Basic Elements of Programming Languages

Ronghui Gu

Spring 2019

Columbia University

* Course website: https://www.cs.columbia.edu/ rgu/courses/4115/spring2019

** These slides are borrowed from Prof. Edwards.

What is a Programming Language?

A programming language is a notation that a person and a computer can both understand.

- It allows you to express what is the **task** to compute
- It allows a computer to **execute** the computation task

Language Specifications

How to Define a Language

When designing a language, it's a good idea to start by sketching forms that you want to appear in your language as well as forms you do not want to appear.

```
int avg(int a, int b)
{
  return (a + b) / 2;
}
```

Examples

```
a int vg(int a,
{
  return (a; + b)
{ {
```

Non-Examples

How to Define a Language

- An official documents, with **informal** descriptions.
- An official documents, with **formal** descriptions.
- A reference implementation, e.g., a compiler.

Some language definitions are sanctioned by an official standards organization, e.g., C11 (ISO/IEC 9899:2011).

```c
int compare()
{
  int a[10], b[10];
  if (a > b)
    return true;
  return false;
}
```

Aspects of Language Specifications

Syntax

Semantics

Pragmatics

- **Syntax**: how characters combine to form a program.
- **Semantics**: what the program *means*.
- **Pragmatics**: common programming idioms; programming environments; the standard library; ecosystems.

Syntax

Syntax is divided into:

- **Microsyntax**: specifies how the characters in the source code stream are grouped into tokens.
- **Abstract syntax**: specifies how the tokens are grouped into phrases, e.g., expressions, statements, etc.

Microsytax

Source program is just a sequence of characters.

```
int avg(int a, int b)
{
  return (a + b) / 2;
}
```

i n t SP a v g (i n t SP a , SP i n t SP b) NL
{ NL
SP SP r e t u r n SP (a SP + SP b) SP / SP 2 ; NL
} NL

Microsytax

```
int avg(int a, int b)
{
  return (a + b) / 2;
}
```

Token	**Lexemes**	**Pattern (as regular expressions)**
ID	avg, a, b	letter followed by letters or digits
KEYWORD	int, return	letters
NUMBER	2	digits
OPERATOR	+, /	+, /
PUNCTUATION	;,(,),{,},	;,(,),{,},

int avg (int a , int b) { return (a + b

) / 2 ; }

Lexical Analysis Gives Tokens

```
int avg(int a, int b)
{
  return (a + b) / 2;
}
```

`int` `avg` `(` `int` `a` `,` `int` `b` `)` `{` `return` `(` `a` `+` `b` `)` `/` `2` `;` `}`

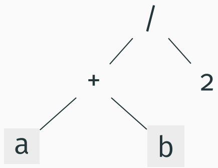
- Throw errors when failing to create tokens: malformed numbers (e.g., 23f465#g) or invalid characters (such as non-ASCII characters in C).

Abstract Syntax

Abstract Syntax can be defined using Context Free Grammar.

```
expr :
    expr OPERATOR expr
  | ( expr )
  | NUMBER
```

Expression $(a + b)/2$ can be parsed into an AST:

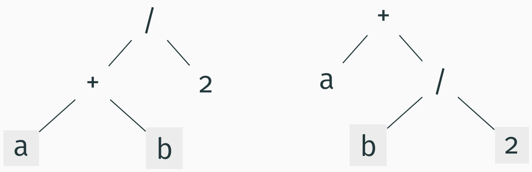


Abstract Syntax

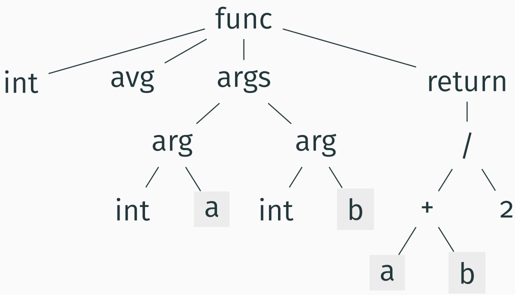
Abstract Syntax can be defined using Context Free Grammar.

```
expr :
    expr OPERATOR expr
  | ( expr )
  | NUMBER
```

Ambiguous! What about $a + b/2$?



Syntax Analysis Gives an Abstract Syntax Tree




```
int avg(int a, int b)
{
  return (a + b) / 2;
}
```

- Syntax analysis will throw errors if “}” is missing. Lexical analysis will not.

Semantics

- **Static Semantics**: deals with legality rules—things you can check before running the code (compile time), e.g., type, scope, for some languages.
- **Dynamic Semantics**: deals with the execution behavior; things that can only be known at runtime, e.g., value.

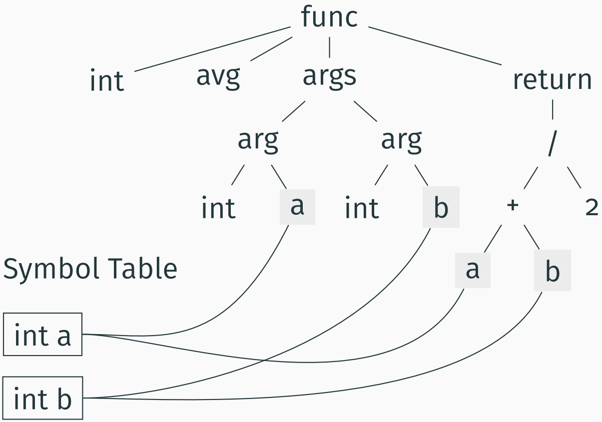
Static Semantics

We can use inference rules to define semantics, e.g., type:

$$\frac{}{\text{NUMBER} \; : \; \textbf{int}} \qquad \frac{\text{expr} \; : \; \textbf{int}}{(\text{expr}) \; : \; \textbf{int}}$$

$$\frac{\text{expr}_1 \; : \; \textbf{int} \quad \text{expr}_2 \; : \; \textbf{int}}{\text{expr}_1 \text{ OPERATOR } \text{expr}_2 \; : \; \textbf{int}}$$

Semantic Analysis: Resolve Symbols; Verify Types



We can use inference rules to define semantics, e.g., value:

$$\frac{}{\textbf{eval}(\text{NUMBER}) = \text{NUMBER}} \qquad \frac{\textbf{eval}(\text{expr}) = n}{\textbf{eval}((\text{expr})) = n}$$

$$\frac{\textbf{eval}(\text{expr}_1) = n_1 \quad \textbf{eval}(\text{expr}_2) = n_2 \quad (n_1 + n_2) = n}{\textbf{eval}(\text{expr}_1 \ + \ \text{expr}_2) = n}$$

Dynamic Semantics

Consider the integer range:

$$\frac{\text{wrap}(\text{NUMBER}) = n}{\textbf{eval}(\text{NUMBER}) = n} \qquad \frac{\textbf{eval}(\text{expr}) = n}{\textbf{eval}((\text{expr})) = n}$$

$$\frac{\textbf{eval}(\text{expr}_1) = n_1 \quad \textbf{eval}(\text{expr}_2) = n_2 \quad \text{wrap}(n_1 + n_2) = n}{\textbf{eval}(\text{expr}_1 \ + \ \text{expr}_2) = n}$$

Programming Paradigms

Programming Paradigms

A programming paradigm is a style, or "way," of programming. Some languages make it easy to write in some paradigms but not others.

Imperative Programming

An imperative program specifies how a computation is to be done: a sequence of statements that update state.

```
    result = []
    i = 0
    numStu = length(students)
start:
    if i >= numStu goto finished
    name = students[i]
    nameLength = length(name)
    if nameLength <= 5 goto nextOne
    addToList(result, name)
nextOne:
    i = i + 1
    goto start
finished:
    return result
```

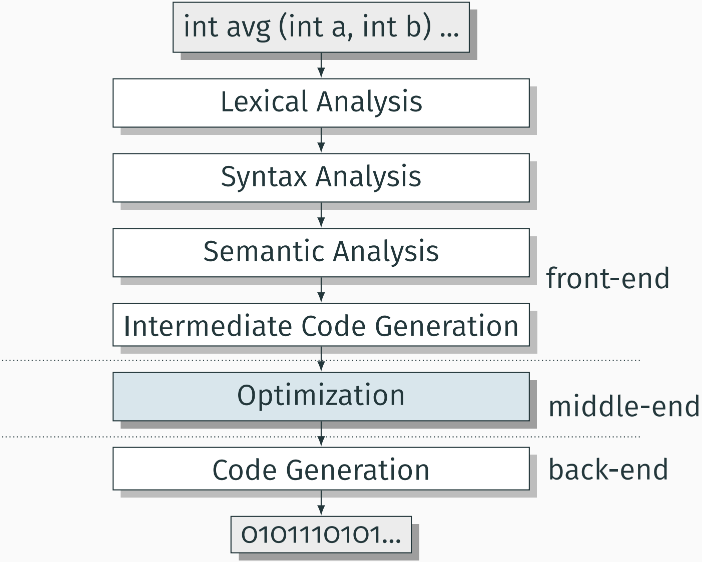
Structured Programming

Programming with clean, goto-free, nested control structures.

Go To Statement Considered Harmful by Dijkstra.

```
    result = []
    i = 0
    numStu = length(students)
start:
    if i >= numStu goto finished
    name = students[i]
    nameLength = length(name)
    if nameLength <= 5 goto nextOne
    addToList(result, name)
nextOne:
    i = i + 1
    goto start
finished:
    return result
```

Optimization



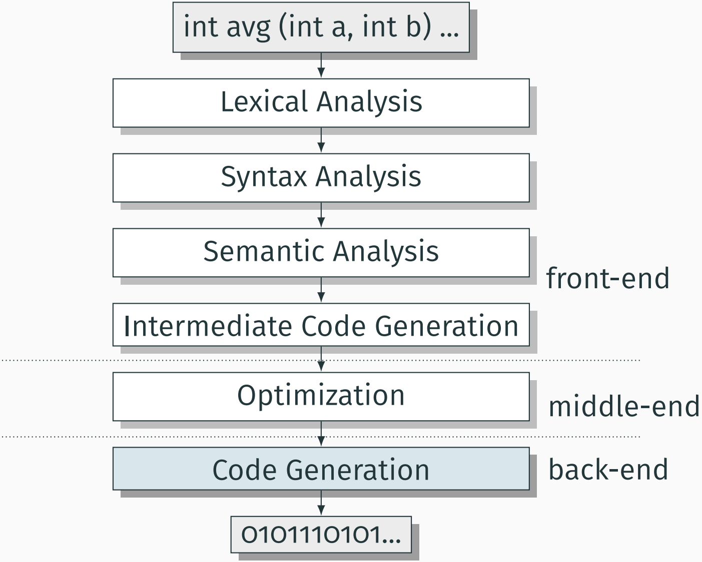
Optimization

```
avg:
  t0 := a + b
  t1 := 2
  t2 := t0 / t1
  ret t2
```

Optimization

```
avg:
  t0 := a + b
  t2 := t0 / 2
  ret t2
```

Code Generation



Generation of x86 Assembly

```
avg:
  t0 := a + b
  t2 := t0 / 2
  ret t2
```

Code Generation

```
avg:    pushl %ebp              # save BP
        movl  %esp,%ebp
        movl  8(%ebp),%eax      # load a from stack
        movl  12(%ebp),%edx     # load b from stack

        addl  %edx,%eax         # a += b
        shr   $1,%eax           # a /= 2
        ret
```