CS 153: Concepts of Compiler Design

September 19 Class Meeting

Department of Computer Science
San Jose State University

Fall 2017
Instructor: Ron Mak

www.cs.sjsu.edu/~mak

Assignment #3

- Invent a new Pascal **when** statement:

```
WHEN
    i = 1 => f := 10;
    i = 2 => f := 20;
    i = 3 => f := 30;
    i = 4 => f := 40;
    OTHERWISE => f := -1
END
```

- New syntax, but old semantics, equivalent to:

```
IF      i = 1 THEN f := 10
ELSE IF i = 2 THEN f := 20
ELSE IF i = 3 THEN t := 30
ELSE IF i = 4 THEN f := 40
ELSE               f := -1;
```

Assignment #3, *cont'd*

- New syntax:
  - Any boolean expression as the selector to the left of **=>**
  - A single statement (which can be compound) to the right of **=>**
- Old semantics:
  - Evaluate the boolean selectors sequentially from first to last.
  - If a selector is true, then execute the corresponding statement to the right of **=>** and then leave the statement.

Assignment #3, *cont'd*

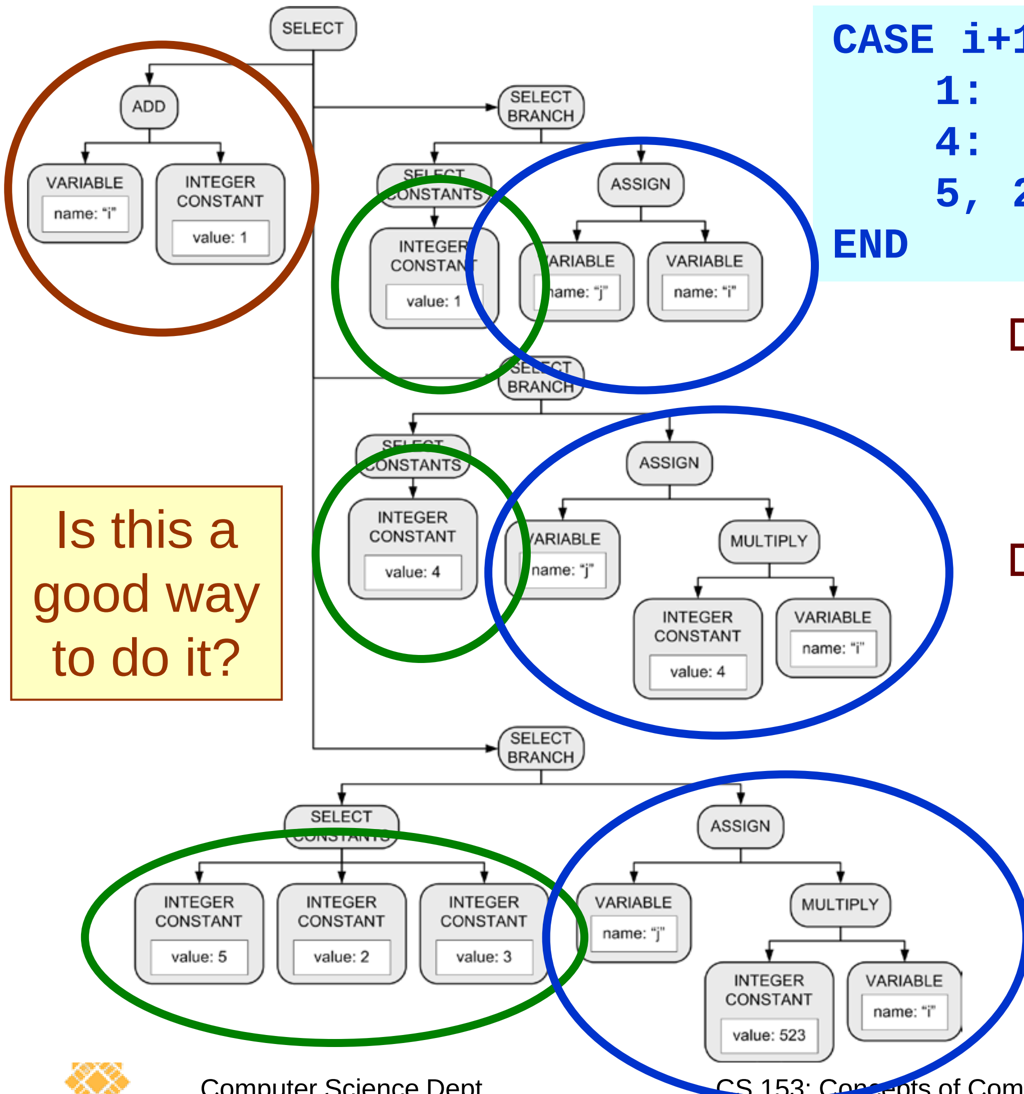
- Draw syntax diagrams for the **when** statement.
- Design the parse tree for the **when** statement.
  - Draw the tree for a simple **when** statement.
- Write the parser for the **when** statement.
  - Test it on some sample **when** statements.
  - Is it building proper trees?

Assignment #3, *cont'd*

- Write the backend executor for the **when** statement.
  - Test it by executing some sample statements.
  - Do you get the expected results?

- Due Monday, October 2.

Executing a SELECT Parse Tree



Is this a good way to do it?

```
CASE i+1 OF
    1:        j := i;
    4:        j := 4*i;
    5, 2, 3: j := 523*i;
END
```

- Evaluate the first child expression subtree to get the selection value.
- Examine each SELECT BRANCH subtree.
  - Look for the selection value in the SELECT CONSTANTS list of each SELECT BRANCH.
  - If there is a match, then execute the SELECT BRANCH's statement subtree.

Executing a SELECT Parse Tree, *cont'd*

- Why is searching for a matching selection value among all the SELECT BRANCHes bad?
  - It's inefficient.
- Selection values that appear earlier in the parse tree are found faster.
  - The Pascal programmer should not have to consider the order of the selection values.

Executing a SELECT Parse Tree, *cont'd*

- A better solution:
  For each SELECT tree, create a jump table implemented as a hash table.

- Build the table from the SELECT parse tree.

- Each jump table entry contains:
  - **Key:** A constant selection value.
  - **Value:** The root node of the corresponding statement subtree.

Executing a SELECT Parse Tree, *cont'd*

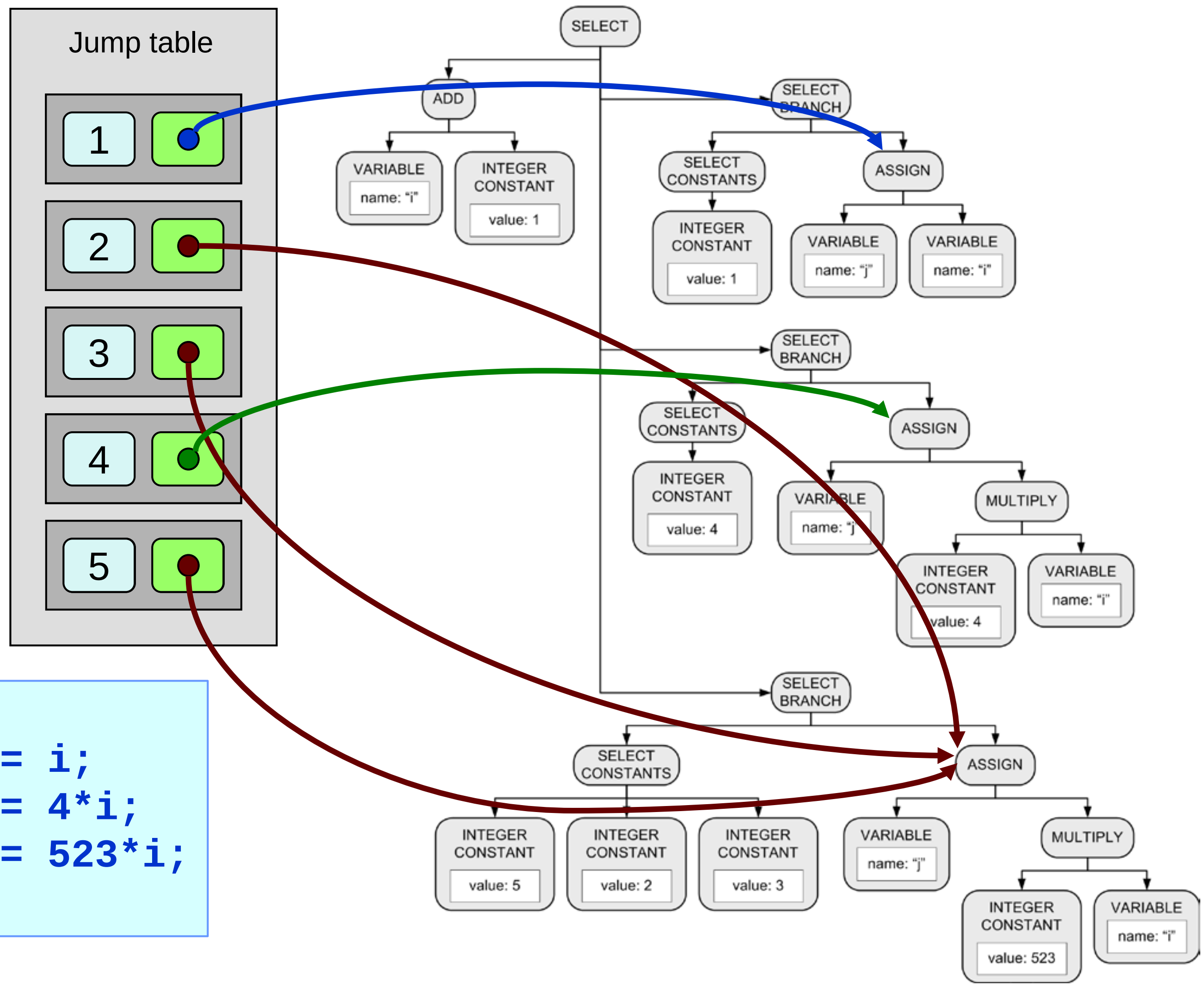
- During execution, the computed selection value is the key that extracts the corresponding statement subtree to execute.

SELECT Jump Table

Key: constant selection value

Value: root node of the corresponding statement

This is an example of **optimization** for **faster execution**.




```
CASE i+1 OF
    1:        j := i;
    4:        j := 4*i;
    5, 2, 3: j := 523*i;
END
```

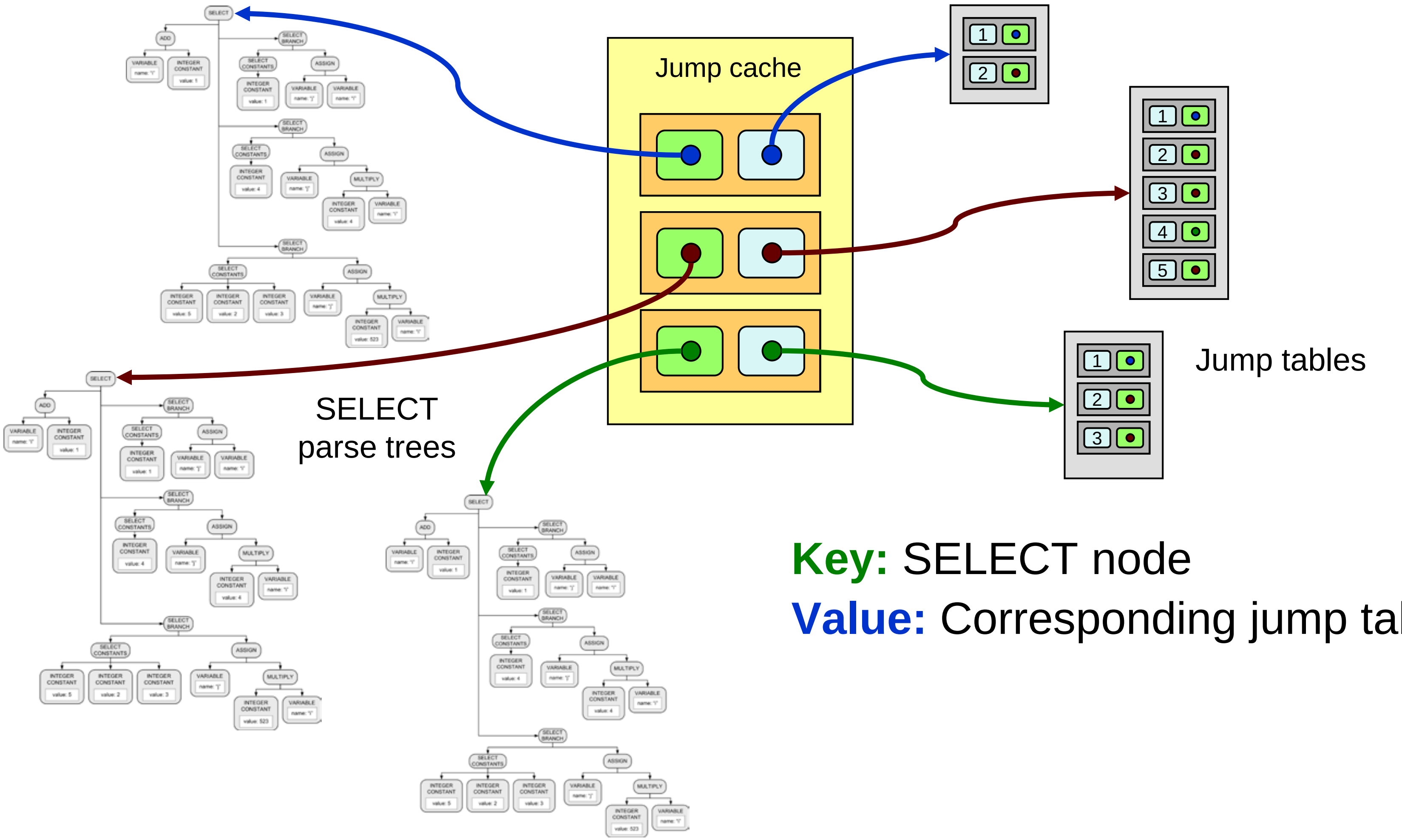
Multiple **CASE** Statements

- If a Pascal source program contains multiple **CASE** statements, there will be multiple SELECT parse trees.

- Create a global jump cache, a hash table of jump tables.

- Each jump cache entry contains:
  - **Key :** The SELECT node of a SELECT parse tree.
  - **Value:** The jump table created for that SELECT parse tree.

Jump Cache of Jump Tables



Key: SELECT node

Value: Corresponding jump table

Class SelectExecutor

- The global jump cache, which contains a jump table for each SELECT tree in the Pascal source program.

```
// Jump cache: entry key is a SELECT node,
//             entry value is the jump table.

// Jump table: entry key is a selection value,
//             entry value is the branch statement root node.

private static HashMap<ICodeNode, HashMap<Object, ICodeNode>> jumpCache =
    new HashMap<ICodeNode, HashMap<Object, ICodeNode>>();
```

Class SelectExecutor

```
public Object execute(ICodeNode node)
{
    HashMap<Object, ICodeNode> jumpTable = jumpCache.get(node);

    ArrayList<ICodeNode> selectChildren = node.getChildren();
    ICodeNode exprNode = selectChildren.get(0);

    ExpressionExecutor expressionExecutor = new ExpressionExecutor(this);
    Object selectValue = expressionExecutor.execute(exprNode);

    ICodeNode statementNode = jumpTable.get(selectValue);
    if (statementNode != null) {
        StatementExecutor statementExecutor = new StatementExecutor(this);
        statementExecutor.execute(statementNode);
    }

    ++executionCount;  // count the SELECT statement itself
    return null;
}
```

Get the right jump table from the jump cache.

Evaluate the selection value.

Get the right statement to execute.

Can we eliminate the jump cache?

Simple Interpreter II

- Demos
  - `java –classpath classes Pascal execute case.txt`

Multipass Compilers

- A compiler or an interpreter makes a "pass" each time it processes the source program.
  - Either the original source text, or
  - The intermediate form (parse tree)

Three-Pass Compiler

- We've designed a 3-pass compiler or interpreter.
- **Pass 1:** Parse the source in order to build the parse tree and symbol tables.
- **Pass 2:** Work on the parse tree to do some optimizations.
  - Example: Create **CASE** statement jump tables.
- **Pass 3:** Walk the parse tree to generate object code (compiler) or to execute the program (interpreter).

Multipass Compilers, *cont'd*

- Having multiple passes breaks up the work of a compiler or an interpreter into distinct steps.
- Front end, intermediate tier, back end:
  - Modularize the structure of a compiler or interpreter
- Multiple passes:
  - Modularize the work of compiling or interpreting a program.

Multipass Compilers, *cont'd*

- Back when computers had very limited amounts of memory, multiple passes were necessary.
- The compiler code for each pass did its work and then it was removed from memory.
- The code for the next pass was loaded into memory to do its work based on the work of the previous pass.

Multipass Compilers, *cont'd*

- Example: The FORTRAN compiler for the IBM 1401 could work in only 8K of memory and made up to 63 passes over a source program.
  - See:

http://www.cs.sjsu.edu/~mak/CS153/lectures/IBM1401FORTRANCompiler.pdf

Scripting Engine

- We now have a simple <u>scripting engine</u>!
  - Expressions
  - Assignment statements
  - Control statements
  - Compound statements
  - Variables that are untyped

What's Next?

- Parse Pascal declarations
- Type checking
- Parse procedure and function declarations
- Runtime memory management
- Interpret <u>entire</u> Pascal programs.

Parsing Declarations

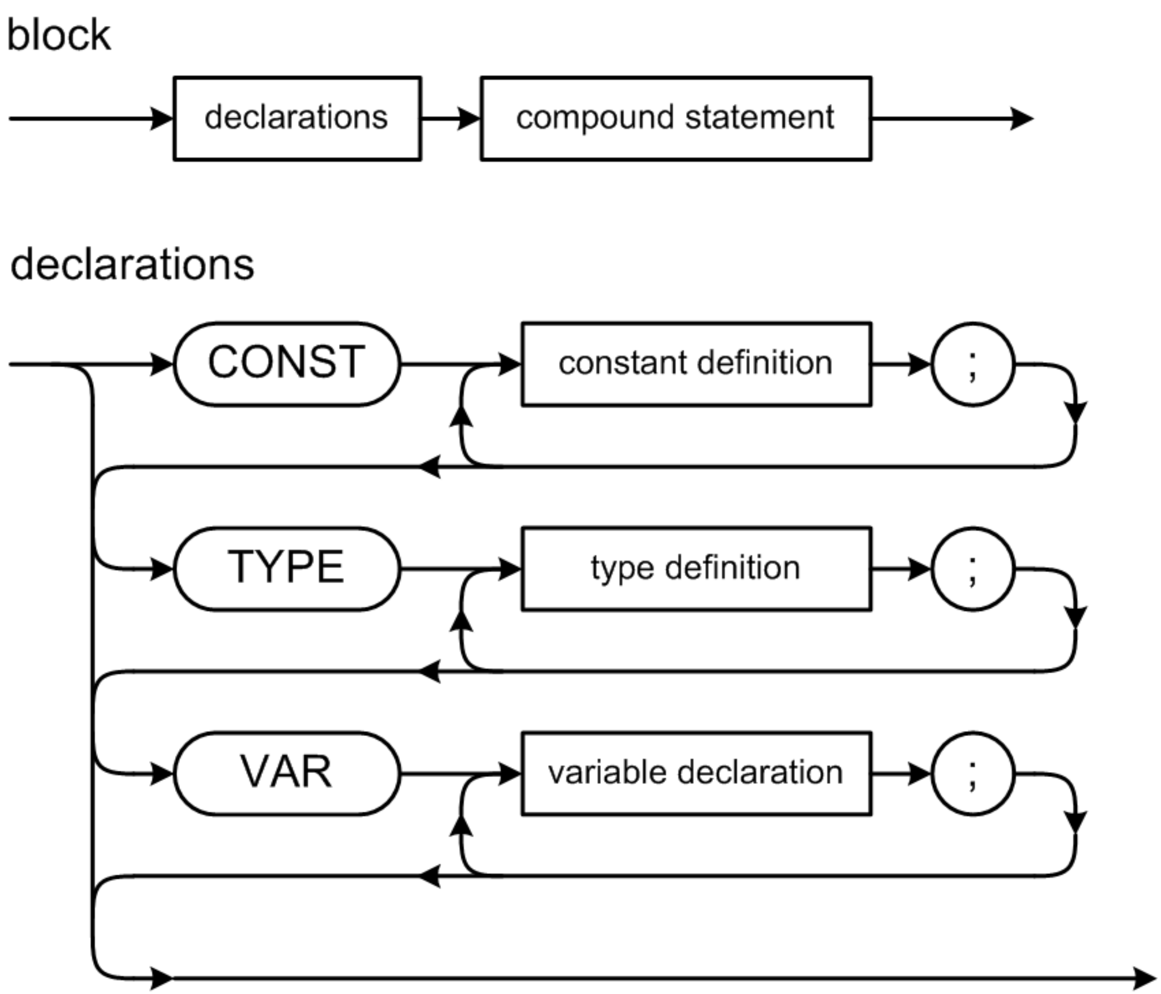
- The declarations of a programming language are often the most challenging to parse.

- Declarations syntax can be difficult.
- Declarations often include recursive definitions.
- You must keep of track of diverse information.
- Many new items to enter into the symbol table.

Pascal Declarations

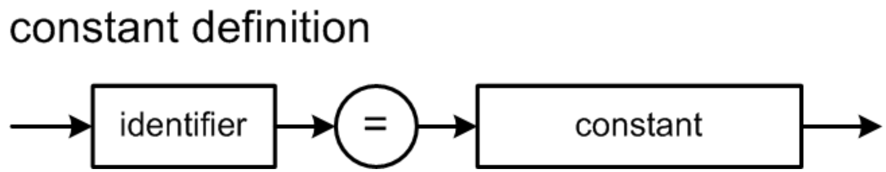
- Classic Pascal declarations consist of 5 parts, each optional, but <u>always in this order</u>:
  1. Label declarations
  2. Constant definitions
  3. Type definitions
  4. Variable declarations
  5. Procedure and function declarations
- We will examine 2, 3, and 4 next.
  - We'll do procedures and functions in a couple of weeks.

Pascal Declarations



- The **CONST**, **TYPE**, and **VAR** parts are optional, but they must come in this order.
- Note that constants and types are **defined**, but variables are **declared**.
- Collectively, you refer to all of them as declarations.

Pascal Constant Definitions

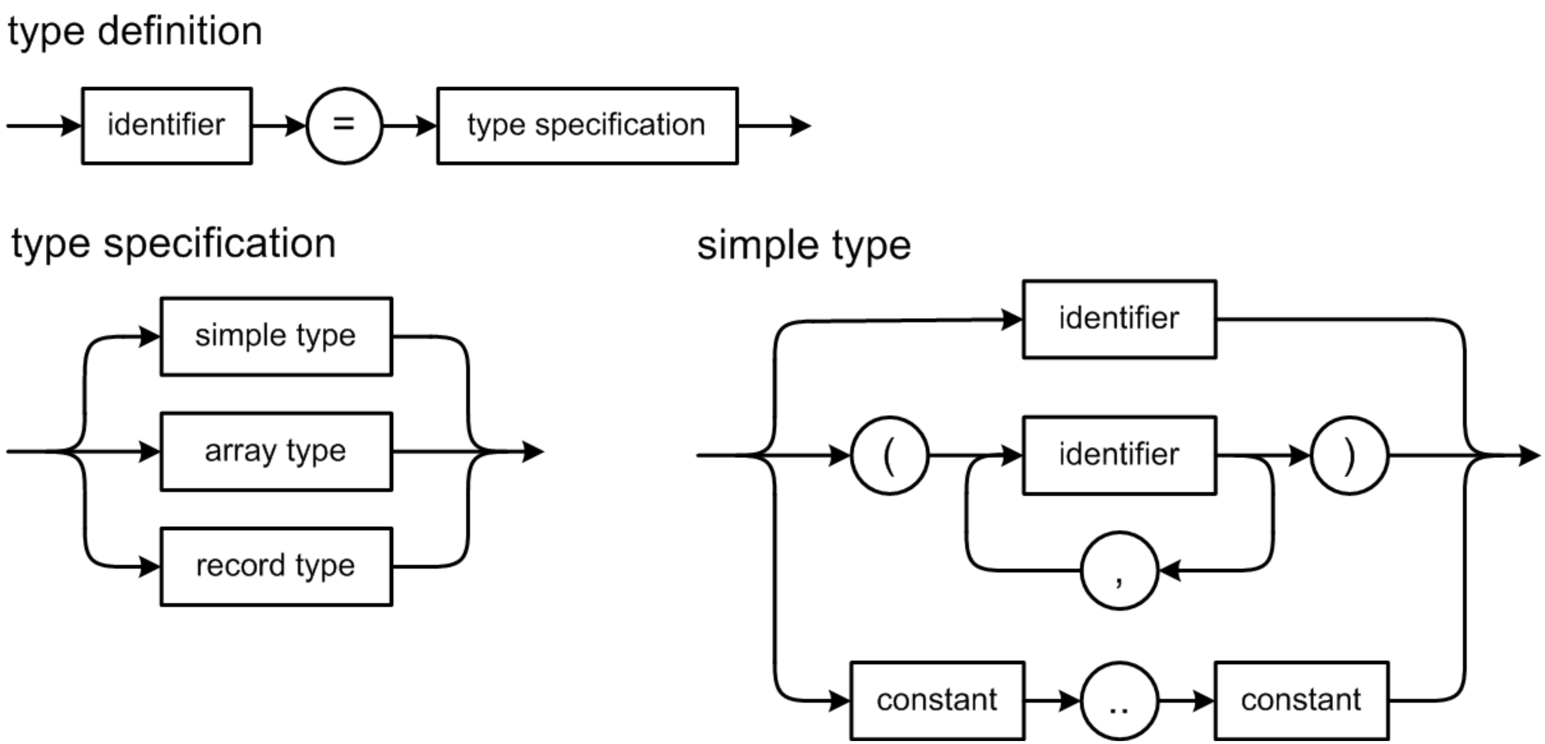


- Example constant definition part:

```
CONST
    factor = 8;
    epsilon = 1.0e-6;
    ch = 'x';
    limit = -epsilon;
    message = 'Press the OK button to confirm your selection.';
```

- Classic Pascal only allows a constant value after the **=** sign.
  - No constant expressions.

Pascal Type Definitions



- A Pascal simple type can be:
  - scalar (**`integer`**, **`real`**, **`boolean`**, **`char`**)
  - enumeration
  - subrange

Not reserved words!

Pascal Simple Type Definitions

- Examples of subrange and enumeration type definitions:

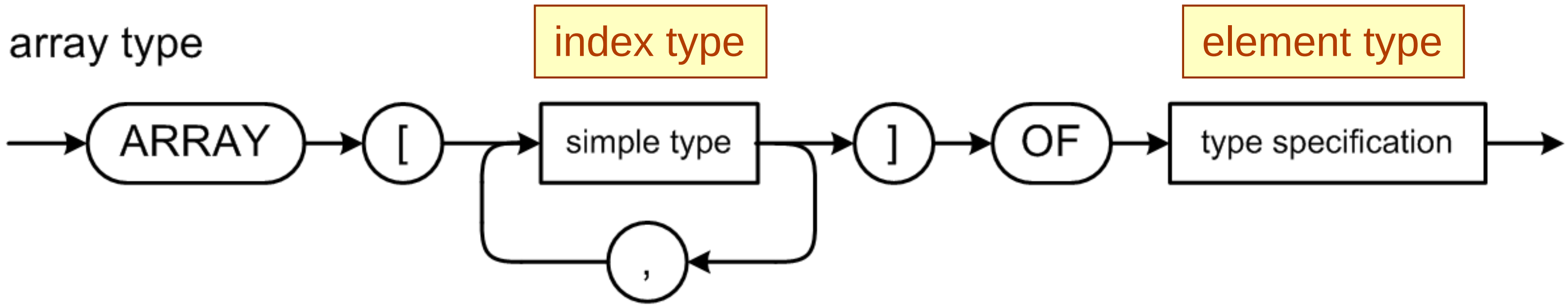
```
CONST
    factor = 8;

TYPE
    range1 = 0..factor; {subrange of integer (factor is constant)}
    range2 = 'a'..'q';  {subrange of char}
    range3 = range1;    {type identifier}

    grades  = (A, B, C, D, F);  {enumeration}
    passing = A..D;             {subrange of enumeration}

    week    = (monday, tuesday, wednesday, thursday,
               friday, saturday, sunday);
    weekday = monday..friday;
    weekend = saturday..sunday;
```

Pascal Array Type Definitions



- An array type specification has an index type and an element type.
- The index type must be a simple type (subrange or enumeration).
- The element type can be any type.
  - Including another array type (multidimensional arrays).

Pascal Array Type Definitions

- Examples of array definitions.

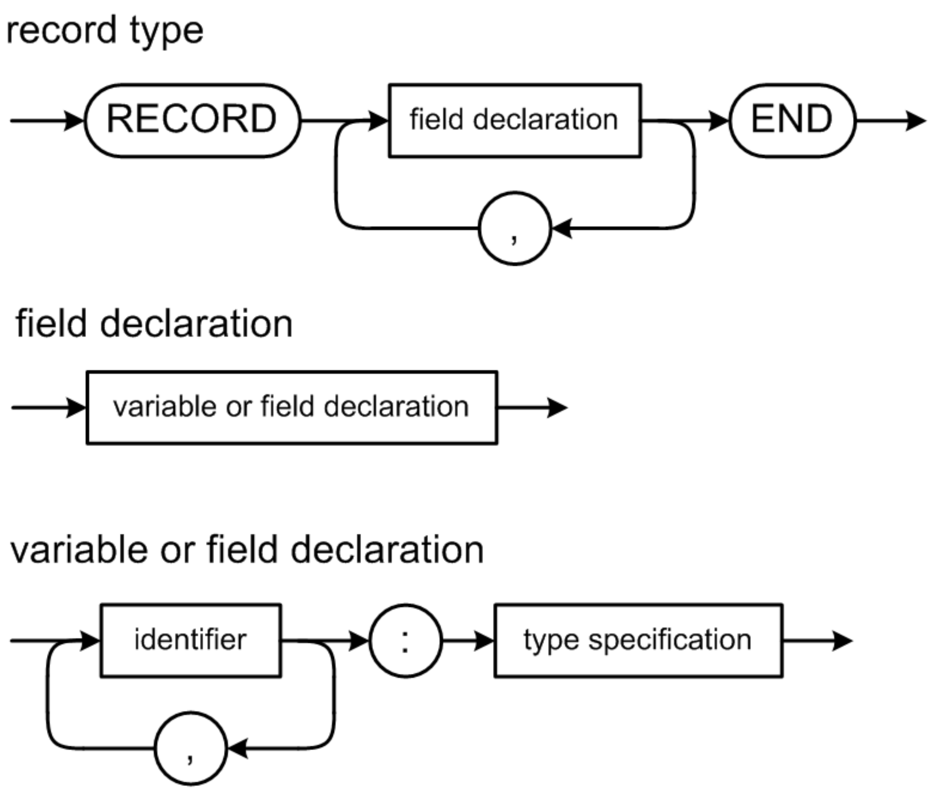
```
TYPE
    ar1 = ARRAY [grades] OF integer;
    ar2 = ARRAY [(alpha, beta, gamma)] OF range2;
    ar3 = ARRAY [weekday] OF ar2;
    ar4 = ARRAY [range3] OF (foo, bar, baz);
    ar5 = ARRAY [range1] OF ARRAY [range2] OF ARRAY[c..e] OF enum2;
    ar6 = ARRAY [range1, range2, c..e] OF enum2;
```

Type definitions **ar5** and **ar6** above are equivalent ways to define a multidimensional array.

- A Pascal string type is an array of characters.
  - The index type must be an integer subrange with a lower limit of 1.

```
str = ARRAY [1..10] OF char;
```

Pascal Record Type Definitions



- A record field can be any type.
  - Including another record type (nested records).

Pascal Record Type Definitions

- Examples of record definitions:

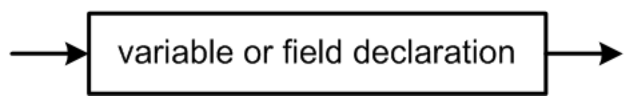
```
TYPE
    rec1 = RECORD
               i : integer;
               r : real;
               b1, b2 : boolean;
               c : char
           END;

    rec2 = RECORD
               ten : integer;
               r : rec1;
               a1, a2, a3 : ARRAY [range3] OF range2;
           END;
```

Pascal Variable Declarations

- Variable declarations are syntactically similar to record field declarations:

variable declaration



- Examples:

```
VAR
    var1 : integer;
    var2, var3 : range2;
    var4 : ar2
    var5 : rec1;
    direction : (north, south, east, west);
```

- Types can be named or unnamed.