantic actions/side-effects.
- ▶ They "emit" the program fragments during "tree traversal".



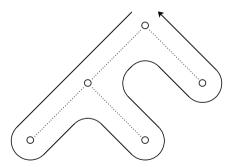


Semantic Action Table: Infix to Postfix Notation

```
PRODUCTIONS WITH SIDE-EFFECTS/ACTIONS
expr \rightarrow expr_1 + term_2 \{ print('+') \}
expr \rightarrow expr_1 - term_2 \{ print('-') \}
expr \rightarrow term
term \rightarrow 0 \{ print('0') \}
term \rightarrow 1 \{ print('1') \}
term \rightarrow 2 \{ print('2') \}
term \rightarrow 3 \{ print('3') \}
term \rightarrow 4 \{ print('4') \}
term \rightarrow 5 \{ print('5') \}
term \rightarrow 6 \{ print('6') \}
term \rightarrow 7 \{ print('7') \}
term \rightarrow 8 \{ print('8') \}
term \rightarrow 9 \{ print('9') \}
```



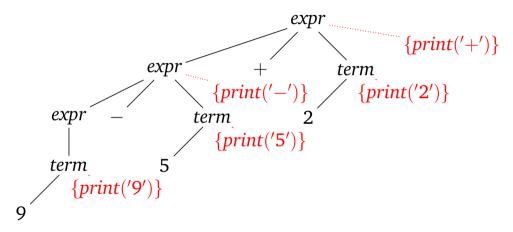








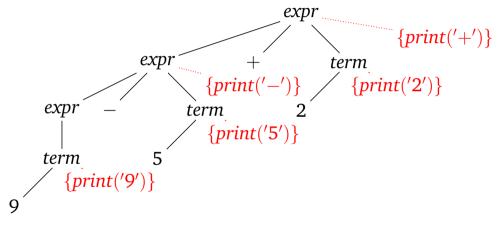
Semantic Actions/Side-Effects in Parse Tree



Input: 9 - 5 + 2. Output (print): 95 - 2 +. Single depth-first

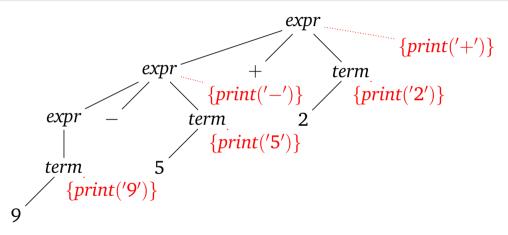


Semantic Actions/Side-Effects in Parse Tree



Input: 9 - 5 + 2. Output (print): 95 - 2 +.

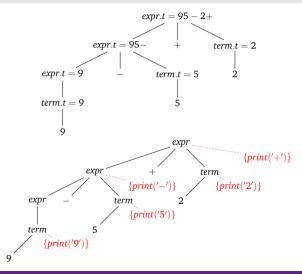




Input: 9 - 5 + 2. Output (print): 95 - 2 +. Single depth-first



Summary



attributes/semantic rules (HACS)

semantic actions/side-effects (yacc)





Tree Traversal and Translation Schemes

When translation is defined in terms of semantic actions, the tree traversal order, that is the order in which the nodes are visited, becomes essential.





Exercise

Construct an SDD that generates an attribute t in prefix notation for each expression expr, for arithmetic expressions. (You only need to consider: '+','-', and '*' expressions). Prefix notation is where the operator comes before its operands; e.g., -xy is the prefix notation for x-y.





Solution: SDD for Infix Notation into Prefix Notation

PRODUCTION	SEMANTIC RULE
$expr ightarrow expr_1 + term_2$	$ expr.t = ' +' expr_1.t term_2.t$
$expr o expr_1 - term_2$	$ expr.t = ' - ' expr_1.t term_2.t$
$expr o expr_1 * term_2$	$ expr.t = '*' expr_1.t term_2.t$
expr o term	expr.t = term.t
term o 0	term.t = 0'
term ightarrow 1	term.t = '1'
term o 9	term.t = '9'





- Introduction
- 2 SDT: Syntax-Directed Translation
- SDD: Syntax-Directed Definition
- 4 HACS
 - Review
 - S-attributed SDDs
 - Recursive Translation Schemes





Syntax-Directed Definition (SDD)

Recall the key definition:

Grammar Given a context-free grammar.

Attributes Each grammar symbol (terminal or nonterminal) has a set of *attributes* associated.

Semantic Rules Each production has a set of *semantic rules* associated for computing the attributes.

If *X* is a symbol, *a* is an attribute, then *X*.*a* denotes the value of *a* at node *X*.





Synthesized and Inherited Attributes

Synthesized attributes at node N are defined only in terms of the attribute values of the children of N, and N itself.

Inherited attributes at node N is defined only in terms of attribute values at N's parent, N itself, and N's siblings.





Example: SDD for Desk Calculator

	PRODUCTION	SEMANTIC RULE
1.	L o E \$	L.val = E.val
2.	$E ightarrow E_1 + T$	$E.val = E_1.val + T.val$
3.	E o T	E.val = T.val
4.	$T o T_1 * F$	$T.val = T_1.val \times F.val$
5.	T o F	T.val = F.val
6.	$F \rightarrow (E)$	F.val = E.val
7.	$F o \mathbf{digit}$	$F.val = \mathbf{digit}.lexval$

'val' and 'lexval' are synthesized attributes.





Characteristics of Desk Calculator

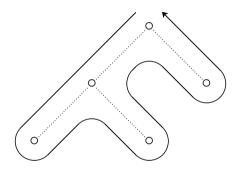
- S-attributed (only synthesized attributes)
- attribute grammar (without side-effects)

Order of attribute evaluation: any bottom-up traversal.





Tree Traversal: Depth-First



Post-order (default). Pre-order. Binary trees: also in-order.



SDD: Syntax-Directed Definition

Exercise: Desk Calculator

Give annotated parse trees for the following expressions:

- (3+4)*(5+6)\$
- 1*2*3*(4+5)\$

(Parse trees showing attributes and their values)





Exercise: Mingling with Inherited Attributes...

	PRODUCTION	SEMANTIC RULES
1.	T o F T'	T'.inh = F.val; T.val = T'.syn
2.	$T' o *FT_1'$	$T_1'.inh = T'.inh \times F.val; T'.syn = T_1'.syn$
3.	$T' o \epsilon$	T'. $syn = T'$. inh
4.	$F o \mathbf{digit}$	$F.val = \mathbf{digit}.lexval$

'inh' is inherited, and 'val', 'syn' are synthesized.





Dependency Graphs

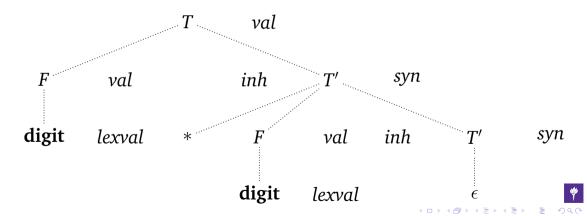
Dependency graphs depicts the flow of information among attributes.





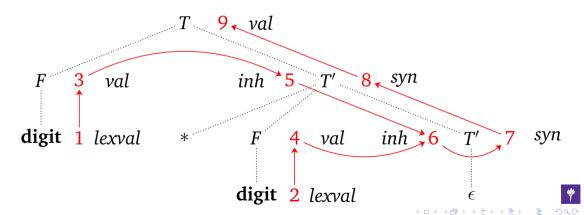
Example: Attributes

Parse tree for 3 * 5 based on SDD:



Example: Attributes with Dependency Graph

Parse tree for 3 * 5 based on SDD:



Dependency Graphs

Central problem: Determining the efficient evaluation order for the attribute instances in a given parse tree.





Calculating Attributes

Cycles possible!

PRODUCTION	N SEMANTIC RULE	
A o B	A.s = B.i; B.i = A.s + 1	

When dependency graphs have cycles, evaluation is "stuck"!





Well-behaved SDD Classes

Given an SDD, hard to tell if there exists a parse tree whose dependency graphs have cycles!





Well-behaved SDD Classes

Given an SDD, hard to tell if there exists a parse tree whose dependency graphs have cycles!

Safe way to go: use classes of SDDs that guarantee an evaluation order:

S-Attributed Definitions all attributes are synthesized. L-Attributed Definitions "left" restrictions on dependency graph edges.





L-Attributed Definitions

Suppose an attribute rule for $A \to X_1 X_2 \dots X_i \dots X_n$ produces inherited $X_i.a$ attribute. The rule may only use:

- ▶ inherited attributes associated with the head (*A*),
- ▶ inherited or synthesized attributes associated with the occurences of symbols $X_1, X_2, ..., X_{i-1}$ located left of X_i .
- ▶ Inherited or synthesized attributes associated with the occurrence of X_i itself, as long as not part of a dependency graph cycle!

or, the produced attribute is synthesized.





Example: Mingling with Inherited Attributes...

Decide if the SDD is L-attributed and argue:

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

'inh' is inherited, and 'val', 'syn' are synthesized.





Example: Mingling with Inherited Attributes...

Decide if the SDD is L-attributed and argue:

	PRODUCTION	SEMANTIC RULES
1.	T o F T'	T'.inh = F.val; T.val = T'.syn
2.	$T' o *FT_1'$	$T'_1.inh = T'.inh \times F.val; T'.syn = T'_1.syn$
3.	$T' o \epsilon$	T'.syn = T' .inh
4.	$F o \mathbf{digit}$	$F.val = \mathbf{digit}.lexval$

'inh' is inherited, and 'val', 'syn' are synthesized.





Example: Mingling with Inherited Attributes...

Decide if the SDD is L-attributed and argue:

	PRODUCTION	SEMANTIC RULES
1.	T o F T'	T'.inh = F.val; T.val = T'.syn
2.	$T' o *FT_1'$	$T'_1.inh = T'.inh \times F.val; T'.syn = T'_1.syn$
3.	$T' o \epsilon$	T'. $syn = T'$. inh
4.	$F o \mathbf{digit}$	F.val = digit.lexval

'inh' is inherited, and 'val', 'syn' are synthesized.





Another example

The following SDD is NOT L-attributed. Why?

PRODUCTION	SEMANTIC RULES	
$A \rightarrow BC$	A.s = B.b	
	B.i = F(C.c, A.s)	

where B.i is inherited, and A.s, C.c are synthesized.





Another example

The following SDD is NOT L-attributed. Why?

PRODUCTION	SEMANTIC RULES	
$A \rightarrow BC$	A.s = B.b	
	B.i = F(C.c, A.s)	

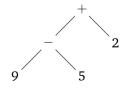




Construction of Abstract Syntax Trees

TREE	INTERIOR NODES	Grammar
parse tree	nonterminals	concrete syntax
abstract syntax tree	programming constructs	abstract syntax

Example (abstract syntax)



$$E \rightarrow E + E \mid E - E \mid \mathbf{digit}$$





Construction of AST

Example 1: S-attributed SDD for simple expresssions.

	Ъ	0 0
	PRODUCTION	Semantic Rule
1.	$E ightarrow E_1 + T_2$	$E.node = $ new $Node('+', E_1.node, T_2.node)$
2.	$E ightarrow E_1 - T_2$	$E.node = \mathbf{new} \ Node('-', E_1.node, T_2.node)$
3.	E o T	E.node = T.node
4.	T o (E)	T.node = E.node
5.	$T o \mathbf{id}$	$T.node = \mathbf{new} \ Leaf(\mathbf{id}, \mathbf{id}.entry)$
6.	$T \rightarrow \mathbf{num}$	$T.node = \mathbf{new} \ Leaf(\mathbf{num}, \mathbf{num}.entry)$

▶ Show the AST for a - 4 + c





Construction of AST

Example 2: L-attributed SDD for simple expresssions.

	PRODUCTION	SEMANTIC RULE
1.	E o T E'	E.node = E'.syn; E'.inh = T.node
2.	$E' ightarrow + T E_1'$	$E'_1.inh = \mathbf{new} \ Node('+', E'.node, T.node)$
	_	$E'.syn = E'_1.syn$
3.	$E' ightarrow - T E_1'$	$E'_1.inh = $ new $Node('-', E'.node, T.node)$
	_	$E^{\prime}.syn=E_{1}^{\prime}.syn$
4.	$E' o \epsilon$	$E'.syn = E^{ar{\prime}}.inh$
5.	T o (E)	T.node = E.node
6.	$T \rightarrow \mathbf{id}$	T.node = new $Leaf($ id , id . $entry)$
7.	$T o \mathbf{num}$	T.node = new $Leaf($ num , num .entry $)$

Show AST and dependency graph for a - 4 + c





Syntax-Directed Translation of Type Expressions

	PRODUCTION	SEMANTIC RULE
1	T o B C	T.syn = C.syn
		C.inh = B.syn
2	$B o {f int}$	B.syn = integer
3	$B \rightarrow \mathbf{float}$	B.syn = integer
4	$C \rightarrow [\mathbf{num}]C_1$	$C.syn = array(\mathbf{num}.val, C_1.syn)$
		$C_1.inh = C.inh$
5	$C o \epsilon$	C.syn = C.inh

► Show AST and dependencies for **int**[2][3]





- Introduction
- SDT: Syntax-Directed Translation
- SDD: Syntax-Directed Definition
- **HACS**
 - Review
 - S-attributed SDDs
 - Recursive Translation Schemes





HACS Review: Lexical Specification

```
space [ t ] ;
token Int |\langle Digit \rangle + ;
token Float | (Int) "." (Int);
token fragment Digit | [0-9] |;
```





$E ightarrow E + T \mid T$ $T ightarrow T * F \mid F$ $F ightarrow (E) \mid \mathbf{int} \mid \mathbf{float}$ (4.1)

sort Exp |
$$[\langle Exp@1 \rangle + \langle Exp@2 \rangle]@1$$

| $[\langle Exp@2 \rangle * \langle Exp@3 \rangle]@2$
| $[\langle Int \rangle]@3 | [\langle Float \rangle]@3$
| sugar $[(\langle Exp#1@1 \rangle)]@3 \rightarrow Exp#1;$

 $E \rightarrow E + E \mid E * E \mid$ **int** \mid **float**



HACS Review: Parser Specification

$$E \rightarrow E + T \mid T$$
 $T \rightarrow T * F \mid F$
 $F \rightarrow (E) \mid \mathbf{int} \mid \mathbf{float}$ (4.1)

```
sort Exp | [\langle Exp@1 \rangle + \langle Exp@2 \rangle] @1
                      [\langle Exp@2 \rangle * \langle Exp@3 \rangle] @2
                       [\langle Int \rangle] = 3 | [\langle Float \rangle] = 3
                      sugar \lceil (\langle \text{Exp} # 1@1 \rangle) \rceil @ 3 \rightarrow \text{Exp} # 1;
```



HACS Review: Parser Specification

$$E \rightarrow E + T \mid T$$
 $T \rightarrow T * F \mid F$
 $F \rightarrow (E) \mid \mathbf{int} \mid \mathbf{float}$ (4.1)

sort Exp |
$$[\langle \text{Exp@1} \rangle + \langle \text{Exp@2} \rangle]$$
@1
| $[\langle \text{Exp@2} \rangle * \langle \text{Exp@3} \rangle]$ @2
| $[\langle \text{Int} \rangle]$ @3 | $[\langle \text{Float} \rangle]$ @3
| sugar $[(\langle \text{Exp#1@1} \rangle)]$ @3 \rightarrow Exp#1;

$$E \rightarrow E + E \mid E * E \mid$$
int \mid **float**



This Week

- S-attributed SDDs.
- ▶ Recursive Translation Schemes.





HACS

S-attributed SDD

Only synthesized attributes.





HACS

Example: Type Synthesis SDD

PRODUCTION	SEMANTIC RULES	
$E \rightarrow E_1 + E_2$	$E.t = \text{Unif}(E_1.t, E_2.t)$	(1)
$\mid E_1 * E_2 \mid$	$E.t = \text{Unif}(E_1.t, E_2.t)$	(2)
int	E.t = Int	(3)
float	E.t = Float	(4)





Example: Type Synthesis HACS Sorts

```
sort Type | Int | Float;
```





Example: Type Synthesis HACS Unification Rules

```
sort Type | scheme Unif(Type,Type);
Unif(Int, Int) \rightarrowInt;
Unif(Float, \#) \rightarrow Float;
Unif(#, Float) \rightarrowFloat;
```





$$E \rightarrow E_1 + E_2 \quad | \quad E.t = \text{Unif}(E_1.t, E_2.t)$$
 (1)

$$[(Exp#1 + type(\#t1)) + (Exp#2 + type(\#t2))] \uparrow type(Umf(\#t1, \#t2))$$





$$E \rightarrow E_1 + E_2 \quad | \quad E.t = \text{Unif}(E_1.t, E_2.t)$$
 (1)

$$[\![\langle Exp\#1\uparrow type(\#t1)\rangle + \langle Exp\#2\uparrow type(\#t2)\rangle]\!]\uparrow type(Unif(\#t1,\#t2))$$





$$E \rightarrow E_1 + E_2 \quad | \quad E.t = \text{Unif}(\underline{E_1.t}, E_2.t)$$
 (1)

$$[\![\langle Exp\#1 \uparrow type(\#t1)\rangle + \langle Exp\#2 \uparrow type(\#t2)\rangle]\!] \uparrow type(Unif(\#t1, \#t2))$$





$$E \rightarrow E_1 + E_2 \quad | \quad E.t = \text{Unif}(E_1.t, \underline{E_2.t})$$
 (1)

$$[\![\langle Exp\#1 \uparrow type(\#t1)\rangle + \langle Exp\#2 \uparrow type(\#t2)\rangle]\!] \uparrow type(Unif(\#t1, \#t2))$$





$$E \rightarrow E_1 + E_2 \quad | \quad E.t = \operatorname{Unif}(E_1.t, E_2.t)$$
 (1)

$$[\![\langle Exp\#1 \uparrow type(\#t1)\rangle + \langle Exp\#2 \uparrow type(\#t2)\rangle]\!] \uparrow type(Unif(\#t1, \#t2))$$





Example: Type Synthesis HACS

```
attribute ↑type(Type);
sort Exp; ↑type;
   \langle \text{Exp}\#1 \uparrow \text{type}(\#t1) \rangle * \langle \text{Exp}\#2 \uparrow \text{type}(\#t2) \rangle ] \uparrow \text{type}(\text{Unif}(\#t1,\#t2));
[ \langle Int # \rangle ] \uparrow type(Int);
 \langle \text{Float} \# \rangle \uparrow \text{type}(\text{Float});
```





Recursive Translation Scheme

// New syntax for value conversion from Int to Float:

```
sort Exp | [float (Exp)];
```





Recursive Translation Scheme (II)

// New scheme for inserting all needed int—to—float conversions:

```
sort Exp | scheme I2F(Exp);
```



HACS



September 25, 2014 58/62

Recursive Translation Scheme (III)

```
// Cases for +:
I2F([\langle Exp\#1\uparrow type(Int)\rangle + \langle Exp\#2\uparrow type(Int)\rangle]]\uparrow type(\#t))
   \rightarrow \mathbb{I}\langle \text{Exp I2F}(\#1)\rangle + \langle \text{Exp I2F}(\#2)\rangle \mathbb{I}\uparrow \text{type}(\#t);
\frac{12F([\langle Exp\#1\uparrow type(Float)\rangle + \langle Exp\#2\uparrow type(Int)\rangle] \uparrow type(\#t))}{}
   \rightarrow [\langle \text{Exp I2F}(\#1) \rangle + (\text{float}\langle \text{Exp I2F}(\#2) \rangle)] \uparrow \text{type}(\#t):
I2F([\langle Exp\#1\uparrow type(Int)\rangle + \langle Exp\#2\uparrow type(Float)\rangle] \uparrow type(\#t))
   \rightarrow [(\text{float}(\text{Exp I2F}(\#1))) + (\text{Exp I2F}(\#2))] \uparrow \text{type}(\#t);
I2F([\langle Exp#1\uparrow type(Float)\rangle + \langle Exp#2\uparrow type(Float)\rangle] \uparrow type(#t))
   \rightarrow [\langle \text{Exp I2F}(\#1) \rangle + \langle \text{Exp I2F}(\#2) \rangle] \uparrow \text{type}(\#t);
```



Recursive Translation Scheme (IV)

```
// Cases for *:
\frac{12F([\langle Exp\#1\uparrow type(Int)\rangle *\langle Exp\#2\uparrow type(Int)\rangle]]\uparrow type(\#t))}{}
   \rightarrow \mathbb{I}\langle \text{Exp I2F}(\#1)\rangle * \langle \text{Exp I2F}(\#2)\rangle \mathbb{I}\uparrow \text{type}(\#t);
I2F([\langle Exp\#1\uparrow type(Float)\rangle *\langle Exp\#2\uparrow type(Int)\rangle]]\uparrow type(\#t))
   \rightarrow [\langle \text{Exp I2F}(\#1) \rangle * (\text{float}\langle \text{Exp I2F}(\#2) \rangle)] \uparrow \text{type}(\#t):
\frac{12F}{([\langle Exp\#1\uparrow type(Int)\rangle *\langle Exp\#2\uparrow type(Float)\rangle]]\uparrow type(\#t))}
   \rightarrow \|(\text{float}(\text{Exp I2F}(\#1))) * (\text{Exp I2F}(\#2))\| \uparrow \text{type}(\#t):
I2F([\langle Exp#1 \uparrow type(Float) \rangle * \langle Exp#2 \uparrow type(Float) \rangle]] \uparrow type(#t))
   \rightarrow [\langle \text{Exp I2F}(\#1) \rangle * \langle \text{Exp I2F}(\#2) \rangle] \uparrow \text{type}(\#t);
```



Recursive Translation Scheme (II)

```
// Cases for literals:
```

```
\frac{12F(\lceil \langle Int\#1 \rangle \rceil \uparrow type(\#t))}{} \rightarrow \lceil \langle Int\#1 \rangle \rceil \uparrow type(\#t);
I2F([\langle Float \# 1 \rangle]] \uparrow type(\# t)) \rightarrow [\langle Float \# 1 \rangle]] \uparrow type(\# t);
```





krisrose@cs.nyu.edu evarose@cs.nyu.edu



