

# Language Engineering Notes

*paraphrased by* Tyler Wright

*An important note, these notes are absolutely **NOT** guaranteed to be correct, representative of the course, or rigorous. Any result of this is not the author's fault.*

# 1 Syntax and Semantics

## 1.1 What is Syntax?

Syntax is the grammatical structure of a program. For example, for the program  $x = y; y = z; z = x;$ , syntactic analysis of this program would conclude that we have three statements concluded with ';'. Each of said statements are variables followed by the composite symbol '=' and another variable.

## 1.2 What are Semantics?

The semantics of a program are what the program evaluates to or rather, the meaning of a syntactically correct program. For example,  $x = y$  evaluates to setting the value of  $x$  to the value of  $y$ .

# 2 Mathematical Notation

## 2.1 Relations

For two sets  $X, Y$  we have that  $f$  is a relation from  $X$  to  $Y$  if  $f \subseteq X \times Y$ .

## 2.2 Total Functions

A total function from  $X$  to  $Y$  is a function  $f : X \rightarrow Y$  that maps each value in  $X$  to a value in  $Y$ .

## 2.3 Partial Functions

A partial function from  $X$  to  $Y$  is a function  $f : X \hookrightarrow Y$  where for some  $X' \subseteq X$  we have that  $f : X' \rightarrow Y$  is total.

## 3 The While Language

### 3.1 Syntactic Categories

For this language, we have five syntactic categories:

- **Numerals** (Num), denoted by  $n$
- **Variables** (Var), denoted by  $x$
- **Arithmetic Expressions** (Aexp), denoted by  $a$
- **Boolean Expressions** (Bexp), denoted by  $b$
- **Statements** (Stm), denoted by  $S$ .

We assume that the numerals and variables are defined elsewhere (for example, the numerals could be strings of digits and the variables could be strings of letters). The other structures are detailed below:

$$\begin{aligned} a &:= n \mid x \mid a_1 + a_2 \mid a_1 \star a_2 \mid a_1 - a_2 \\ b &:= \text{true} \mid \text{false} \mid a_1 == a_2 \mid a_1 \leq a_2 \mid \neg b \mid b_1 \wedge b_2 \\ S &:= x = a \mid \text{skip} \mid S_1; S_2 \mid \text{if } b \text{ then } S_1 \text{ else } S_2 \mid \text{while } b \text{ do } S. \end{aligned}$$

We define representations of each structure in terms of itself as composite elements and the rest are basis elements.

### 3.2 State

When considering a statement, it is almost always important to consider the state of the program as the role of statements is to change the state of the program. For example, when evaluating  $x + 1$ , we must consider what  $x$  is. This requires us to develop the notion of state.

We define state as a set of mappings from variables to values:

$$\mathbf{State} = \{f : \mathbf{Var} \rightarrow \mathbb{Z}\}.$$

This is commonly listed out. For example, if  $x$  maps to 4 and  $y$  maps to 5, we have our state equal to  $\{x \mapsto 4, y \mapsto 5\}$ .

### 3.3 Semantic Functions

The use of semantic functions is to convert syntactic elements into its meaning. Each of the semantic styles will be used to define said semantic functions for variable and statements. However, for numerals, arithmetic and boolean expressions, they are defined as follows.

#### 3.3.1 Numerals

If we assume our numerals are in base-2 (binary), we can define a numeral as follows:

$$n := 0 \mid 1 \mid n0 \mid n1 \mid.$$

In order to determine the value represented by a numeral, we define a *semantic function*  $\mathcal{N} : \mathbf{Num} \rightarrow \mathbb{Z}$ :

$$\begin{aligned}\mathcal{N}(0) &= 0 \\ \mathcal{N}(1) &= 1 \\ \mathcal{N}(n0) &= 2 \cdot \mathcal{N}(n) \\ \mathcal{N}(n1) &= 2 \cdot \mathcal{N}(n) + 1\end{aligned}$$

This definition is called *compositional* as it simply defines an output for each way of constructing a numeral.

*Much like Linear Algebra, it is important to differentiate the numeral '1' and the integer 1 and similarly for zero.*

#### 3.3.2 Arithmetic Expressions

For a given arithmetic expression, it may be necessary to analyse the state to determine the result. Thus, the semantic expression for arithmetic expressions is a function  $\mathcal{A} : \mathbf{Aexp} \rightarrow (\mathbf{State} \rightarrow \mathbb{Z})$ .

This means providing  $\mathcal{A}$  with an arithmetic expression gives us a function from a state to a value. So, for a state  $s$ , we can define  $\mathcal{A}$  as follows:

$$\begin{aligned}\mathcal{A}(n)(s) &= \mathcal{N}(n) \\ \mathcal{A}(x)(s) &= s(x) \\ \mathcal{A}(a_1 + a_2)(s) &= \mathcal{A}(a_1)(s) + \mathcal{A}(a_2)(s) \\ \mathcal{A}(a_1 \star a_2)(s) &= \mathcal{A}(a_1)(s) \star \mathcal{A}(a_2)(s) \\ \mathcal{A}(a_1 - a_2)(s) &= \mathcal{A}(a_1)(s) - \mathcal{A}(a_2)(s).\end{aligned}$$

Similarly to the numerical semantic function, it's important to differentiate between the syntactic  $+$ ,  $\star$ ,  $-$  in the arithmetic expressions like  $a_1 + a_2$  and the usual arithmetic operations  $+$ ,  $\star$ ,  $-$  between integers.

### 3.3.3 Boolean Expressions

We define the set  $T$  to be  $T = \{\text{tt}, \text{ff}\}$  where  $\text{tt}$  represents truth and  $\text{ff}$  represents falsity.

Similarly to arithmetic expressions we can define  $\mathcal{B} : \mathbf{Bexp} \rightarrow (\mathbf{State} \rightarrow \mathbb{Z})$  for a state  $s$ :

$$\begin{aligned} \mathcal{B}(\text{true})(s) &= \text{tt} \\ \mathcal{B}(\text{false})(s) &= \text{ff} \\ \mathcal{B}(a_1 == a_2)(s) &= \begin{cases} \text{tt} & \text{if } \mathcal{A}(a_1)(s) = \mathcal{A}(a_2)(s) \\ \text{ff} & \text{otherwise} \end{cases} \\ \mathcal{B}(a_1 \leq a_2)(s) &= \begin{cases} \text{tt} & \text{if } \mathcal{A}(a_1)(s) \leq \mathcal{A}(a_2)(s) \\ \text{ff} & \text{if } \mathcal{A}(a_1)(s) > \mathcal{A}(a_2)(s) \end{cases} \\ \mathcal{B}(\neg b)(s) &= \begin{cases} \text{tt} & \text{if } \mathcal{B}(b)(s) = \text{ff} \\ \text{ff} & \text{if } \mathcal{B}(b)(s) = \text{tt} \end{cases} \\ \mathcal{B}(b_1 \wedge b_2)(s) &= \begin{cases} \text{tt} & \text{if } \mathcal{B}(b_1)(s) = \text{tt} \text{ and } \mathcal{B}(b_2)(s) = \text{tt} \\ \text{ff} & \text{if } \mathcal{B}(b_1)(s) = \text{ff} \text{ or } \mathcal{B}(b_2)(s) = \text{ff} \end{cases} \end{aligned}$$

## 3.4 Free Variables

We will later be interested in the 'free variables' of an expression (the set of variables occuring within it).

### 3.4.1 Arithmetic Expressions

We define  $F_V : \mathbf{Aexp} \rightarrow \{\mathbf{Var}\}$  by:

$$\begin{aligned} F_V(n) &:= \emptyset \\ F_V(x) &:= \{x\} \\ F_V(a_1 + a_2) &:= F_V(a_1) \cup F_V(a_2) \\ F_V(a_1 \star a_2) &:= F_V(a_1) \cup F_V(a_2) \\ F_V(a_1 - a_2) &:= F_V(a_1) \cup F_V(a_2). \end{aligned}$$

Thus, we have developed a way to decompose expressions and generate all the variables said expression depends on. This is formalised for states  $s_1, s_2$  and an arithmetic expression  $a$ :

$$[\forall x \in F_V(a)][s_1(x) = s_2(x)] \implies [\mathcal{A}(a)(s_1) = \mathcal{A}(a)(s_2)].$$

### 3.4.2 Boolean Expressions

We define  $F_V : \mathbf{Bexp} \rightarrow \{\mathbf{Var}\}$  by:

$$\begin{aligned} F_V(\mathbf{true}) &:= \emptyset \\ F_V(\mathbf{false}) &:= \emptyset \\ F_V(a_1 == a_2) &:= F_V(a_1) \cup F_V(a_2) \\ F_V(a_1 \leq a_2) &:= F_V(a_1) \cup F_V(a_2) \\ F_V(\neg b) &:= F_V(b) \\ F_V(b_1 \wedge b_2) &:= F_V(b_1) \cup F_V(b_2). \end{aligned}$$

Similarly, for states  $s_1, s_2$  and an boolean expression  $b$ :

$$[\forall x \in F_V(b)][s_1(x) = s_2(x)] \implies [\mathcal{B}(b)(s_1) = \mathcal{B}(b)(s_2)].$$

## 3.5 Substitutions

Within expressions, we will be interested in swapping the occurrences of a variable with another expression. For an expression  $e$  with  $x$  in  $F_V(e)$  and another expression  $e_0$ , we write  $e[x \mapsto e_0]$  for the expression  $e$  where  $x$  is substituted for  $e_0$ .

### 3.5.1 Arithmetic Expressions

For an arithmetic expression  $a_0$ , we can define substitution on arithmetic expressions:

$$\begin{aligned} n[y \mapsto a_0] &:= n \\ x[y \mapsto a_0] &:= \begin{cases} a_0 & \text{if } x = y \\ x & \text{otherwise} \end{cases} \\ (a_1 + a_2)[y \mapsto a_0] &:= (a_1[y \mapsto a_0]) + (a_2[y \mapsto a_0]) \\ (a_1 \star a_2)[y \mapsto a_0] &:= (a_1[y \mapsto a_0]) \star (a_2[y \mapsto a_0]) \\ (a_1 - a_2)[y \mapsto a_0] &:= (a_1[y \mapsto a_0]) - (a_2[y \mapsto a_0]). \end{aligned}$$

### 3.5.2 Boolean Expressions

For an arithmetic expression  $a_0$ , we can define substitution on boolean expressions:

$$\begin{aligned}(\mathbf{true})[x \mapsto a_0] &:= \mathbf{true} \\(\mathbf{false})[x \mapsto a_0] &:= \mathbf{false} \\(a_1 == a_2)[x \mapsto a_0] &:= (a_1[x \mapsto a_0]) == (a_2[x \mapsto a_0]) \\(a_1 \leq a_2)[x \mapsto a_0] &:= (a_1[x \mapsto a_0]) \leq (a_2[x \mapsto a_0]) \\(\neg b)[x \mapsto a_0] &:= \neg(b)[x \mapsto a_0] \\(b_1 \wedge b_2)[x \mapsto a_0] &:= (b_1)[x \mapsto a_0] \wedge (b_2)[x \mapsto a_0].\end{aligned}$$

### 3.5.3 States

A similar process applies to states, for a state  $s$  and the variables  $x, v$ :

$$(s[y \mapsto v])(x) = \begin{cases} v & \text{if } x = y \\ s(x) & \text{otherwise.} \end{cases}$$

## 4 Operational Semantics

### 4.1 Overview of Operational Semantics

An operational explanation of the meaning of a construct will explain how to execute said construct. For example, in C, the semicolons provide chronology and the  $=$  symbol demonstrates assignment. The semantics are not merely interested in the results of a program.

More specifically, this means we are interested in how the state of the program changes. To this end, there are two approaches we shall consider:

- **Natural Semantics:** to describe how the overall results of executions are obtained
- **Structural Operational Semantics:** to describe how the individual steps of computation take place.

### 4.2 Rules and Derivation Trees

#### 4.2.1 Transitions

A transition is a construct representing the transition of a state  $s$  to a state  $s'$  via some statement  $S$ , denoted by  $\langle S, s \rangle \rightarrow s'$ .

### 4.2.2 Rules

We can define 'rules' which are of the form:

$$\frac{\langle S_1, s_1 \rangle \rightarrow s_2, \dots, \langle S_n, s_n \rangle \rightarrow s'}{\langle S, s \rangle \rightarrow s'} \quad \text{if } \dots$$

Where the premises consisting of immediate constituents of a statement  $S$  or compositions of said immediate constituents, are written above the conclusion. Conditions may be written to the right of the rule which are necessary for when the rule is applied.

### 4.2.3 Axioms

For a rule where the set of premises is empty, we call such a rule an *axiom*. Axioms with meta-variables are called *axiom schema* where we can obtain an *instance* of an axiom by defining particular variables.

### 4.2.4 Derivation Trees

A program's execution can be modelled by a 'derivation tree' formed from rules. For example, the program  $x = y; y = z; z = x;$  can be written as:

$$\frac{\frac{\langle z = x, s_{570} \rangle \rightarrow s_{575} \quad \langle x = y, s_{575} \rangle \rightarrow s_{775}}{\langle z = x; x = y, s_{570} \rangle \rightarrow s_{775}} \quad \langle y = z, s_{775} \rangle \rightarrow s_{755}}{\langle x = y; y = z; z = x, s_{570} \rangle \rightarrow s_{755}}.$$

Where  $s_{abc}$  represents the state where  $\{x \mapsto a, y \mapsto b, z \mapsto c\}$ .

This tree represents the decomposition of the composite statement at the bottom of the tree  $x = y; y = z; z = x;$ . As  $z = x$  transforms the state  $s_{570}$  to  $s_{575}$  and  $x = y$  transforms the state  $s_{575}$  to  $s_{775}$ , we know that the composite statement  $z = x; x = y$  transforms  $s_{570}$  to  $s_{775}$ .



### 4.3 Natural Semantics

During this section, we will use  $\langle S, s \rangle$  representing the statement  $S$  being executed from the state  $s$  and simply  $s$  to represent a terminal state.

We can define some natural semantics for the While language as follows:

$$\begin{array}{ll}
\text{[assignment]} & \langle x = a, s \rangle \rightarrow s[x \mapsto \mathcal{A}(a)(s)] \\
\\
\text{[skip]} & \langle \text{skip}, s \rangle \rightarrow s \\
\\
\text{[composition]} & \frac{\langle S_1, s \rangle \rightarrow s' \quad \langle S_2, s' \rangle \rightarrow s''}{\langle S_1; S_2, s \rangle \rightarrow s''} \\
\\
\text{[if]} & \left\{ \begin{array}{ll} \frac{\langle S_1, s \rangle \rightarrow s'}{\langle \text{if } b \text{ then } S_1 \text{ else } S_2, s \rangle \rightarrow s'} & \text{if } \mathcal{B}(b)(s) = \text{tt} \\ \frac{\langle S_2, s \rangle \rightarrow s'}{\langle \text{if } b \text{ then } S_1 \text{ else } S_2, s \rangle \rightarrow s'} & \text{if } \mathcal{B}(b)(s) = \text{ff} \end{array} \right. \\
\\
\text{[while]} & \left\{ \begin{array}{ll} \frac{\langle S, s \rangle \rightarrow s' \quad \langle \text{while } b \text{ do } S, s' \rangle \rightarrow s''}{\langle \text{while } b \text{ do } S, s \rangle \rightarrow s''} & \text{if } \mathcal{B}(b)(s) = \text{tt} \\ \langle \text{while } b \text{ do } S, s \rangle \rightarrow s & \text{if } \mathcal{B}(b)(s) = \text{ff} \end{array} \right.
\end{array}$$

We have that the semantics detailed above are deterministic, that is, for identical inputs, we get identical outputs.

As there will be more than one definition of these semantics we will subscript these definitions with  $ns$  where appropriate.

### 4.3.1 The Semantic Function

We can now condense the meaning of statements into a partial function from **State** to **State**. We define the semantic function as follows:

$$\mathcal{S} : \mathbf{Stm} \rightarrow (\mathbf{State} \hookrightarrow \mathbf{State}).$$

Thus for a given statement  $S$ , we have that:

$$\mathcal{S}(S) : \mathbf{State} \hookrightarrow \mathbf{State}$$

$$s \mapsto \begin{cases} s' & \text{if } \langle S, s \rangle \rightarrow s' \\ \text{undefined} & \text{otherwise.} \end{cases}$$

We that this function is partial because some statements may never terminate. Thus giving an indeterminate answer.

As there will be more than one semantic function (for our different types of semantics) we denote this function as  $\mathcal{S}_{ns}$ .

## 4.4 Structural Operational Semantics

In structural operation semantics, there is a particular emphasis on the individual steps of execution. The transition relation in these semantics is of the form:

$$\langle S, s \rangle \Rightarrow \gamma,$$

where  $\gamma$  is of the form:

$\langle S', s' \rangle$	denoting that $S$ and $s$ have transformed in some way and that the execution is not yet complete,
$s'$	denoting that the execution has completed with some terminal state $s'$ .

We shall say that  $\langle S, s \rangle$  is **stuck** if there does not exist a  $\gamma$  such that  $\langle S, s \rangle \Rightarrow \gamma$  and **unstuck** otherwise.

We can define some structural operational semantics for the While language as follows:

$$[\text{assignment}] \quad \langle x = a, s \rangle \Rightarrow s[x \mapsto \mathcal{A}(a)(s)]$$

$$[\text{skip}] \quad \langle \text{skip}, s \rangle \Rightarrow s$$

$$[\text{composition}] \quad \frac{\langle S_1, s \rangle \Rightarrow \langle S'_1, s' \rangle}{\langle S_1; S_2, s \rangle \Rightarrow \langle S'_1; S_2, s' \rangle} \quad \text{or} \quad \frac{\langle S_1, s \rangle \Rightarrow s'}{\langle S_1; S_2, s \rangle \Rightarrow \langle S_2, s' \rangle}$$

$$[\text{if}] \quad \begin{cases} \langle \text{if } b \text{ then } S_1 \text{ else } S_2, s \rangle \Rightarrow \langle S_1, s \rangle & \text{if } \mathcal{B}(b)(s) = \text{tt} \\ \langle \text{if } b \text{ then } S_1 \text{ else } S_2, s \rangle \Rightarrow \langle S_2, s \rangle & \text{if } \mathcal{B}(b)(s) = \text{ff} \end{cases}$$

$$[\text{while}] \quad \langle \text{while } b \text{ do } S, s \rangle \Rightarrow \langle \text{if } b \text{ then } (S; \text{while } b \text{ do } S) \text{ else skip}, s \rangle$$

We can see by the definition of composition that with these semantics we have to fully decompose and execution  $S_1$  before we can begin execution of  $S_2$ .

As there will be more than one definition of these semantics we will subscript these definitions with  $sos$  where appropriate.

## 4.5 Derivation Sequences

For a statement  $S$  and state  $s$ , we can consider the sequence of configurations of  $S$  in  $s$ . The length of such a sequence is either:

- **Finite:**  $\gamma_0, \gamma_1, \gamma_2, \dots, \gamma_k$  for some  $k$  in  $\mathbb{Z}_{\geq 0}$  where each  $i$  in  $\{0, 1, \dots, k-1\}$  satisfies  $\gamma_0 = \langle S, s \rangle$ ,  $\gamma_i \Rightarrow \gamma_{i+1}$  where  $\gamma_k$  is terminal or stuck
- **Infinite:**  $(\gamma_k)_{k \geq 0}$  where for each  $i$  in  $\mathbb{Z}_{\geq 0}$  satisfies  $\gamma_0 = \langle S, s \rangle$ ,  $\gamma_i \Rightarrow \gamma_{i+1}$ .

For two configurations  $\gamma_i, \gamma_k$ , we write  $\gamma_i \Rightarrow^k \gamma_j$  to denote that there are  $k$  steps in execution between the configurations or  $\gamma_i \Rightarrow^* \gamma_j$  to denote that there's some finite number of steps between the two configurations.

We say for a statement  $S$  and state  $s$ , the execution of  $S$  on  $s$ :

- **Terminates:** if and only if there's a finite derivation sequence starting with  $\langle S, s, \rangle$
- **Loops:** if and only if there's a infinite derivation sequence starting with  $\langle S, s \rangle$ .

Thus, statements can loop on some states and terminate on others. Similarly, some states always terminate and always loop.