Compilation process

How a program is born

CocoaHeads Berlin, 2015

> whoami

Twitter:
@1101_debian

Github:
@AlexDenisov

Freenode:
AlexDenisov

Outline

- Compilation process
- LLVM/Clang
- Q & A

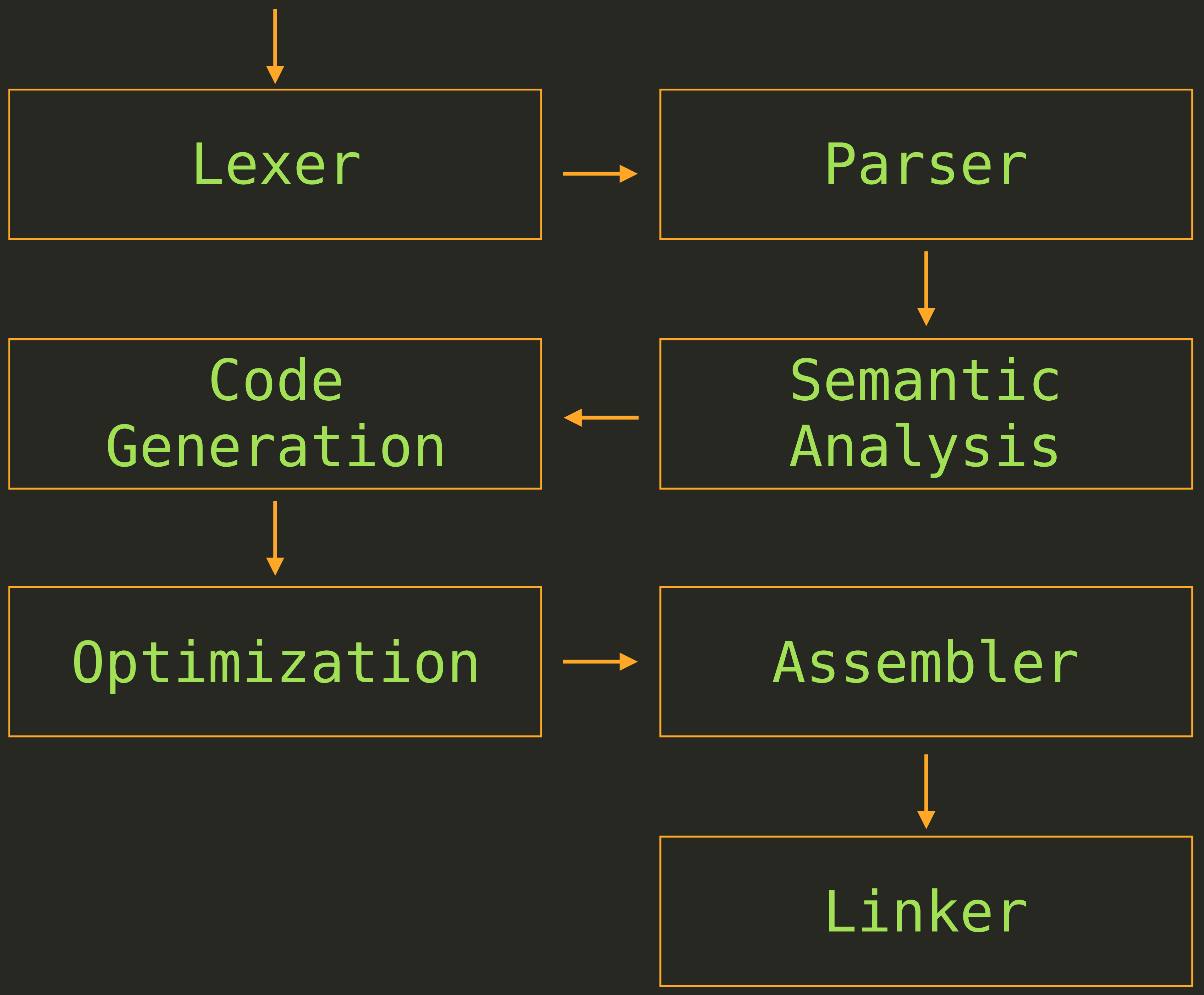
Compilation Process

```c
int main(){
  return 0;
}
```

```
0000000 cf fa ed fe 07 00 00 01 03 00 00 80 02 00 00 00
0000010 0f 00 00 00 38 03 00 00 85 00 20 00 00 00 00 00
0000020 19 00 00 00 48 00 00 00 5f 5f 50 41 47 45 5a 45
0000030 52 4f 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0000040 00 00 00 00 01 00 00 00 00 00 00 00 00 00 00 00
0000050 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0000060 00 00 00 00 00 00 00 00 19 00 00 00 38 01 00 00
0000070 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
0000080 00 00 00 00 01 00 00 00 00 10 00 00 00 00 00 00
0000090 00 00 00 00 00 00 00 00 00 10 00 00 00 00 00 00
00000a0 07 00 00 00 05 00 00 00 03 00 00 00 00 00 00 00
00000b0 5f 5f 74 65 78 74 00 00 00 00 00 00 00 00 00 00
00000c0 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
00000d0 98 0f 00 00 01 00 00 00 08 00 00 00 00 00 00 00
00000e0 98 0f 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000f0 00 04 00 80 00 00 00 00 00 00 00 00 00 00 00 00
0000100 5f 5f 75 6e 77 69 6e 64 5f 69 6e 66 6f 00 00 00
0000110 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
0000120 a0 0f 00 00 01 00 00 00 48 00 00 00 00 00 00 00
0000130 a0 0f 00 00 02 00 00 00 00 00 00 00 00 00 00 00
0000140 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0000150 5f 5f 65 68 5f 66 72 61 6d 65 00 00 00 00 00 00
0000160 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
0000170 e8 0f 00 00 01 00 00 00 18 00 00 00 00 00 00 00
0000180 e8 0f 00 00 03 00 00 00 00 00 00 00 00 00 00 00
0000190 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00001a0 19 00 00 00 48 00 00 00 5f 5f 4c 49 4e 4b 45 44
00001b0 49 54 00 00 00 00 00 00 00 10 00 00 01 00 00 00
00001c0 00 10 00 00 00 00 00 00 00 10 00 00 00 00 00 00
00001d0 d8 00 00 00 00 00 00 00 07 00 00 00 01 00 00 00
00001e0 00 00 00 00 00 00 00 00 22 00 00 80 30 00 00 00
00001f0 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
```

```c
int main(){
  return 0;
}
```

```
0000000 cf fa ed fe 07 00 00 01 03 00 00 80 02 00 00 00
0000010 0f 00 00 00 38 03 00 00 85 00 20 00 00 00 00 00
0000020 19 00 00 00 48 00 00 00 5f 5f 50 41 47 45 5a 45
0000030 52 4f 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0000040 00 00 00 00 01 00 00 00 00 00 00 00 00 00 00 00
0000050 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0000060 00 00 00 00 00 00 00 00 19 00 00 00 38 01 00 00
0000070 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
0000080 00 00 00 00 01 00 00 00 00 10 00 00 00 00 00 00
0000090 00 00 00 00 00 00 00 00 00 10 00 00 00 00 00 00
00000a0 07 00 00 00 05 00 00 00 03 00 00 00 00 00 00 00
00000b0 5f 5f 74 65 78 74 00 00 00 00 00 00 00 00 00 00
00000c0 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
00000d0 98 0f 00 00 01 00 00 00 08 00 00 00 00 00 00 00
00000e0 98 0f 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00000f0 00 04 00 80 00 00 00 00 00 00 00 00 00 00 00 00
0000100 5f 5f 75 6e 77 69 6e 64 5f 69 6e 66 6f 00 00 00
0000110 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
0000120 a0 0f 00 00 01 00 00 00 48 00 00 00 00 00 00 00
0000130 a0 0f 00 00 02 00 00 00 00 00 00 00 00 00 00 00
0000140 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
0000150 5f 5f 65 68 5f 66 72 61 6d 65 00 00 00 00 00 00
0000160 5f 5f 54 45 58 54 00 00 00 00 00 00 00 00 00 00
0000170 e8 0f 00 00 01 00 00 00 18 00 00 00 00 00 00 00
0000180 e8 0f 00 00 03 00 00 00 00 00 00 00 00 00 00 00
0000190 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
00001a0 19 00 00 00 48 00 00 00 5f 5f 4c 49 4e 4b 45 44
00001b0 49 54 00 00 00 00 00 00 00 10 00 00 01 00 00 00
00001c0 00 10 00 00 00 00 00 00 00 10 00 00 00 00 00 00
00001d0 d8 00 00 00 00 00 00 00 07 00 00 00 01 00 00 00
00001e0 00 00 00 00 00 00 00 00 22 00 00 80 30 00 00 00
00001f0 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00
```


Lexer
Parser
Semantic Analysis
Code Generation
Optimization
Assembler
Linker

```
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

Lexer

```
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
(KW ‘const’)
```

```
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
(KW ‘const’), (TYPE ‘float’)
```

```
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
(KW ‘const’), (TYPE ‘float’), (ID ‘factor'),
(EQ ‘=‘), (NUM ’42.f’), (SEMI ‘;’)
```

```
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

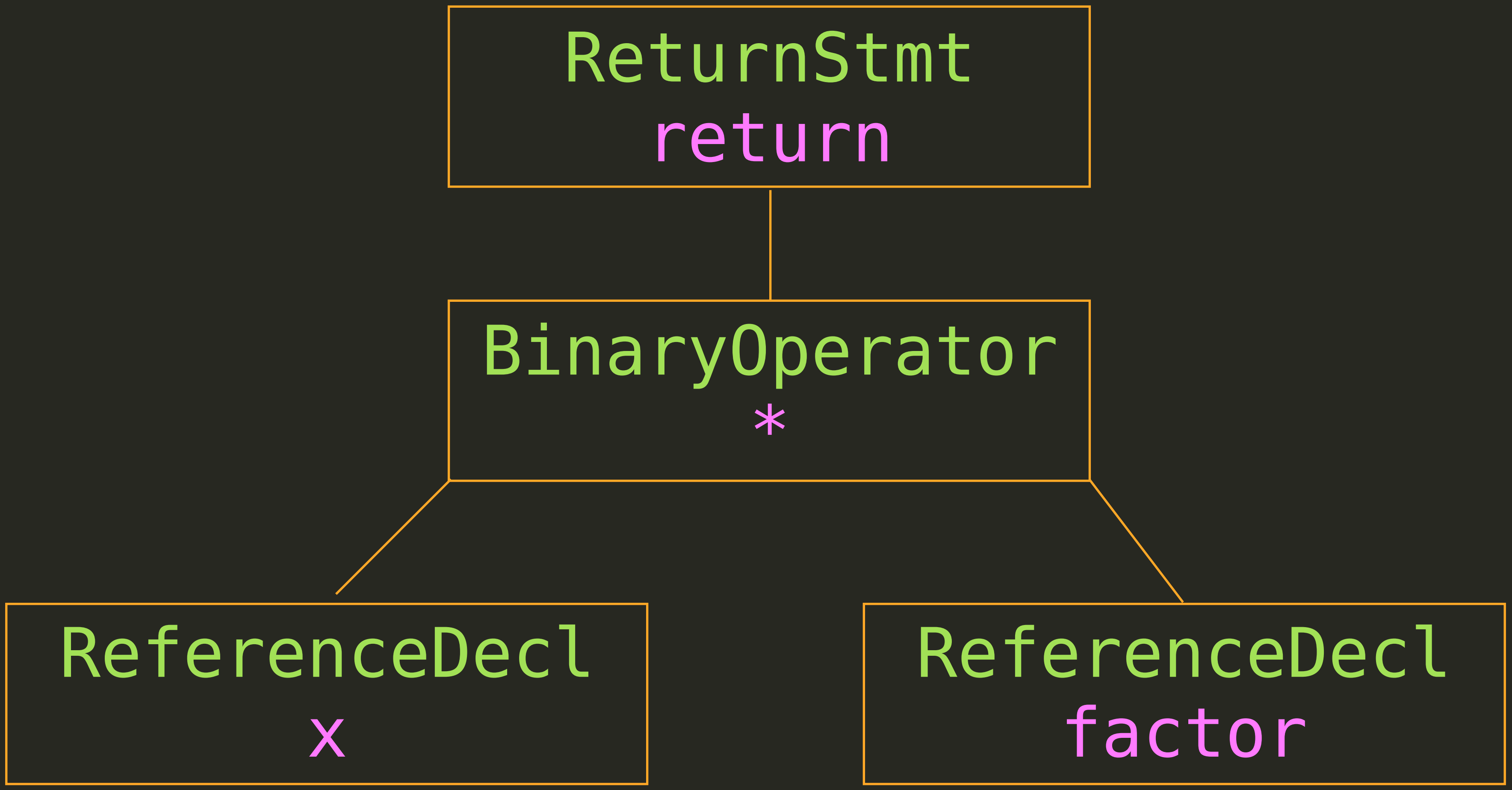
$\downarrow$

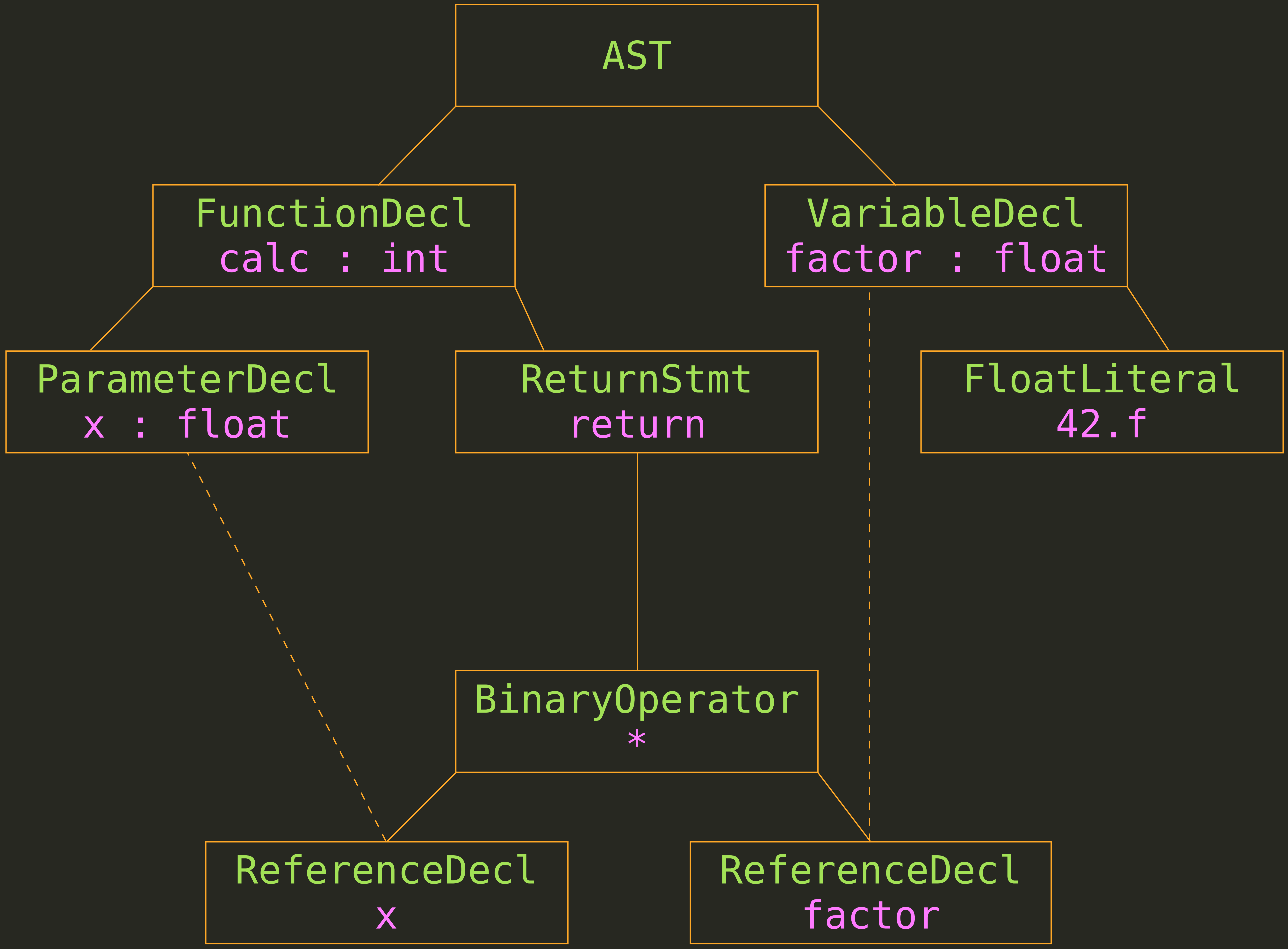
```
(KW ‘const’), (TYPE ‘float’), (ID ‘factor'),
(EQ ‘=‘), (NUM ’42.f’), (SEMI ‘;’), (TYPE ‘int’),
(ID ‘calc’), (L_PAREN ‘(‘), (TYPE ‘float’), (ID 'x')
(R_PAREN ‘)’), (L_BRACE ‘{‘), (KW ‘return’),
(ID ‘factor’), (STAR ‘*’), (ID ‘x’), (SEMI ‘;’),
(R_BRACE ‘}’), (EOF '')
```

Parser

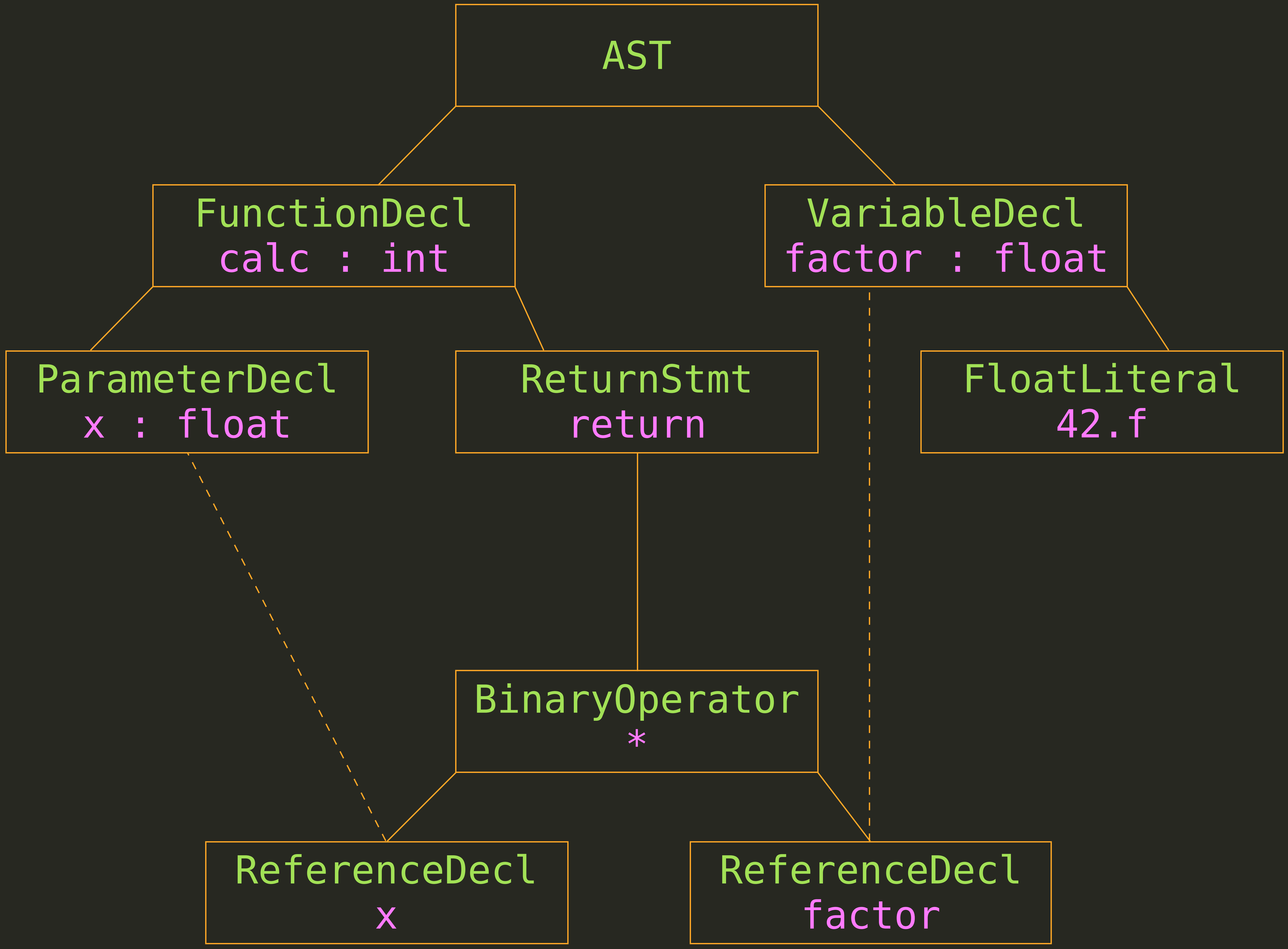
```
(KW ‘return’) (ID ‘factor’) (STAR ‘*’) (ID ‘x’)
```

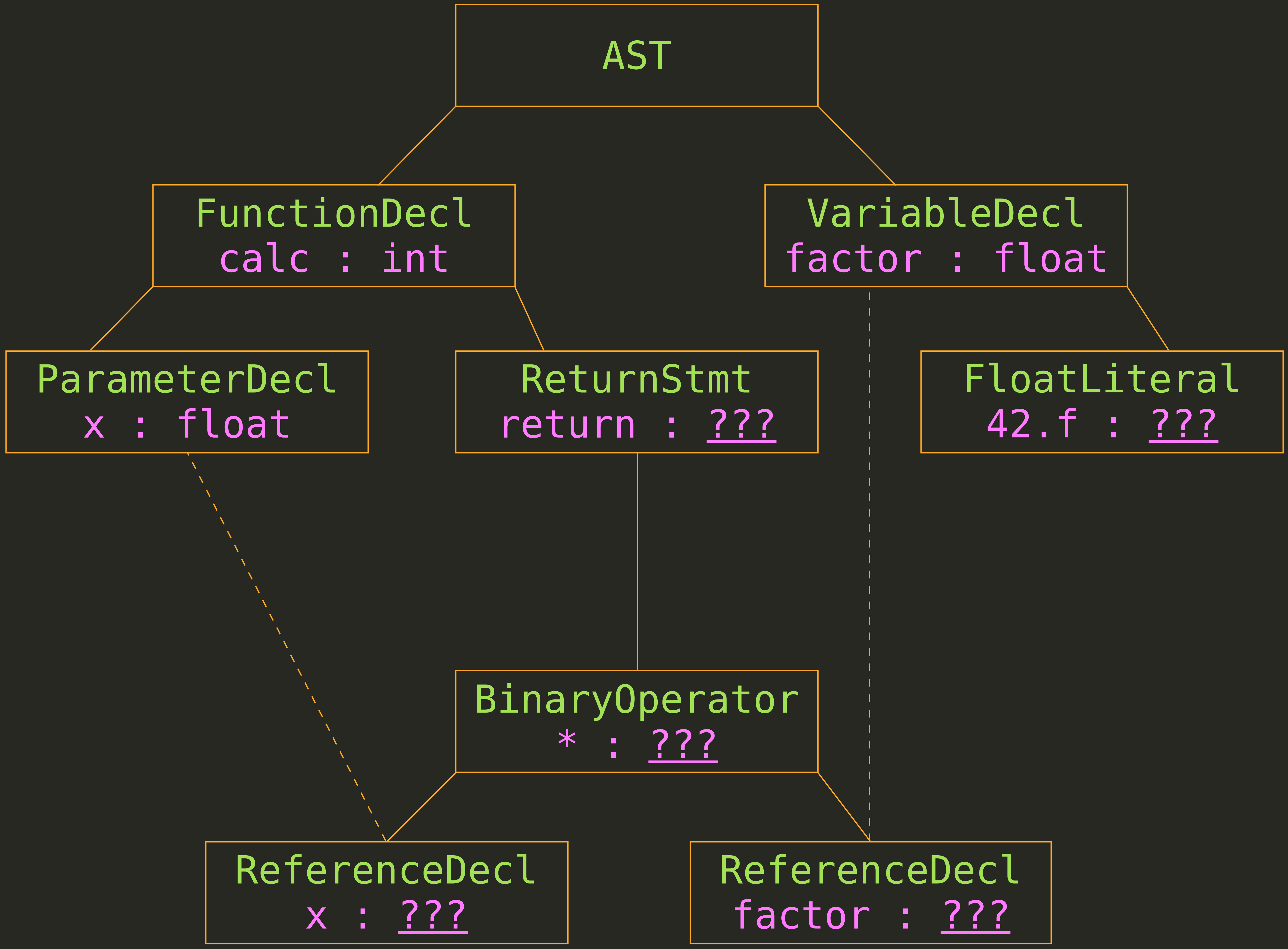
```
(KW ‘return’) (ID ‘factor’) (STAR ‘*’) (ID ‘x’)
```

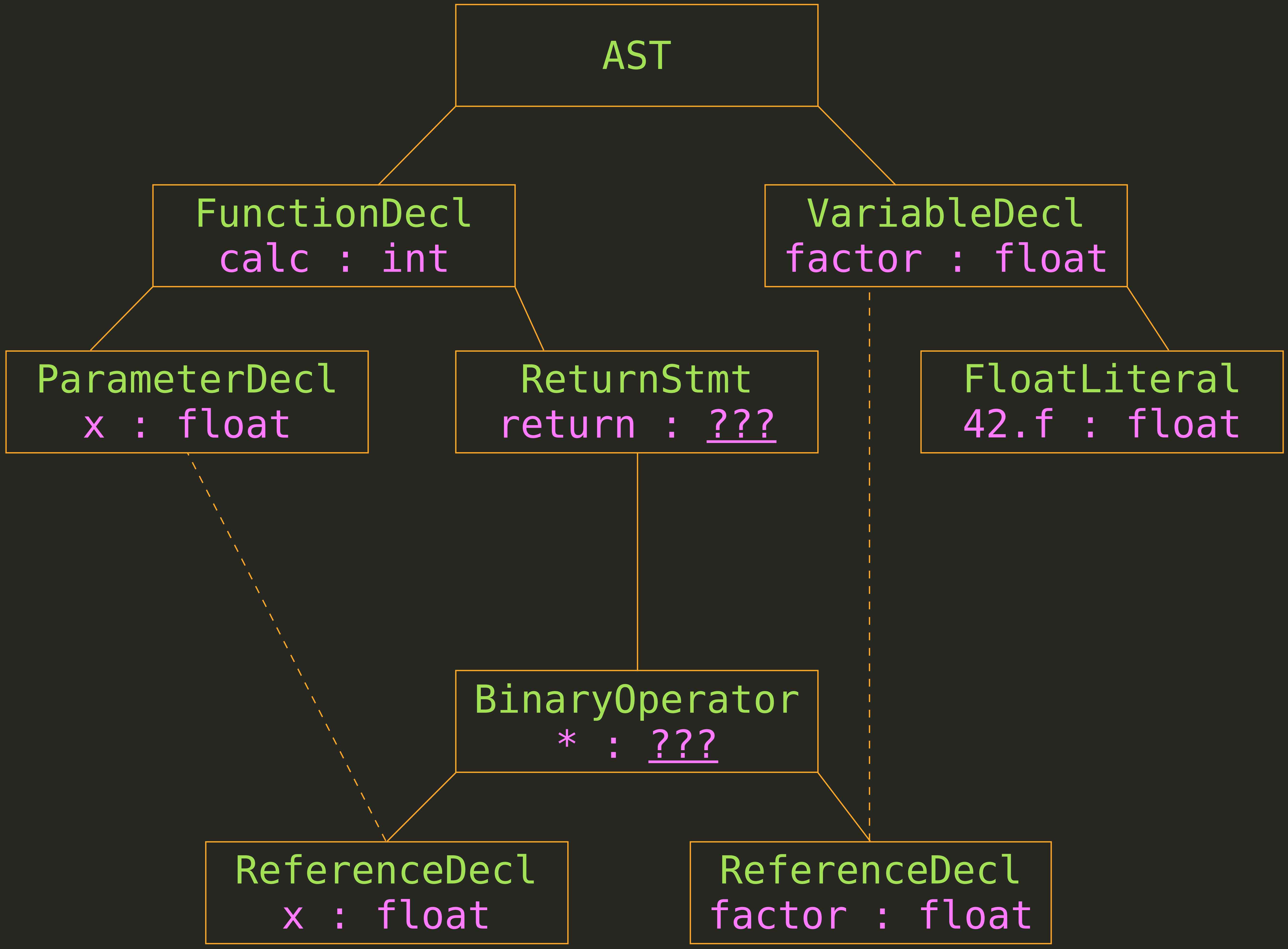


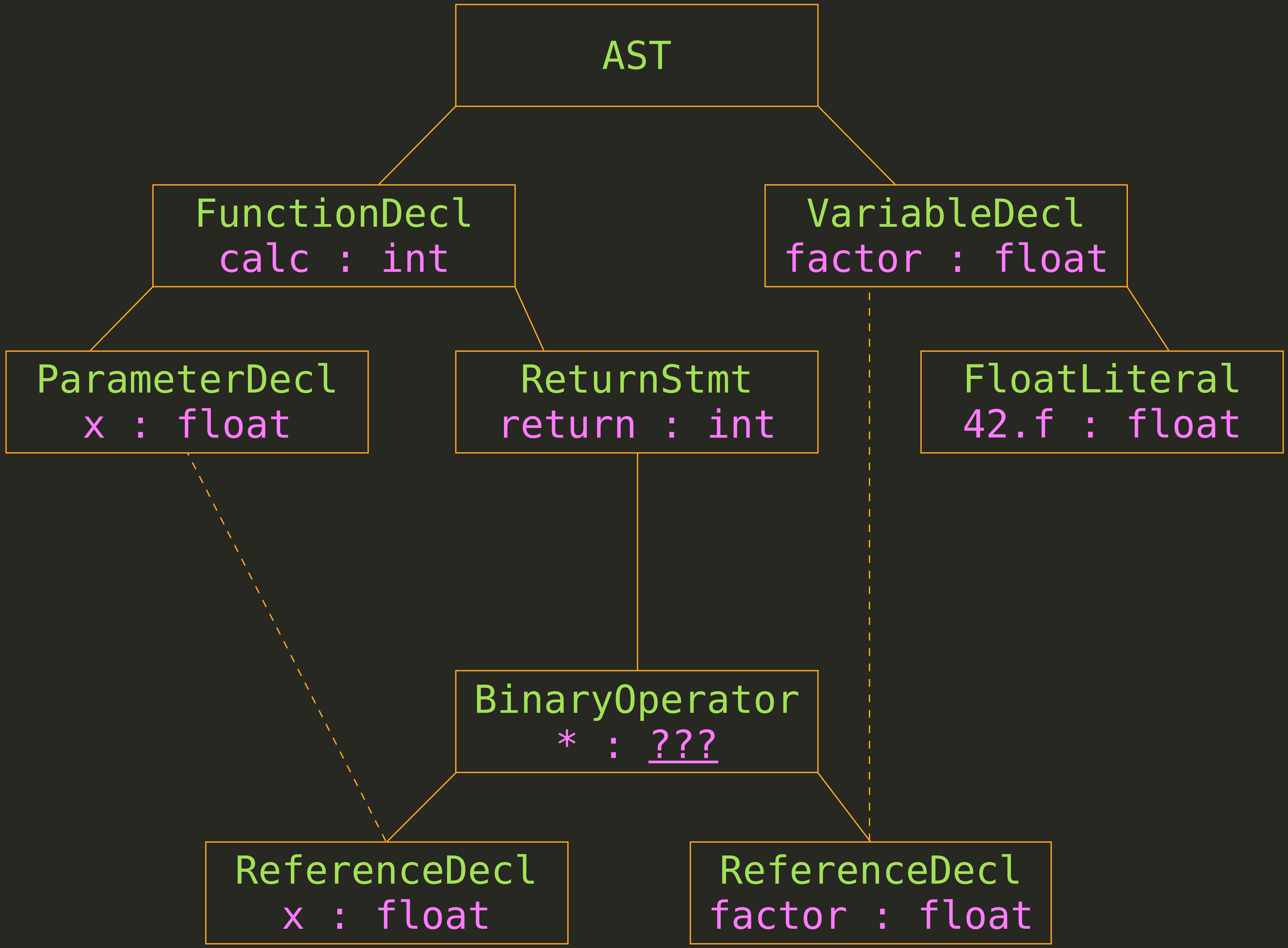

AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return
FloatLiteral
42.f
BinaryOperator
*
ReferenceDecl
x
ReferenceDecl
factor

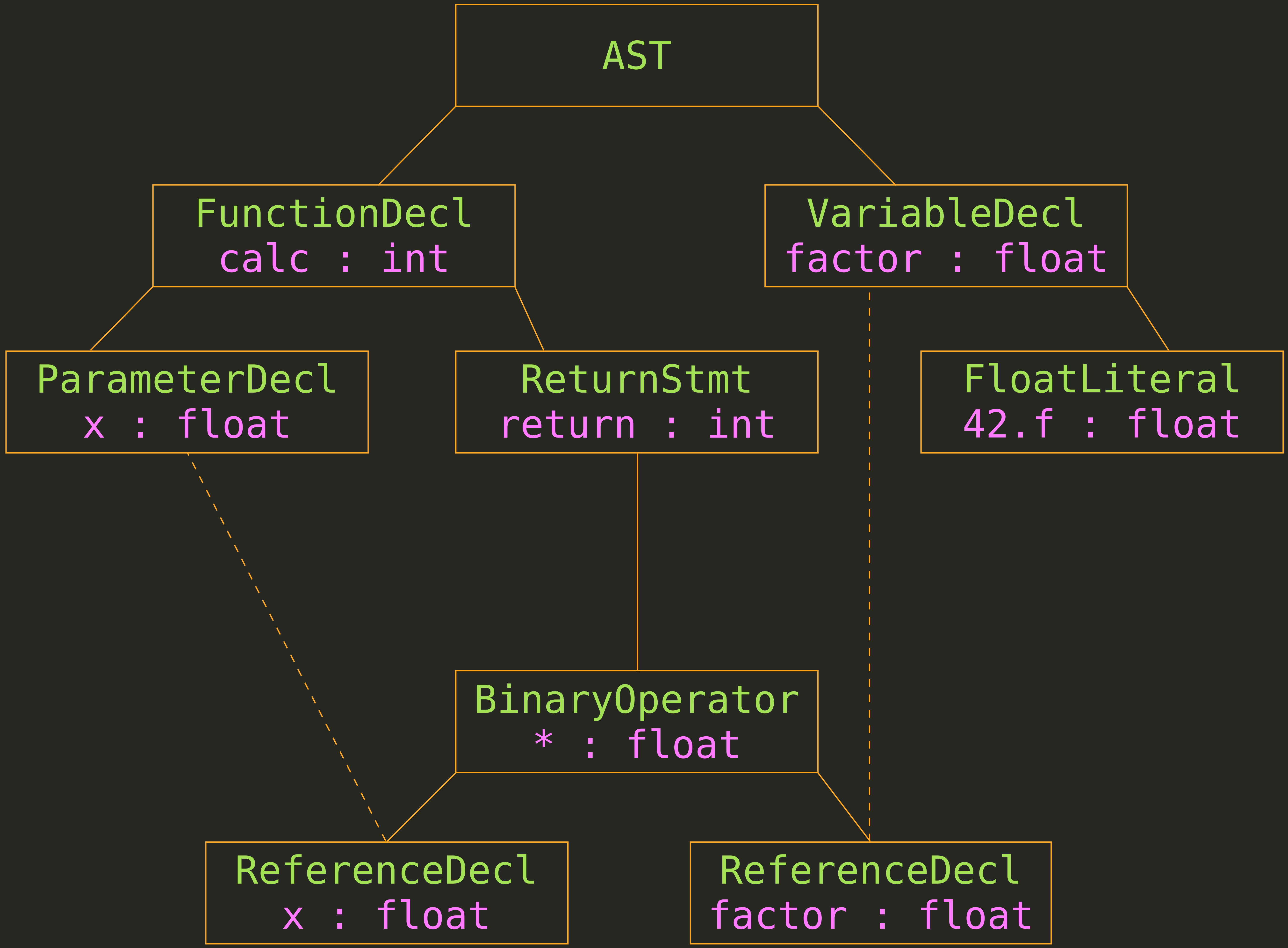
Semantic Analysis

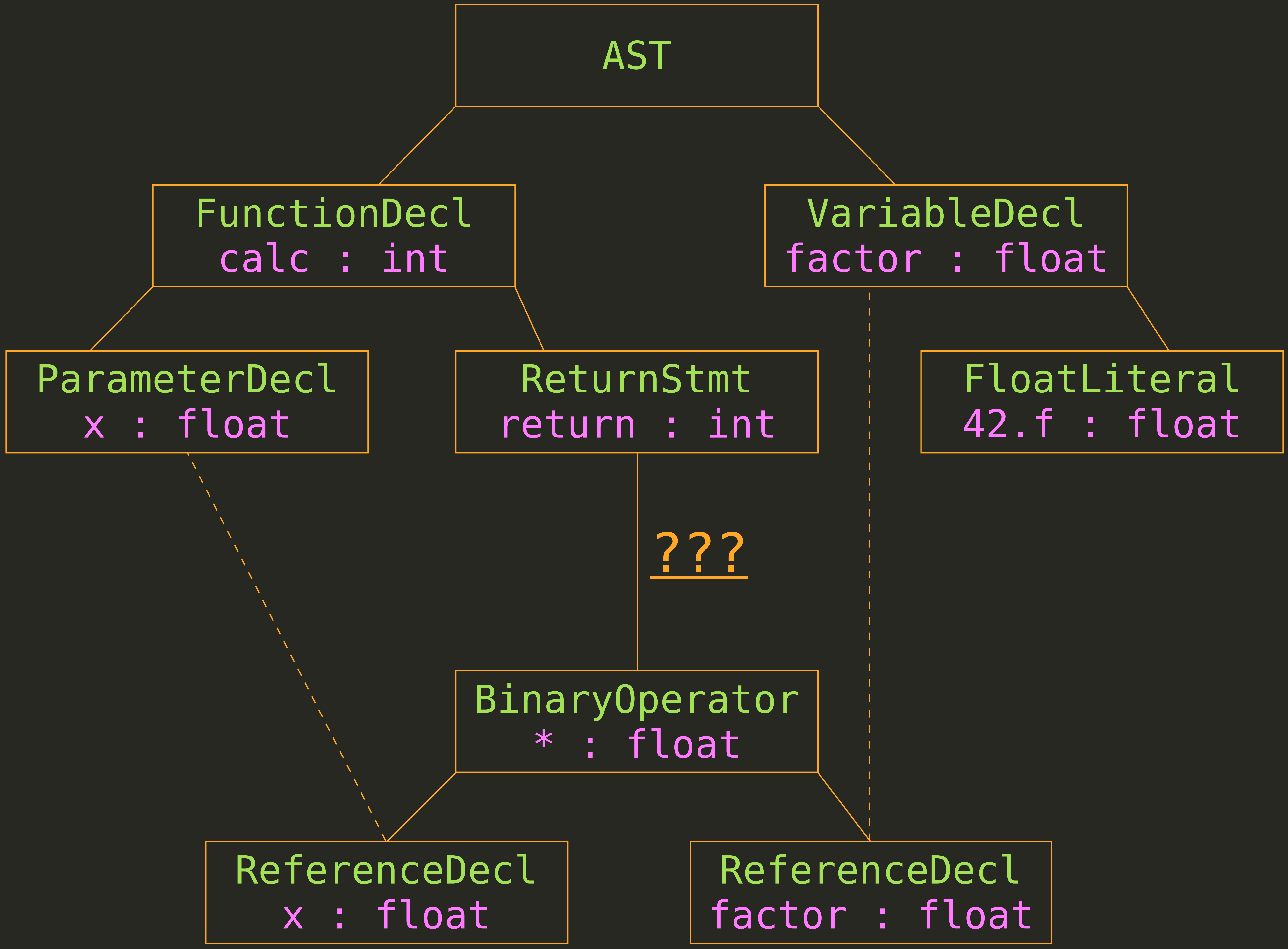

AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return
FloatLiteral
42.f
BinaryOperator
*
ReferenceDecl
x
ReferenceDecl
factor

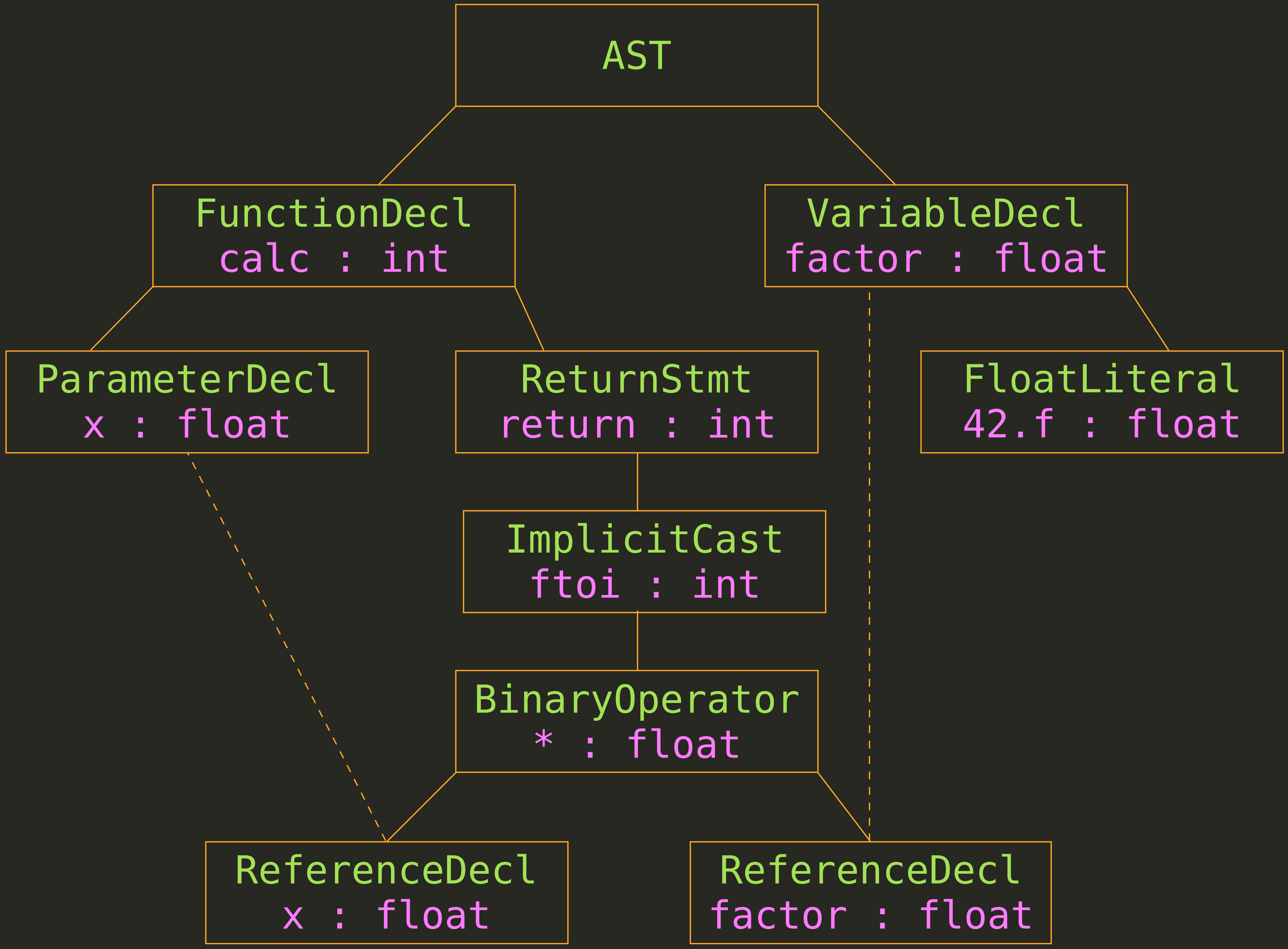

AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return : ???
FloatLiteral
42.f : ???
BinaryOperator
* : ???
ReferenceDecl
x : ???
ReferenceDecl
factor : ???


AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return : ???
FloatLiteral
42.f : float
BinaryOperator
* : ???
ReferenceDecl
x : float
ReferenceDecl
factor : float


AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return : int
FloatLiteral
42.f : float
BinaryOperator
* : ???
ReferenceDecl
x : float
ReferenceDecl
factor : float


AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return : int
FloatLiteral
42.f : float
BinaryOperator
* : float
ReferenceDecl
x : float
ReferenceDecl
factor : float


AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return : int
FloatLiteral
42.f : float
???
BinaryOperator
* : float
ReferenceDecl
x : float
ReferenceDecl
factor : float


AST
FunctionDecl
calc : int
VariableDecl
factor : float
ParameterDecl
x : float
ReturnStmt
return : int
FloatLiteral
42.f : float
ImplicitCast
ftoi : int
BinaryOperator
* : float
ReferenceDecl
x : float
ReferenceDecl
factor : float

Code Generation

```
@factor = constant float 42.0

define calc(float %x) {
entry:
  movf %x, %r1
  movf @factor, %r2
  %r3 = fmul %r1, %r2
  movf %r3, %r0
  ret
}
```

Optimization

```
@factor = constant float 42.0

define calc(float %x) {
entry:
  movf %x, %r1
  movf @factor, %r2
  %r3 = fmul %r1, %r2
  movf %r3, %r0
  ret
}
```

```
@factor = constant float 42.0

define calc(float %x) {
entry:
  %r0 = fmul @factor, %x
  ret
}
```

Assembler

```
_calc:
    push {r7, lr}
    mov  r7, sp
    mov  r1, #36175872
    orr  r1, r1, #1073741824
    bl ___mulsf3
    bl ___fixsfsi
    pop  {r7, lr}
    mov  pc, lr

    .section  __TEXT,__const
    .globl  _factor @ @factor
    .align 2
_factor:
    .long  1109917696 @ float 42
```

Linker

```c
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
 > clang -c calc.c -o calc.o
```

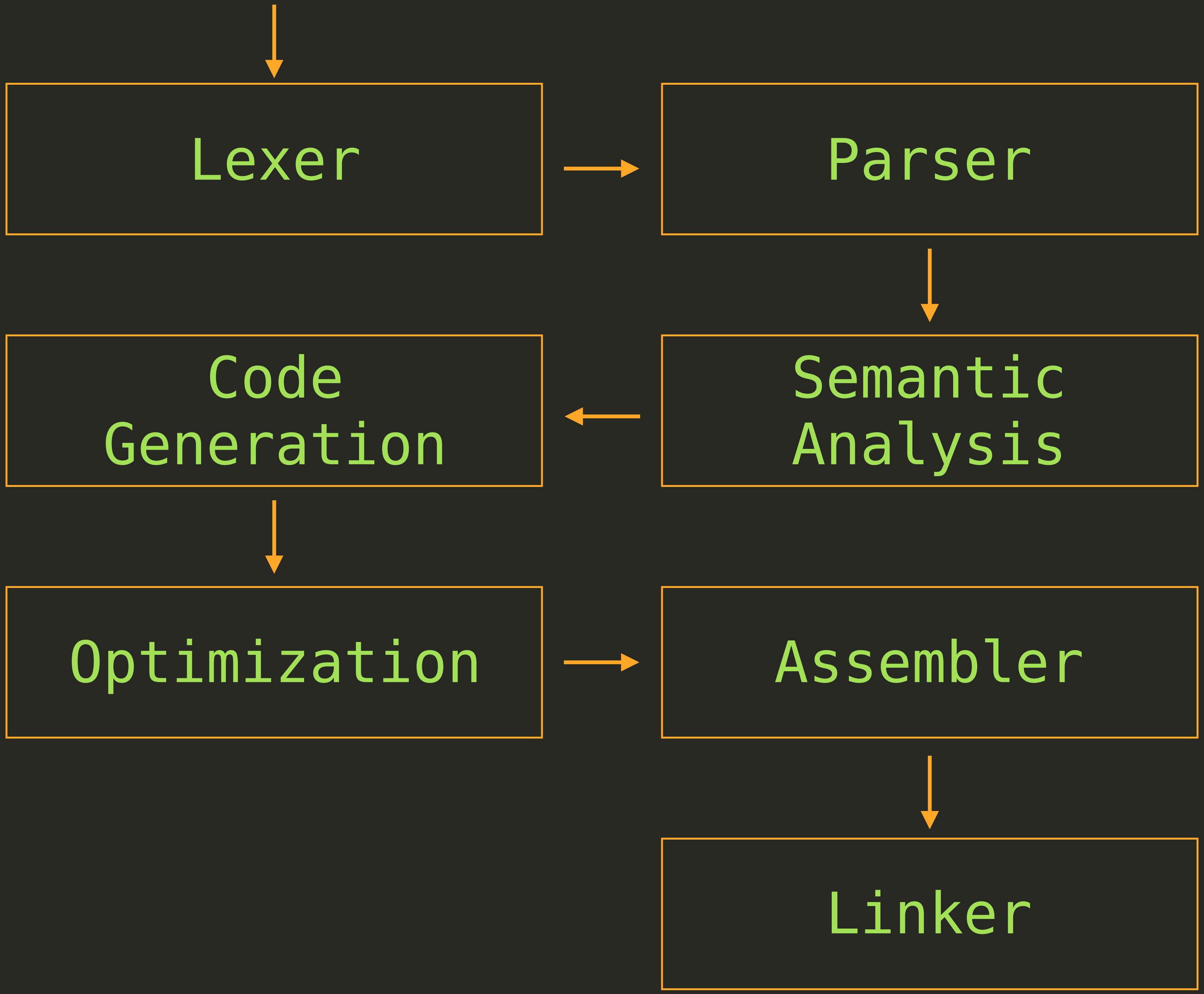
```c
extern int calc(float);

int main() {
  printf(“%d\n”, calc(2.f));
  return 0;
}
```

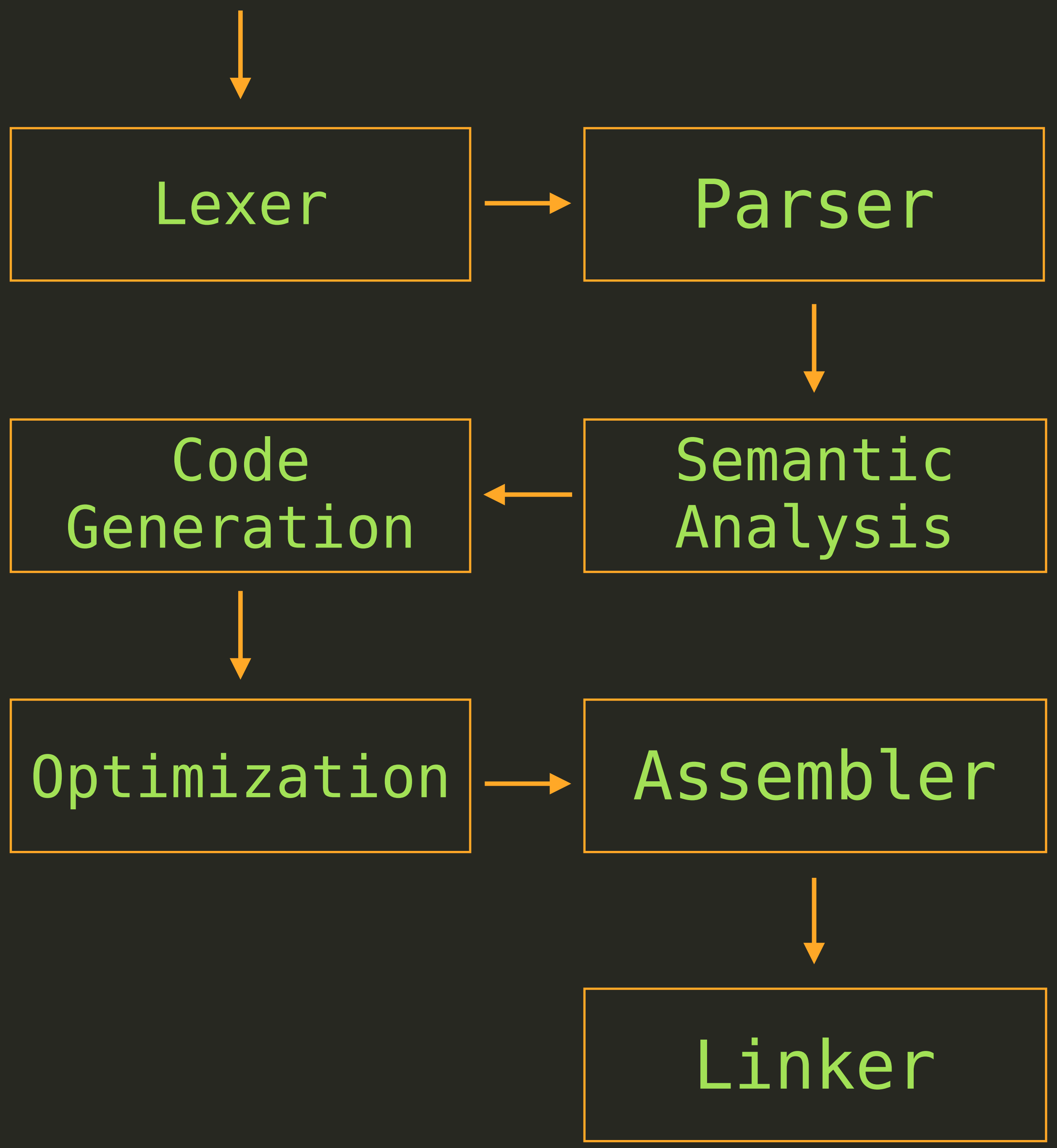
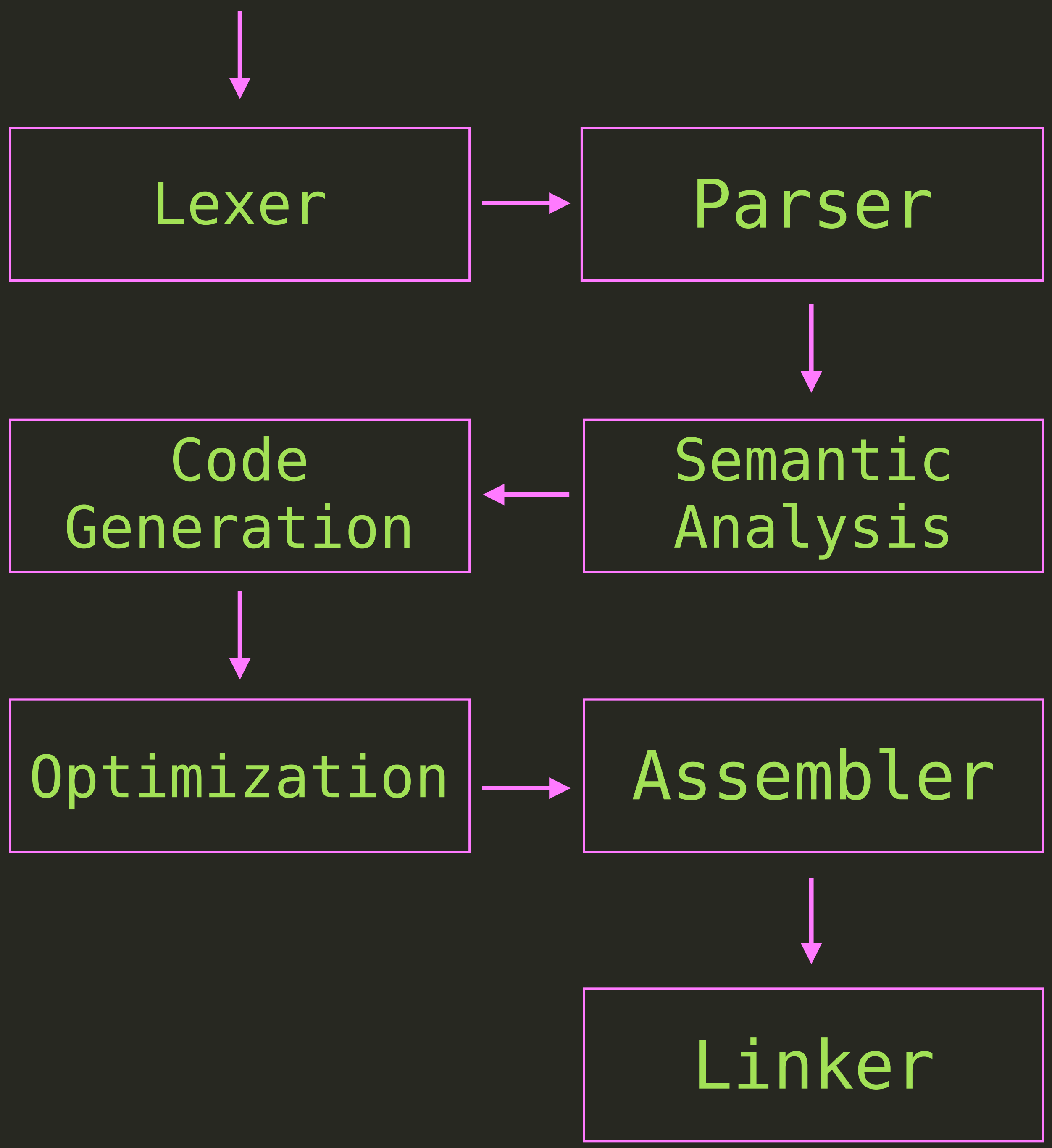
```
  > clang -c main.c -o main.o
```

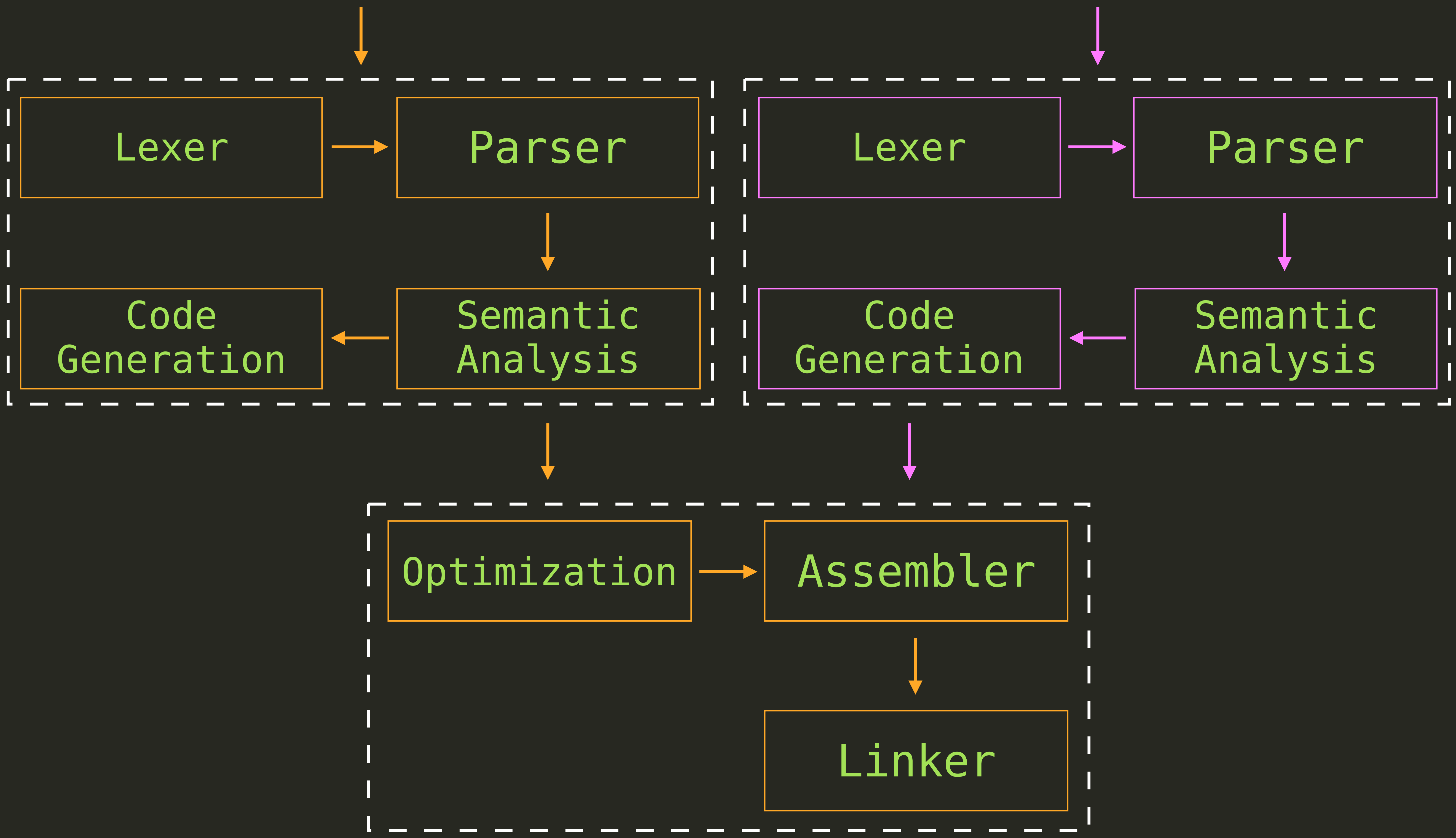
```
> nm main.o
                 U _calc
0000000000000000 T _main
                 U _printf
```

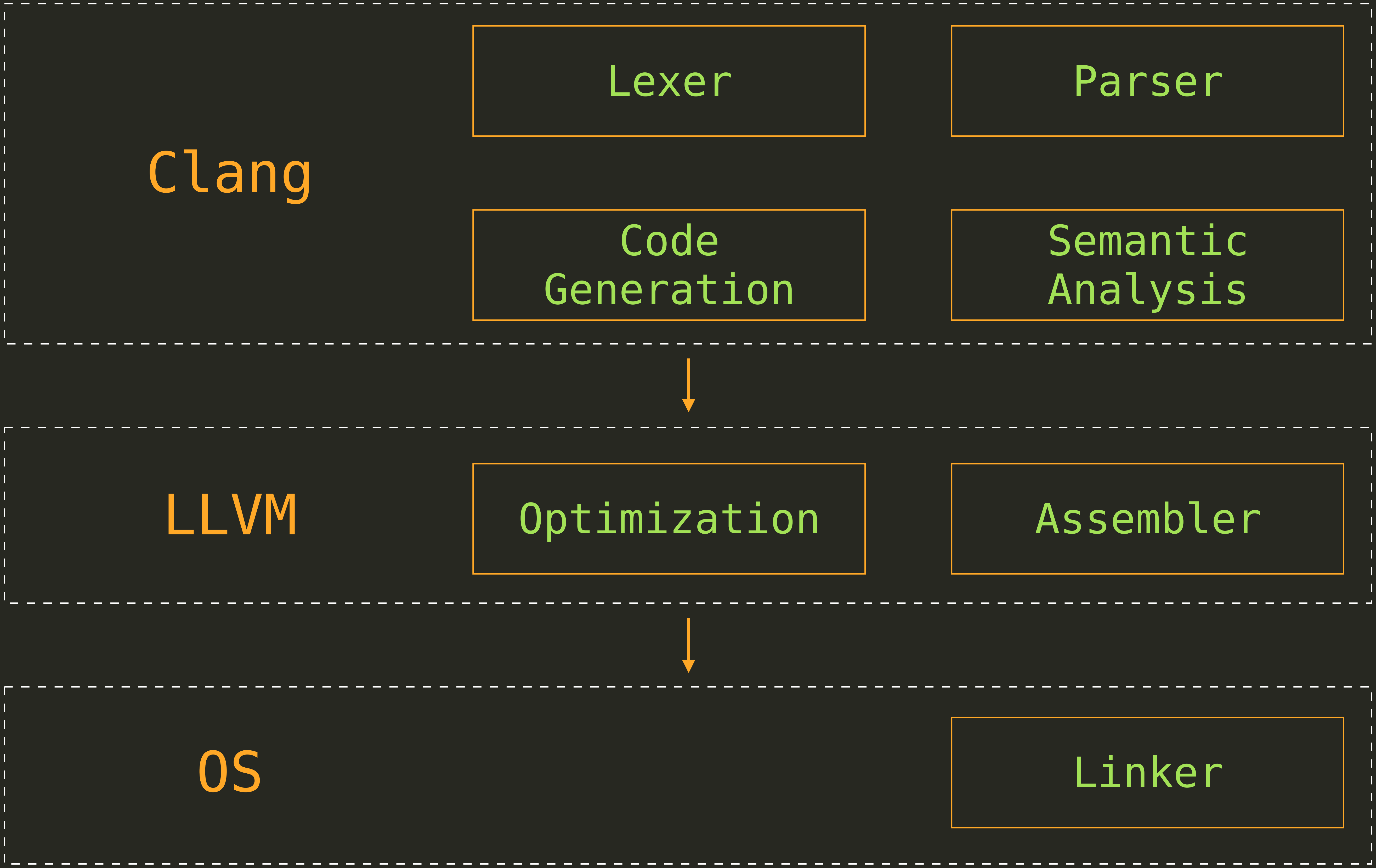
```
> nm main.o
                 U _calc
0000000000000000 T _main
                 U _printf

> ld -lc calc.o main.o -o main

> nm main
0000000000001f30 T _calc
0000000000001fc8 S _factor
0000000000001f60 T _main
                 U _printf
```

LLVM & Clang


Lexer
Parser
Semantic Analysis
Code Generation
Optimization
Assembler
Linker


Lexer
Parser
Semantic Analysis
Code Generation
Optimization
Assembler
Linker

Lexer
Parser
Semantic Analysis
Code Generation
Optimization
Assembler
Linker


Lexer
Parser
Semantic Analysis
Code Generation
Lexer
Parser
Semantic Analysis
Code Generation
Optimization
Assembler
Linker


Clang
Lexer
Parser
Code Generation
Semantic Analysis
LLVM
Optimization
Assembler
OS
Linker

libclangLex

```c
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
> clang -cc1 -dump-tokens calc.c
```

```
const 'const'    [StartOfLine]  Loc=<calc.c:1:1>
float 'float'    [LeadingSpace] Loc=<calc.c:1:7>
identifier 'factor' [LeadingSpace] Loc=<calc.c:1:13>
equal '='  [LeadingSpace] Loc=<calc.c:1:20>
numeric_constant '42.f'    [LeadingSpace] Loc=<calc.c:1:22>
semi ';'     Loc=<calc.c:1:26>
int 'int'  [StartOfLine]  Loc=<calc.c:3:1>
identifier 'calc'   [LeadingSpace] Loc=<calc.c:3:5>
l_paren '('     Loc=<calc.c:3:9>
float 'float'      Loc=<calc.c:3:10>
identifier 'x'   [LeadingSpace] Loc=<calc.c:3:16>
r_paren ')'     Loc=<calc.c:3:17>
l_brace '{'   [LeadingSpace] Loc=<calc.c:3:19>
return 'return'  [StartOfLine] [LeadingSpace]   Loc=<calc.c:4:3>
identifier 'factor' [LeadingSpace] Loc=<calc.c:4:10>
star '*'   [LeadingSpace] Loc=<calc.c:4:17>
identifier 'x'   [LeadingSpace] Loc=<calc.c:4:19>
semi ';'     Loc=<calc.c:4:20>
r_brace '}'   [StartOfLine]  Loc=<calc.c:5:1>
eof ''    Loc=<calc.c:6:1>
```

libclangParse/
libclangSema

```c
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
> clang -cc1 -ast-dump calc.c
```

```
TranslationUnitDecl <<invalid sloc>> <invalid sloc>
|-VarDecl <calc.c:1:1, col:22> col:13 used factor 'const float' cinit
| `-FloatingLiteral <col:22> 'float' 4.200000e+01
`-FunctionDecl <line:3:1, line:5:1> line:3:5 calc 'int (float)'
  |-ParmVarDecl <col:10, col:16> col:16 used x 'float'
  `-CompoundStmt <col:19, line:5:1>
    `-ReturnStmt <line:4:3, col:19>
      `-ImplicitCastExpr <col:10, col:19> 'int' <FloatingToIntegral>
        `-BinaryOperator <col:10, col:19> 'float' '*'
          |-ImplicitCastExpr <col:10> 'float' <LValueToRValue>
          | `-DeclRefExpr <col:10> 'const float' lvalue Var 'factor' 'const float'
          `-ImplicitCastExpr <col:19> 'float' <LValueToRValue>
            `-DeclRefExpr <col:19> 'float' lvalue ParmVar 'x' 'float'
```

libclangCodeGen

```c
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
> clang -S -emit-llvm calc.c
```

```
@factor = constant float 4.200000e+01, align 4

define i32 @calc(float %x) #0 {
entry:
  %x.addr = alloca float, align 4
  store float %x, float* %x.addr, align 4
  %0 = load float* %x.addr, align 4
  %mul = fmul float 4.200000e+01, %0
  %conv = fptosi float %mul to i32
  ret i32 %conv
}
```

libclangCodeGen + opt

```c
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
> clang -S -emit-llvm -O1 calc.c
```

```
@factor = constant float 4.200000e+01, align 4

define i32 @calc(float %x) #0 {
entry:
  %mul = fmul float %x, 4.200000e+01
  %conv = fptosi float %mul to i32
  ret i32 %conv
}
```

libLLVMAsmPrinter

```c
const float factor = 42.f;

int calc(float x) {
    return factor * x;
}
```

```
> clang -S -arch arm -O0 calc.c
```

```
_calc:
   push {r7, lr}
   mov  r7, sp
   mov  r1, #36175872
   orr  r1, r1, #1073741824
   bl   ___mulsf3
   bl   ___fixsfsi
   pop  {r7, lr}
   mov  pc, lr

   .section  __TEXT,__const
   .globl _factor @ @factor
   .align 2
_factor:
   .long  1109917696 @ float 42
```

Summary

Summary

- Learn your tools

Summary

- Learn your tools
- Provide feedback, don’t make complaints

Summary

- Learn your tools
- Provide feedback, don’t make complaints
- Give back to community

Questions?

Twitter:
@1101_debian

Slides:
https://speakerdeck.com/alexdenisov/compilation-process