CS846
Machine Learning for Software Engineering

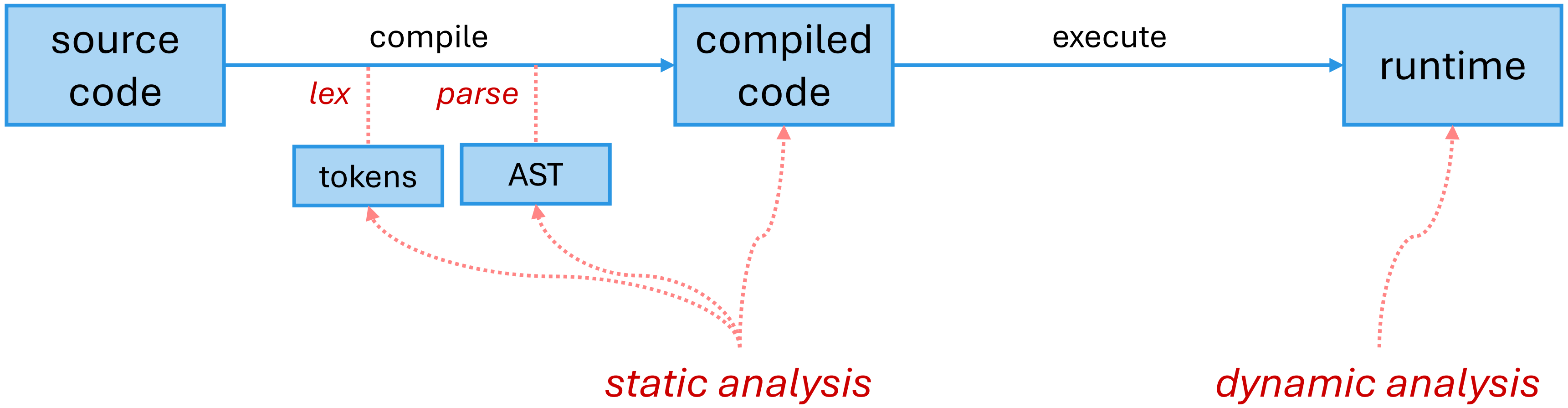
Pengyu Nie

Parsing and Build Systems

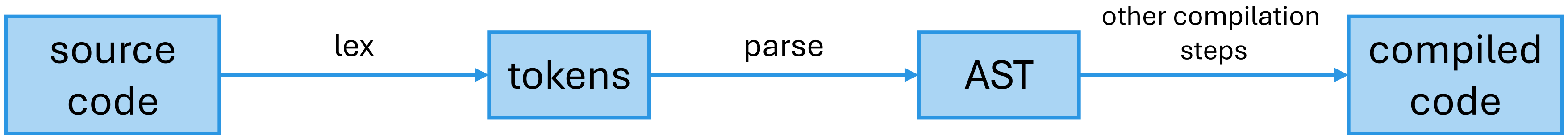
Context-free grammar & parser generators

Build system

Program Analyses Overview



Lexing and Parsing


source code

lex

tokens

parse

AST

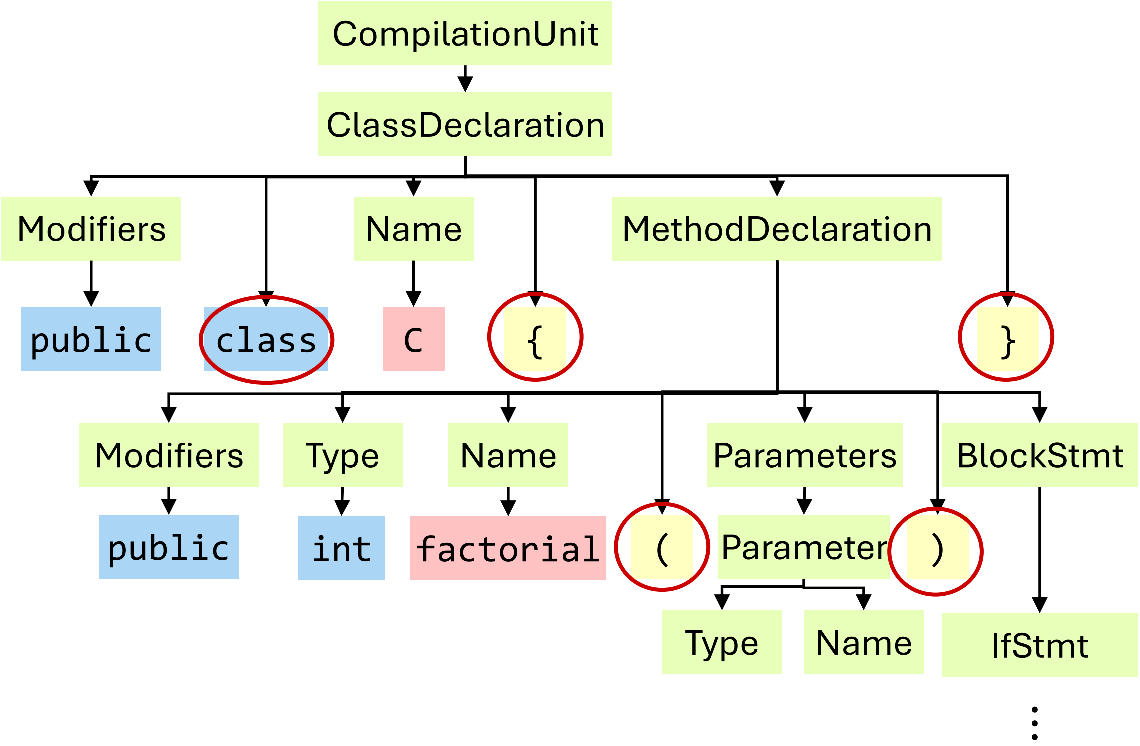
other compilation steps

compiled code

```
public class C {
  public int factorial(int n) {
    if (n > 0) {
      return n * factorial(n-1);
    } else {
      return 0;
    }
  }
}
```

public	KW
class	KW
C	ID
{	SYM
public	KW
int	KW
factorial	ID
(	SYM
int	KW
n	ID
)	SYM
{	SYM
if	KW
...	

PL tokens != ML tokens


CompilationUnit

ClassDeclaration

Modifiers

Name

MethodDeclaration

public

class

C

{

}

Modifiers

Type

Name

Parameters

BlockStmt

public

int

factorial

(

Parameter

)

Type

Name

IfStmt

⋮

concrete AST: all tokens explicitly appear as nodes

Context-Free Grammar

production rule

LHS can be expanded to RHS

context-free

✅ S: s1 s2

❌ A S B: A s1 s2 B

lexer rules

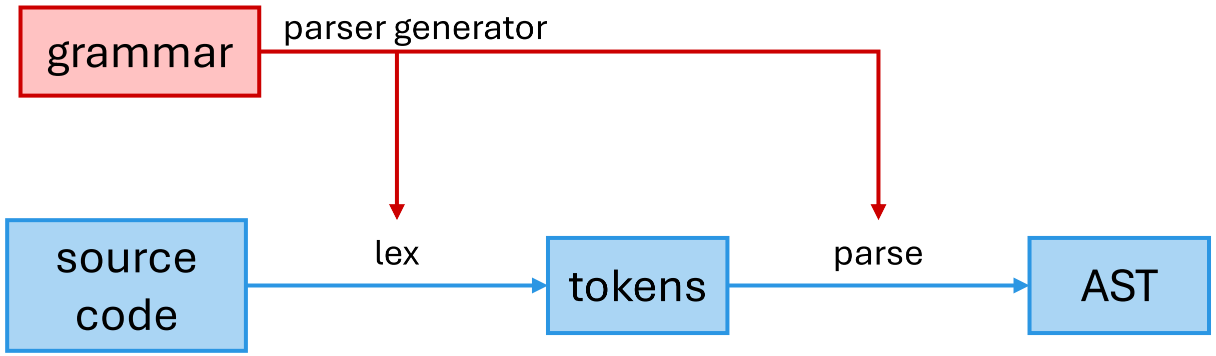
```
IDENTIFIER: Letter LetterOrDigit*;
Letter: [a-zA-Z$_];
LetterOrDigit: Letter | [0-9];
```

parser rules

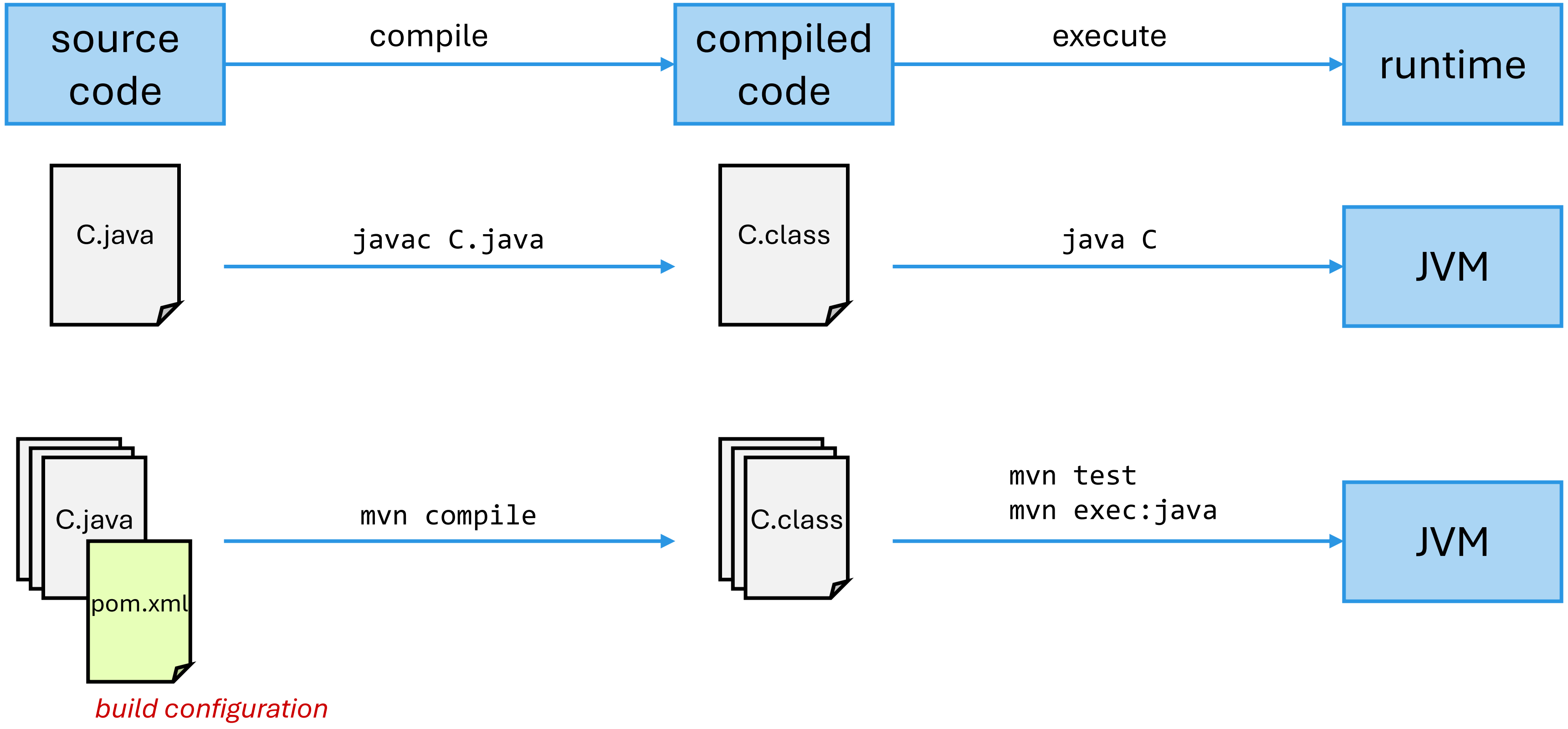
```
methodDeclaration
    : typeTypeOrVoid identifier formalParameters ('[' ']')* (THROWS qualifiedNameList)? methodBody
    ;

typeTypeOrVoid
    : typeType
    | VOID
    ;

typeType
    : annotation* (classOrInterfaceType | primitiveType) (annotation* '[' ']')*
    ;

primitiveType: BOOLEAN | CHAR | BYTE | SHORT | INT | LONG | FLOAT | DOUBLE;
```

Parser Generator

- Generate (lexers and) parsers given grammar rules
- Examples
  - ANTLR [tool](https://github.com/antlr/antlr4) [grammars](https://github.com/antlr/grammars-v4)
  - Tree-sitter [tool](https://github.com/tree-sitter/tree-sitter) [grammars](https://github.com/tree-sitter/tree-sitter/wiki/List-of-parsers)



Build System Overview



Build System Examples

- Java
  - build system (configuration file):
    Maven (pom.xml) https://maven.apache.org/
    Gradle (gradle.build) https://gradle.org/
  - online repository:
    Maven central https://mvnrepository.com/repos/central
- Python
  - build system (configuration file):
    Setuptools, hatch, poetry… (pyproject.toml, setup.py, requirements.txt…)
    https://packaging.python.org/en/latest/overview/
  - online repository:
    pypi https://pypi.org/