NATURAL LANGUAGE PROGRAMMING

WEIHAN HUANG

Master of Computer Science Department, State University of New York at Buffalo, U.S.A.
E-mail: weihanh@yahoo.com.tw

Abstract - In contrast with the mathematical syntax of functions, a natural language syntax is proposed for the programming statements. For Object Oriented Programming, I create new-instance statements and data of objects in natural language programming. New structures of Throw-Variable and For-Each loop statement are also provided. Lastly, I compare this work with other four natural language programmings.

Keywords - Language, Natural, Programming, Syntax

I. NATURAL LANGUAGESYNTAX FOR PROGRAMMING LANGUAGES

In Java[1], C[2], or C++[3], Pascal[4], prolog[5], they adopt the mathematical syntax of functions, i.e. f(arg1, arg2). For example, bool father(x,y).

I think we can have a more flexible syntax such that the expression will be closer to the natural language syntax.

e.g. $x is father of $y

The flexibility comes from the allowance for the arguments to appear in any place of the whole function expression. And this syntax is called the natural language programming syntax.

Other examples are

1. add $object to $list
2. stop for $number seconds
3. search $object in $array

II. FORMAL CONTEXT FREE GRAMMAR OF NATURAL LANGUAGE PROGRAMMING

In contrast with mathematical function syntax

f(x,y)

we abbreviate the natural language programming(NLP) syntax as

A $x B $y C

The formal context free grammar of natural language programming statements is as follows :

sentence := term
sentence := term sentence
term := variable
term := word
term := function Call
variable := $word
function Call := (sentence)

III. ADVANTAGES OF NATURAL LANGUAGE PROGRAMMING (NLP)

1. This syntax change can solve a classical syntax problem of constructors in Object Oriented Programming languages, which will be discussed later in the object oriented programming section IV.
2. Codes in NLP are easy to understand and easy to maintain. For example of "$x is father of $y". On the contrary, for the example of father(x,y), it is hard to tell if "x is father of y" or "y is father of x". If person A is going to maintain person B's code, it is not easy for A to understand. And even for B, after many years, B could forget which the correct interpretation is.
3. Natural language syntax changes slowly with time, while programming language syntax change quickly with time. The code written in natural language syntax can survive for a long run. It can be read and understood after 50-100 years.
4. Pseudo-codes now can be compiled and run. We used to write pseudo-codes in books to illustrate algorithms, and now these pseudo-codes are runnable.
5. Few comments need to be written. In java or C++, we often write natural languagedocumentations beside the codes. But in natural language programming(NLP), it merges the comments to the programming language syntax.
6. Natural language programming(NLP) is more "high-level". It is close to human habits in life. Thus programmer can write programs in a natural way. Reading programs is like reading articles. e.g.fill(array,10) // C, Pascal array.fill(10) // Java, C++➔fill $array with 10 // NLP
7. In addition to English, the flexible syntax can allow for many other natural languages to be realized in programming. For Chinese, it only need to have a small modification for the variable syntax:
   English $x ➔ Chinese x This will be discussed in section VII.
8. It is good for education. For little children, beginners of programming, and people who are

not in the science area, math notation could be formidable. Natural languages are easier for them to use and understand. And for Chinese people, elementary school students do not know English, and hence unable to learn programming. If natural language programming can be realized in Chinese, then it will be very easy for Chinese elementary school students to learn programming.

9. The flexible syntax gives rich expression. For example, the complex number in math can be written as 3+4i in NLP.
10. Spell check or grammar check could possibly be used by natural language programming compilers to help debugging.
11. The definitions of functions in NLP provide functional and procedural semantics for natural language sentences.

IV. OBJECT ORIENTED PROGRAMMING (OOP)

A. Constructor syntax for OOP

new [className] $variable words-variables.

For example, the constructor for a new window :
newwindow $wwith size(1024,768)
And this can solve a classical problem in constructors in previously created programming languages like Java or C++.
Constructor window(intposX, intposY)
Constructor window(intsizeX, intsizeY)

The problem is that the two constructors cannot be distinguished by the compilers. But with natural language programming, it can be solved by defining two constructors respectively
new window $window with position $array
new window $window with size $array

B. Data syntax for OOP in NLP

The access of data member of an object can be done by calling
dataName of $object
for example : name of $student.
let (name of $student) be “Weihan”;
write(name of $student);

V. THROW VARIABLES

The example of multiple outputs of a function is to search an element in a list, and returns

1. Found or not
2. The element
3. The index of the element

In Java, if only two things need to be returned, it can be implemented by returning the “null” when not found, and returning the element when found. But if the function need to return 3 things as above, then it is not easy for Java to do it. (Another way could return these things in an array.)

In C, or C++, it can use “pointers” or “ references” to make an argument as an output.

e.g. declare : bool search(Object object, Object * element, int * index)

calllist.search(object,&element, &index)

In natural language programming, it can be as follows :

Line 1 : search $object in $list;
Line 2 : if(#found), then {
Line 3 :write #element;
Line 4 :write #index;
Line 5 :}

This is closer to natural language writing. The function call in line 1 is much shorter and more understandable than the function call in C++ because the NLP function does not include the outputs. And also Line1 does not use the concepts of “pointers” and “references” which are more complicated concepts.

The variables starting with # in Line 2,3,4 are outputs thrown out by the function in Line 1. They are scoped variables generated in the block after the function“search $object in $list” is called.

The advantages are in the following :

1. Multiple outputs can be done in a natural way, and in a high-level way.
2. No need of using complicated concepts “pointers” and “references” as in C++.
3. No need to use an array or “null” for multiple outputs as in Java.

VI. NATURALIZE THE SYNTAX OF LOOP STRUCTURE

A. For looping list, array or any indexed objects, we can have the loop like for each student $x in $list, do {…..} An indexed object is defined such that its class contains 3 functions :length of $indexedObject :return $number
$indexedObject $indexNumber : return $object set $indexedObject $indexNmber to $object

Then the compiler will use these 3 functions to implement the for-each statement.
B. For looping a number variable from start position to end positon, we have loop $x from 1 to 100, do {……}

VII. VARIABLE SYNTAX IN CHINESE

For Chinese, there are no blank between words, so we

need two $ to surround a variable x. for example :
在$串列$中找尋$物件$
Without two $, only one $ will make mistakes by showing
在$串列中找尋$物件
This is because Chinese characters are glued together without blanks inside the sentence. So we need a change of variable syntax due to this special characteristic of Chinese language.

VIII. CHINESE PROGRAMMING LANGUAGE (CPL)

I started to implement the compiler of Chinese Programming Language since 2003 and finished the implementation in 2009.
And I have done 4 large applications using Chinese Programming Language till now.

1. Star Explorer Operating System, finished on 2017 Oct.
2. CPL IDE(Integrated Development Environment), finished on 2017 Nov.
3. Mahjung game, finished on 2019 Jan.
4. Tensor Calculation in mathematics of physics, finished on 2021 Feb.

IX. DIFFERENCES BETWEEN CHINESE PROGRAMMING LANGUAGE AND OTHER 4 NATURAL LANGUAGE PROGRAMMINGS :HYPERTALK[6], APPLESCRIPT[7], INFORM 7[8], COBOL[9].

1. Variables in other 4 NLPs do not start with $.

Missing the $ sign for marking variable will result in ambiguous parsing of statement. An example is like this :
A(variable) B : is a function
A B(variable) : is a function
A(variable) B(variable) : is a function
If we call
A B
The 3 functions are all candidates of the call. Hence ambiguous parsing happens. And this is why CPL marks the variable with $ sign. In CPL, the case is written as
$A B : is a function F1
A $B : is a function F2
$A $B : is a function F3
Call A B => no function defined
Call $A B =>F1 is chosen
Call A $B => F2 is chosen
Call $A $B => F3 is chosen
Therefore no ambiguity happens with the mark of $ for variables in CPL.

2. Other 4 NLPs do not have parenthesis () enclosing function calls

Missing the parenthesis () enclosing function calls will also lead to ambiguous parsing of statement. An example is as follows :

A B : is a function call that returns X
B C : is a function call that returns Y
X C : is a function call
A Y : is a function call

A statement
A B C
will result in ambiguous parses :
(A B) C => X C
A (B C) => A Y

However in CPL, programmers must specify the function calls by enclosing the expressions by parenthesis (). If choosing to call A B, then enclose it by (A B).
So (A B) C =>parsed to be X C
Similarly, if choosing to call B C, then enclose it by (B C).
So A (B C) => parsed to be A Y

Hence no ambiguity happens by enclosing function calls by parenthesis () in CPL.

1. Other 4 NLPs do not have user defined throw-variables.
2. Other 4 NLPs do not have for-each loop for user defined indexed objects.

X. CONCLUSION

A natural language programming syntax is proposed in contrast with the mathematical function syntax. I use the new constructor syntax in natural language programming to solve a classical problem of constructor in Object Oriented Programming. And then a natural-language-like structure“throw variable” is introduced. Also I need to change the variable syntax in Chinese programming language due to the special characteristic of Chinese language. Finally I compare this work with other four natural languageprogrammings.

REFERENCES

[1] [1] Java https://en.wikipedia.org/wiki/Java_(programming_language)
[2] C https://en.wikipedia.org/wiki/The_C_Programming_Language
[3] C++ https://en.wikipedia.org/wiki/The_C%2B%2B_Programming_Language
[4] Pascal https://en.wikipedia.org/wiki/Pascal_(programming_language)
[5] Prolog https://en.wikipedia.org/wiki/Prolog
[6] HyperTalk https://en.wikipedia.org/wiki/HyperTalk
[7] AppleScript https://en.wikipedia.org/wiki/AppleScript
[8] Inform 7 https://en.wikipedia.org/wiki/Inform#Inform_7
[9] COBOL https://en.wikipedia.org/wiki/COBOL