**What is a *programming language*?**

A programming language is the collective of syntactic rules, keywords, naming structures, data structures, expression and control structures that is intended for the expression of computer programs and that is capable of expressing any computer program. Programming Language is a formal method that:

* Describe a solution to a problem
* Organize a solution to a problem
* Reason about a solution to a problem
* Interface between user and machine

Programming languages trade-off:

* Ease of use - high level
* Efficiency - low level

**Attributes of a Good Programming Language**

**1. Clarity, Simplicity and Unity**

A programming Language provides a medium to conceptual thinking of new algorithms and also a medium to Execute your thought Process into real Coding Statements. For algorithms to be implemented on a Language it’s a basic need is that the language is quite *clear, simple and Unified* in structure. Such that the Primitives of language can be utilized to develop algorithms. It is desirable to have a minimum number of different concepts, so that combining multiple concepts won’t be that complex in nature. It should be simple and regular as possible. This attribute of a language is known as conceptual Integrity.

The main concern of a language now a day is its readability. The syntax of language affects the ease with which programs are written, tested and later used for knowledge or research purpose. A complex syntax language may be easy to write program in, but it proves to be difficult to read and debug the code for later sessions. **For example** APL programs are so complicated that even the own developers find it difficult to understand after 1-2 months. The language should be simple enough to understand or point out errors.

**2. Orthogonality**

The term orthogonality refers to the attribute of being able to combine various features of a language in all possible combinations, with every combination being meaningful. Language design must follow orthogonality principle i.e. independent functions should be controlled by independent mechanisms. **For example,** suppose a language provides with an expression let’s say an arithmetical calculation operator .Taking Another Expression facilitated by the language like conditional Statement, which has 2 outputs either 0 or 1 (in some cases TRUE or FALSE). Now the language should support combination of these two expressions. So that new statement can be formed, and this orthogonality helps to develop many new algorithms.

**3. Naturalness for the application**

A language needs a Syntax that, when applied properly, allows the program Structure to reflect the logical structure what a programmer wanted it to. *Arithmetic Algorithms ,concurrent algorithms , logic Algorithms* and other type of statement have differing natural structures, that can be represented by the Program in that language. The language should provide appropriate *data structures, operations, control structures and a natural syntax* for the problem to be solved.

**For Example:** Consider a real life condition of plates being placed above each plate, this structure is known as Stack. This Stack can be implemented into programming world also. This is used as a data Structure in most of the Languages.

**4. Support for Abstraction**

Many times languages fail to implement many real life problems into Programs. There is always a gap between abstract data structures and operations. Even most natural Programming language fails to bridge the gap. **For Example:** Consider a situation where a scheduling is to be done for college *student for attending a lecture in a class section, teacher.* Suppose the requirement is to assign a student a section lecture and teacher to attend, which are common task for natural application, but are not provided by C.

The need of point is to design an appropriate abstraction for the problems solution and then implementing these abstraction using most primitive features of a language. Ideally, the language should provide the data structures, data types and operations to maintain such abstractions .C++ is one of the most used language, that provide such facilities.

**5. Ease of Program Verification**

The reliability of a programming Language written in a language is always a central Concern. There are many techniques which can be used to keep track of correct functionality of a language. Sometimes testing the Program with random values of the inputs and obtaining corresponding outputs. Program verification should be provided by languages to check and minimize the errors.

**6. Programming Environment**

The environment also plays a vital role in success of a Language. The environment which is technically weak may get a bad response of Programmer, rather than a language that has less facility than the former but its environment is Technically Good. Some of the Good featured of an environment are Special editors and testing packages tailored to the language may greatly speed up the creation and testing of Programs.

**7. Portability of Programs**

The important criterion for many programming projects is the Transportability of the resulting program from one computer to other systems. A language which is widely available and does not support different features on different computer System, which may have different hardware, is considered a good language. **For Example** *C, C++* and most of the language now days are Portable in nature.

**8. Cost of Use**

The trickiest point that always matter a lot in any system that uses resources. It’s a major element to decide the Evaluation of any programming language, but cost means many different things.

**(a) Cost of Program Execution**

Program Execution cost is total amount which has been used to implement the program. The research work on design, optimizing compilers, data allocation registers etc. This is the basic things which come under the cost of Program Execution.

**(b) Cost of Program Translation**

The next concern is program compilation. The program is compiled many times than it is being executed. In such case, it is important to have a speed and efficient compiler to handle this Job.

**(c) Cost of Program Creation, Testing and Use**

Another aspect of Cost management. This includes the cost which a programmer charges for his work of creating Project with the Specified features, the cost involving the Testing issues.

**(d) Cost of Program Maintenance**

After a program is being installed in a System, then after certain intervals it needs maintenance to run smoothly. The maintenance includes the rectification of Error propagated in real time, the updating of Program as need of time.

**Program DESIGN Notations**

Complexity in programming and understanding programs led to development of *program design notations.* These were designed to help the programmer, not to be interpreted by computers.

Some of these notations helped the programmer to design the:

Memory layout

Control flow

•  *Flow Diagrams*

•  Later: Flowcharts

Mnemonics

• Helps to remember instruction codes like assembly language today

**SHORTCOMING OF EARLIER COMPUTER HARDWARE**

**a) Floating Point Arithmetic**

Floating point is harder to support by hardware due to its complex representation. The  earliest built-in floating point support was provided by IBM 704 in 1953. Before that, it had to be *simulated using following steps:*

* Manual scaling, a technique in which numbers were multiplied by scale factor in order to keep them within the range of the integer arithmetic facilities of the computer
* The difficulty of manual scaling led to the development of floating-point subroutines, that is, of subroutines for performing basic floating-point operations(addition, subtraction, multiplication, division, square root, etc)
* Multiply by constant factor
* Use integer processor
* Manually scale back result
* Complicated and error-prone process

**b) Indexing**

Array is one of most common data structures used in the computer programming, useful for various computations. Array uses a technique called indexing to store and retrieve data from consecutive memory locations. Indexing is like “Adding a variable index quantity to a fixed address in order to access the element of an array”. Indexing was not supported by early computers. They used address modification by:

* Altering the program’s own data accessing instruction
* It also consumed much of the scarce memory with this address modification code
* For this reason it was common to use subroutines to perform indexing
* Most of the actual execution time was spent inside the floating-point and indexing subroutines.

However it was very error prone process.

This justified the use of pseudo-code interpreters, which we will discuss next.

**PseudoCode**

It is an instruction code that is different than that provided by the real machine. Pseudo-code offered floating point support and indexing which was actually not supported by earlier generation computers. It provided entirely new instruction set not offered by the real hardware.

**Need of PseudoCode**

During first generation of computer, programming was very difficult as the programmer need to know about the hardware specification of every machine that the program was intended for. For example in 1950’s for IBM 650 which has following characteristics:

* No programming language was available (not even assembler)
* Memory was only a few thousand words.
* Stored program and data on rotating drum.
* Instructions included address of next instruction so that rotating drum was under next instruction to execute and no full rotations were wasted.

**PseudoCode Interpreters**

Pseudo-Code Interpreter is an interpretive subroutine (The **subroutine** is an important part of any **computer** system's **architecture**. A **subroutine** is a group of instructions that usually performs one task, it is a reusable section of the software that is stored in memory once, but used as often as necessary.) developed to run the pseudocode. They are used for saving memory since the pseudocode is more compact than machines real program code. It implements a virtual computer which allows us to use functionality with its own data types (e.g. floating point) and operations (e.g. indexing) not provided by the actual hardware in which the virtual machine resides but abstracted by the virtual machine. It has own data types and operations and we can view all programming languages this way. The pseudocode was at the higher level and provided facilities more suitable to applications and it eliminated many details from programing. In general it is an example of “Automation Principle” of programming language.

Pseudo-Code interpreters were commonly used to perform floating-point operations and indexing. Consistent use of these simplified the programming process and this simulated instructions not provided by the hardware.

Pseudo-Code interpreter (a primitive, interpreted programming language) implements:

* A virtual computer
* New instruction set
* New data structures

Virtual computer:

* Higher level than actual hardware
* Provides facilities more suitable to applications
* Abstracts away hardware details

PseudoCode follows two basic *Principles of Programming*

***The Automation Principle***

Automate mechanical, tedious, or error prone activities.

***The Regularity Principle***

Regular rules, without exceptions, are easier to learn, use, describe, and implement.

**Design of a Pseudo-Code**

The design of PseudoCode is based on the capabilities and constraints of the first generation computers. In 1950 Capabilities expected by the programmers and not support by the hardware at that time are:

* Floating point operation support (+,-,\*,/,…)
* Comparisons (=,≠,<,≤,>,≥)
* Indexing
* Transfer of control
* Input/output

**Hardware Assumptions**

The IBM 650 will serve as the hardware

* 1 word: 10 decimal digits + 1 sign
* 2000 byte memory

•  1000 for data

•  1000 for program

***The design of pseudocode must follow impossible error principle because “making errors impossible to commit is preferable to detecting them after their commission”.***E.g.: Cannot modify the program accidentally, since memory modifying operations are for “data memory” only.

**Language Design**

1 word can be enough to specify a 3- operand instruction. Operation contains a sign (+, –, etc.) and 1 digit. It supports 20 operations. Three 3-digit operands each accesses memory locations in data area.

***Language design should be orthogonal which means:***

* Operations should be more intuitive than machine code
* Use the *sign* to get more orthogonality

***Language design must follow orthogonality principle i.e. independent functions should be controlled by independent mechanisms.***

***Instruction format:***

**Operand src1 src2 dst**

E.g.: x + y 🡪 z : +1 010 150 200

This instruction instructs to “Add values at location 010 and 150, and save it to location 200”

To follow Orthogonal design the subtract operation should be ‘-1’

**Arithmetic Operations**

**Comparisons**

Comparisons alter control flow of the program

If *x* < *y* then go to *z*

For example, the instruction

+4 200 201 035

Means: if the contents of (data) location 200 equal the contents of (data) location 201, then go to the instruction in (program) location 035.

**Moving**

Data could be moved by “adding 0” to an address, but that operation is expensive so the new operation should be assigned to move data.

+ 0 150 000 200 (First operand is source, third operand is destination while second operand is not used)

**Indexing**

One of the justifications for our pseudo-code was that it provided built-in indexing, so we will turn to the design of this facility next. To perform indexing we will need the address of the array and the index variable, thus consuming two of the three address fields in the instruction. Therefore, the only operations we can perform directly on array elements are to move them to or from other locations. We can use the codes +6 and -6 to move from or to an array: xi  z and x yi.

The formats of these operations are

*+6 xxx iii zzz*

*-6 xxx yyy iii*

E.g. if there is a 100-element array beginning at location 250 in data memory, and location 050 contains 17, then

+6 250 050 803

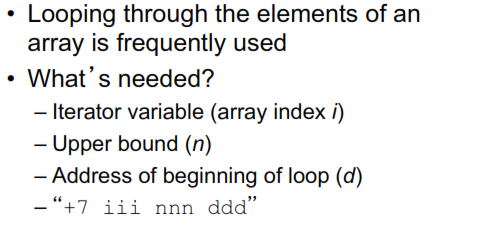
will move the contents of location 267(=250+17) to location 803.

Similarly,

-6 722 250 050

will move the contents of location 722 to location 267.

**Looping**



The operation increments location iii and loops to instruction ddd if the result is less than the contents of nnn.

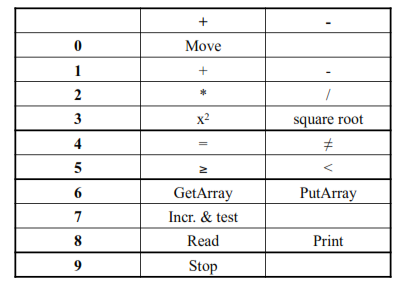
**Input/output**

Program needs to read data from input and write data to output.

* Needs only a memory location to read from or write to
* Read: “+8 000 000 dst”
* Print: “-8 000 000 src”

***Language design must follow the abstraction principle i.e. avoid requiring something to be stated more than once; factor out the recurring pattern.***

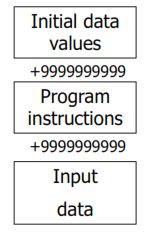
**Complete Pseudo-Code operations**



**Program Structure**

We now know how to write individual instructions, but we have not designed a means of constructing the program as a whole. The simplest solution to this problem is to have interpreter read initialization cards and their content in consecutive memory locations.

Thus, the structure of a program is:



We have used a card containing the “flag value” +9999999999 to separate the initial value from the program and the program from the input data. The loader reads in the initial data values and stores them in consecutive locations (starting with 000) in the data memory. The loader reads in the program instructions and stores them in consecutive locations (starting with 000) in the program memory. The loader does not read the input data; this is read by the user’s program whenever it executes a +8 instruction.

Therefore, the general structure of a program is (1) declarations (2) executable statements, and (3) input data.

**Implementing the Interpreter**

The interpreter for our pseudo-coded program can be implemented as:

* Model interpreter behavior after manual execution
* Cheat: Implement using a high-level language
* We have to simulate the hardware in software

The second option is impossible as high level language is not developed during 1950s. We will focus on the third implementation method.

**Data Structures**

Data structures are needed to simulate the IBM 650 which contains:

* Data memory
* Program memory
* Instruction pointer

**Structure of the Interpreter**

**Data**

**Memory**

**Program**

**Memory**

000 000

999 999

IP

[ Fig. Interpreter Data Structure]

**The Read-Execute Cycle is the Heart of an Iterative Interpreter**

We can now consider how a program is actually interpreted. Roughly, what we will do is read the next instruction to be executed (as indicated by the instruction pointer), determine the operation encoded by the instruction, and then perform that operation. When execution of the operation is completed. This process is called the *read-execute cycle*, and can be summarized as follows:

1. Read the next instruction
2. Decode the instruction
3. Execute the operation
4. Continue from step 1

The instruction pointer (IP) must be updated each time an instruction is read and the best option to that is in step 1 because if we update at last the branch instruction may overwrite the IP.

Typical code for step 1 is

*instruction : = Program[IP];*

*IP : = IP+1;*

**Decoding Instructions by extracting their parts**

PseudoCode has been designed with a regular structure, decoding is simple; we simply extract the sign, operation code, and the three address fields. For example, the destination address could be extracted by

*dest := abs (instruction) mod 1000*

(Where ‘x mod y’ gives the remainder of dividing x by y). We assume that the names of these extracted parts are **sign, op, opnd1, opnd2, and dest**.

The operations types are:

•  Select operation

Switch-statement (case-statement)

•  Arithmetic operations

Straight-forward

•  Control-flow

IP may also need to be altered

**Labeling**

Labeling is the aid to the coding of the program. Labels make programs readable however they are not converted into object codes. Labels are used for replacing absolute addressing in the programs particularly in the loops and branching.

We can define a label in pseudocode using label definition operator as:

  -7 0LL 000 000

defines the statement number, or label, LL.

Only 100 numeric labels are possible (00-99)

The control flow instructions then jump to labels instead of absolute address.

**Interpreting Labels**

Label Interpreting Approaches:

**Look through all instructions from beginning of program?**

Yes, but that is slow. This is how some interpreters work. (BASIC, for instance)

**Create label table with absolute addresses for labels and bind addresses**

|  |  |
| --- | --- |
| **Label** | **Location** |
| 20 | 001 |
| 40 | 005 |
| 50 | 009 |

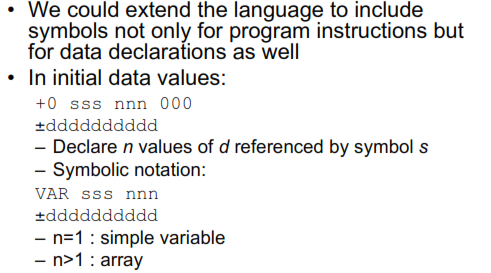
Much faster. Compilers do it this way.

***The labeling principle must be followed i.e. do not require users to know absolute numbers or addresses. Instead associate labels with number or addresses.***

**Data Labels**

We could use labels for variables as well by constructing a symbol table. This idea is easily extended to instructions as well to form a symbolic pseudo-code.

**Data Declaration**



**Debugging?**

Debugging always has to be done and it can be facilitated by debugging by printing instructions executed in order. Also, Interpreter can include *trace*, to get a trace of the execution of the program, that is, a record of the instructions it has executed.

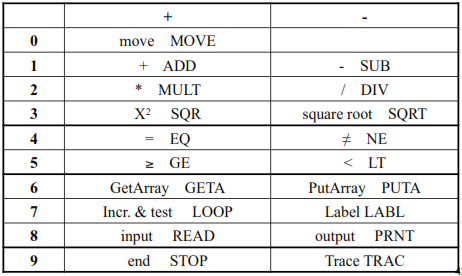
*Read Next Instruction:*

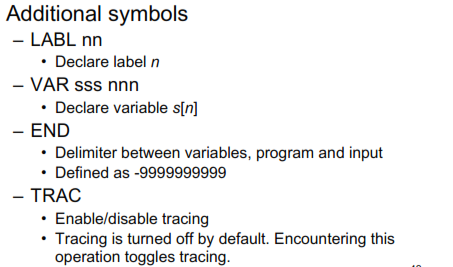
*If trace is enabled then*

*print IP, instruction;*

*IP := Ip+1.*

**Complete Symbolic Language**





**Sample Program**

VAR ZRO 1 Constant Zero

+0000000000

VAR I 1 Index

+0000000000

VAR SUM 1 Sum of array

+0000000000

...

END

READ N Read number of elements

LABL 20

READ TMP Read into TMP

GE TMP ZRO 40 If +ve, skip to 40

SUB ZRO TMP TMP Negate TMP

LABL 40

PUTA TMP DTA I Move TEMP into the ith element

LOOP I N 20 Loop for all array elements

...

STOP

END

***PseudoCode must follow security principle i.e. no program that violates the definition of the language, or its own intended structure, should escape detection.***

**Phenomenology of programming languages**

Programming language is used for solving the problems so it can be taken as a tool. Some of the phenomenon of programming language as tool from the investigations of Don Ihde is:

* ***Tools are Ampliative and Reductive***

To better understand the phenomenology of programming languages, we may begin with a simpler tool. Ihde contrasts the experience of using your hands to pick fruit with that of using a stick to knock the fruit down. On the one hand, the stick is ampliative: it extends your reach to otherwise inaccessible fruit. On the other hand, it is reductive: your experience of the fruit is mediated by the stick, for you do not have the direct experience of grasping the fruit and tugging it off the branch. You cannot feel if the fruit is ripe before you pick it.

“Technological Utopians” tend to focus on the ampliative aspect- the increased reach and power- and to ignore the reductive aspect, whereas “technological dystopian” tend to focus on the reductive aspect- the loss of direct, sensual experience - and to diminish the practical advantages of the tool.

* ***Fascination and fear are common to new tools***

When first introduced, programming languages elicited the two typical responses to a new technology: fascination and fear. Utopians tend to become fascinated with the ampliative aspects of new tools, so they embrace the new technology and are eager to use it and to promote it (even where its use is inappropriate); they are also inclined to extrapolation: extending the technology toward further amplification. Dystopian, in contrast, fear the reductive aspects of the tool (so higher level language are fear for their efficiency), or sometimes the ampliative aspects, which may seem dangerous. Ideally, greater familiarity with a technology allows us to grow beyond these reactions.

* ***With mastery, objectification become Embodiment***

A tool replaces immediate (direct) experience with mediated (indirect) experience. Yet, when a good tool is mastered, its mediation becomes transparent. Consider again the stick. If it is a good tool (sufficiently stiff, not too heavy, etc.) and if you know how to use it, then it functions as an extension of your arm, allowing you both to feel the fruit and to act on it. In this way the tool becomes partially embodied. On the other hand, if the stick is unsuitable or you are unskilled in its use, then you experience it as an object separate from your body; you relate to it rather than through it. With mastery a good tool becomes transparent; it is not invisible, for we still experience its ampliative and reductive aspects, but we are able to look through it rather than at it.

As you acquire skill with the language, it becomes transparent so that you can program the machine through the language and concentrate on the project rather than the tool. With mastery, objectification yields (partial) embodiment.

* ***Programming Language influence focus and action***

Tools influence the style of a project. E.g. writing technologies: dip pen, an electric typewriter, and a word processor. In case of dip pen it is slower than the speed of thought, with typewriter the speed is closer to the speed of thought, and with word processor, text can be revised and rearranged in small units, so there is greater tendency to salvage bits of text.

In general, a tool influences focus and action. It influences focus by making some aspects of the situation salient and by hiding others. Like others, programming language influence the focus and actions of programmers and therefore their programming style.