

Chapter 1

Preface

1.1 Motivation

The halting problem is undecidable in general, however this property is often abused to deduce that for all programs. The intent of this project is to explore some context in which the halting property *is* decidable, and to analyze how useful this indeed is.

1.2 Expectations of the reader

The reader is expected to have a background in computer science on a graduate level or higher. In particular, it is expected that the reader is familiar with basic concepts of compilers, computability and complexity, which are subject to basic undergraduate courses at the state of writing. Furthermore, the reader is expected to be familiar with discrete mathematics and the concepts of functional programming languages. Ideally, the reader should know at least one purely functional programming language.

For those still in doubt, it is expected that the following terms can be used without definiiton:

- Algorithm, Recursion, Induction, Big O Notation
- Regular Expressions (preg syntax)
- Backus-Naur Form
- Turing Machine, Halting Problem
- List, Binary tree, Head, Tail
- Discrete mathematics

Chapter 2

On the general uncomputability of the halting problem

2.1 Computable problems and effective procedures

A computable problem is a problem that can be solved by an effective procedure.

A problem can be solved by an effective procedure iff the effective procedure is well-defined for the entire problem domain¹, and iff passing a value from the domain as input to the procedure *eventually* yields a correct result (to the problem) as output of the procedure. That is, an effective procedure can solve a problem if it computes an injective partial function that associates the problem domain with the range of solutions to the problem.

An effective procedure is discrete, in the sense that computing the said function cannot take an infinite amount of time. To do this, an effective procedure makes use of a finite sequence of steps that themselves are discrete. This has a few inevitable consequences for the input and output values, namely that they themselves must be discrete and that there must be a discrete number of them².

Proof. An infinite value cannot be processed nor produced by a finite sequence of discrete steps. □

An effective procedure is also deterministic, in the sense that passing the same input value always yields the same output value. This means that all of the steps of the procedure that are relevant to it's output³ are themselves deterministic.

Proof. If a procedure made use of a stochastic process to yield a result, that stochastic process would have to yield the output for the same input if the global deterministic property of the procedure is to be withheld. This is clearly absurd. □

In effect, a procedure can be said to comprise of a finite sequence of other procedures, which themselves may comprise of other procedures, however, all procedures eventually bottom out, in that a finite sequence of composite procedures can always be replaced by a finite sequence of basic procedures that are implemented in underlying hardware.

- effective procedure
- effectively decidable
- effectively enumerable

¹Invalid inputs are, in this instance, irrelevant.

²A finite sequence of discrete values can be trivially encoded as a single discrete value.

³All other steps can be omitted without loss of generality.

2.2 Enumerability

2.2.1 Enumerable sets

Enumerable sets, or equivalently countable or recursively enumerable sets, are sets that can be put into a one-to-one correspondence to the set of natural numbers \mathbb{N} , more specifically:

Definition 1. *An enumerable set is either the empty set or a set whose elements can be placed in a sequence s.t. each element gets a consecutive number from the set of natural numbers \mathbb{N} .*

2.2.2 Decidability

Definition 2. *A problem is decidable if there exists an algorithm that for any input event*

- Recursively enumerable – countable sets
- Co-recursively enumerable

2.3 Cantor's diagonalization

Cantor's diagonalization argument is a useful argument for proving unenumerability of a set and hence its uncomputability.

The original proof shows that the set of infinite bit-sequences is not enumerable.

Proof. Assume that sequence S is an infinite sequence of infinite sequences of bits. The claim is that regardless of the number of bit-sequences in S it is always possible to construct a bit-sequence not contained in S .

Such a sequence can be represented as a table:

Such a sequence is constructable by taking the complements of the elements along the diagonal of all

□

2.4 The halting problem

2.5 Rice's statement

2.6 Primitive recursion

All primitive recursive programs terminate.

2.7 Introduction to size-change termination

The size change termination .. why values should be well-founded

2.8 The language to be defined

The soft version.

Chapter 3

D

For the purposes of describing size-change termination we'll consider a language D. The following chapter describes the syntax and semantics of the language.

3.1 General properties

The intent of the language is for it be used to explain concepts such as size-change termination. One of the fundamental concepts required of the language of application is that its datatypes are well-founded. That is, any subset S of the range of values of some well-defined type has a value s s.t. $\forall s' \in S \ s \leq s'$. This makes it ideal to choose some oversimplistic data type structure rather than an army of basic types. Besides, an appropriately defined basic data type should be able to represent arbitrarily complex data values.

The language is initially first-order since the size-change termination principle is first described for first-order programs later on in this work. However, the language is designed so that it is easy to turn it into a high-level language without much effort. This may prove necessary as we try to expand size-change termination to higher-order programs.

The language is a call-by-value and purely functional to avoid any problems that could arise from regarding lazy programs or where the notion of a global state of the machine is relevant. Simply put, this is done to ensure elegance of further proof with the help of the language.

3.2 Data

D is a simple language where the emphasis is on the sizes of data. Hence, the way that data values are constructed does not have to be particularly practical, but all values have to be well founded and easily comparable.

The language D is untyped, and represents all data in terms of *unlabeled ordered binary trees*, henceforth referred to as simply, *binary trees*. Such a tree is recursively defined as follows:

Definition 3. A binary tree is a set that is either empty or contains a single unlabeled node with two binary trees as its left and right child, respectively.

For simplicity, we'll sometimes refer to the empty set as a *leaf*, and the unlabeled node simply as *node*.

Definition 4. The set of all possible binary trees is denoted by \mathbb{B} .

To operate on such trees we'll require a few basic things, namely a representation of leaves, recursive construction and destruction of nodes, as well as a way to tell nodes and leaves apart. Most of these we'll be derived in the operational semantics § 3.4 (6), however, we do require the following basic definitions:

Definition 5. *The leaf with the atom 0. Clearly, $0 \in \mathbb{B}$.*

Definition 6. *The function $\cdot : \mathbb{B} \times \mathbb{B} \rightarrow \mathbb{B}$ constructs a node with the two arguments as it's left and right child, respectively.*

3.3 Syntax

We describe the syntax of D in terms of an extended Backus-Naur form¹. This is a core syntax definition, and other, more practical, syntactical features may be defined later on as needed. The initial non-terminal is `<program>`.

$$\langle \text{program} \rangle ::= \langle \text{declaration} \rangle^* \langle \text{expression} \rangle \quad (3.1)$$

$$\langle \text{expression} \rangle ::= \langle \text{element} \rangle (\langle \text{'.'} \rangle \langle \text{expression} \rangle) ? \quad (3.2)$$

$$\langle \text{element} \rangle ::= \langle \text{'0'} \rangle \mid \langle \text{'('} \rangle \langle \text{element} \rangle \langle \text{'')'} \rangle \mid \langle \text{'name'} \rangle \mid \langle \text{application} \rangle \quad (3.3)$$

$$\langle \text{application} \rangle ::= \langle \text{'name'} \rangle \langle \text{expression} \rangle^+ \quad (3.4)$$

$$\langle \text{declaration} \rangle ::= \langle \text{'name'} \rangle \langle \text{pattern} \rangle^+ \langle \text{' := ' } \rangle \langle \text{expression} \rangle \quad (3.5)$$

$$\langle \text{pattern} \rangle ::= \langle \text{pattern-value} \rangle (\langle \text{'.'} \rangle \langle \text{pattern} \rangle) ? \quad (3.6)$$

$$\langle \text{pattern-value} \rangle ::= \langle \text{'0'} \rangle \mid \langle \text{'_'} \rangle \mid \langle \text{'('} \rangle \langle \text{pattern} \rangle \langle \text{'')'} \rangle \mid \langle \text{'name'} \rangle \quad (3.7)$$

$$\langle \text{'name'} \rangle ::= [\langle \text{'a'} \rangle - \langle \text{'z'} \rangle] ([\langle \text{'-' } \rangle \langle \text{'a'} \rangle - \langle \text{'z'} \rangle]^* [\langle \text{'a'} \rangle - \langle \text{'z'} \rangle]) ? \quad (3.8)$$

0-ary declarations are disallowed to avoid having to deal with constants in general.

The term `'_'` in `<pattern-value>` is the conventional wildcard operator; it indicates a value that won't be used by the declaration, but allows us to keep the same declaration signature. We hence define the *signature* of a declaration as follows:

Definition 7. *A declaration signature in D consists of the function name and the number of parameters it has.*

We'll adopt the Erlang-like notation when talking about function signatures, i.e. if we have a function `less` having two arguments in its signature, we'll refer to it as `less/2`.

3.4 Semantics

In the following section we describe the semantics of D using structured operational semantics. The syntax used to define the reduction rules is largely equivalent to the Aarhus report[3], but differs slightly². The most notable about the syntax used here is the following:

- Rules should be read in increasing order of equation number.
- If some rule with a lower equation number makes use of an undefined reduction rule, it is because the reduction rule is defined under some higher equation number.
- Rules can be defined in terms of themselves, i.e. they can be recursive, even mutually recursive.

The syntax aside, Table 3.1 (7) defines a few lower-letter shorthands for various constructs. Additionally, we'll let the capital equivalents of these letters represent a set of the respective construct, as well as let the atoms 0 and `_` represent themselves in the reduction rules. It is also worth noting that $\forall v \in V : v \in \mathbb{B}$.

¹The extension lends some constructs from regular expressions to achieve a more concise dialect. The extension is described in detail in Appendix A (15).

²The syntax applied here is described in further detail in Appendix B (17).

Notation	Description
x	expression
e	element
p	pattern
v	value
n	name

Table 3.1: Overview of the shorthands used in the reduction rules for D.

3.4.1 The memory model

The memory is considered in terms of a set of stacks σ . Every stack has a unique identifier $n \in N$, that is, a stack represents a variable name. As we give a variable a value, we push that value on top of the corresponding stack. We pop the value off the corresponding stack as we leave the scope of the variable value.

Due to D being a first-order language, we should also separate the function and variable spaces. We'll represent these by ϕ and β , respectively. Whenever we use σ , ϕ or β in set notation, we imply the sets of the names of functions and variables, and not the stacks themselves corresponding to those names. Hence, $\sigma = \phi \cup \beta$, and to keep D first-order we add the limitation, $\phi \cap \beta = \emptyset$.

Making D higher order

The only change that this would require is to let $\phi = \beta = \sigma$.

3.4.2 Declaration

A declaration with a name n , a *non-empty* pattern list $[p]$ and an expression e is stored in the function space ϕ :

$$\frac{\langle \phi(n) \leftarrow \langle P, x \rangle \rangle \Rightarrow \phi'}{\langle n, P, x, \phi \rangle \Rightarrow \phi'} \quad (3.9)$$

3.4.3 Expression evaluation

An expression x is either the element e , or a construction of an element e' with another expression x' . That is, the binary infix operator \cdot is right-associative, and has the following operational semantics:

$$\frac{\frac{x \Rightarrow e}{\langle e, \sigma \rangle \Rightarrow \langle v, \sigma \rangle} \vee \frac{x \Rightarrow e' \cdot x'}{\frac{\langle e', \sigma \rangle \Rightarrow \langle v', \sigma \rangle \wedge \langle x', \sigma \rangle \Rightarrow \langle v'', \sigma \rangle}{v' \cdot v'' \Rightarrow v}}}{\langle x, \sigma \rangle \Rightarrow \langle v, \sigma \rangle} \quad (3.10)$$

3.4.4 Element evaluation

According to the syntax specification, an element of an expression can either be the atom 0, or an application. We'd like to distinguish between variables and functions, and we do that

$$\frac{(e \Rightarrow 0 \wedge v \equiv 0) \vee \frac{e \Rightarrow n}{\beta(n) \Rightarrow v} \vee \frac{e \Rightarrow \langle n, X \rangle}{\langle n, X, \sigma \rangle \Rightarrow \langle v, \sigma \rangle}}{\langle e, \sigma \rangle \Rightarrow \langle v, \sigma \rangle} \quad (3.11)$$

3.4.5 Function application

$$\frac{\frac{\langle n, \phi \rangle \Rightarrow \langle P, x, \phi \rangle}{\langle P, X, \sigma \rangle \Rightarrow \sigma'} \quad \frac{\langle x, \sigma' \rangle \Rightarrow \langle v, \sigma' \rangle}{\langle n, X, \sigma \rangle \Rightarrow \langle v, \sigma \rangle}}{\langle n, X, \sigma \rangle \Rightarrow \langle v, \sigma \rangle} \quad (3.12)$$

3.4.6 Pattern matching

$$\frac{\frac{\langle P_{head}, X_{head}, \sigma \rangle \Rightarrow \sigma''}{\langle P_{tail}, X_{tail}, \sigma'' \rangle \Rightarrow \sigma'} \quad \langle P, X, \sigma \rangle \Rightarrow \sigma'}{\langle P, X, \sigma \rangle \Rightarrow \sigma'} \quad (3.13)$$

$$\frac{\langle I, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle \vee \langle Z, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle \vee \langle N, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle \vee \langle P, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle}{\langle p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle} \quad (3.14)$$

For the sake of an elegant notation, we'll override the function \cdot for patterns.

Definition 8. A pattern is an unlabeled of binary tree which is either empty or consists of an unlabeled node with a 0, $_$ name, or a pattern as it's left and right child.

Definition 9. Let the set of all possible patterns be denoted by \mathbb{P} .

Definition 10. The function $\cdot : \mathbb{P} \times \mathbb{P} \rightarrow \mathbb{P}$ constructs a pattern node with the two arguments as it's left and right child, respectively.

$$\frac{p \Rightarrow _ \cdot p' \wedge x \Rightarrow e \cdot x' \wedge \sigma \Rightarrow \sigma'}{\langle \text{UNDERSCORE}, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle} \quad (3.15)$$

$$\frac{p \Rightarrow 0 \cdot p' \wedge x \Rightarrow e \cdot x' \wedge \sigma \Rightarrow \sigma'}{\langle \text{ZERO}, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle} \quad (3.16)$$

$$\frac{\frac{p \Rightarrow n \cdot p' \wedge x \Rightarrow e \cdot x'}{\langle e, \sigma \rangle \Rightarrow \langle v, \sigma \rangle} \quad \frac{\langle \sigma(n) \leftarrow v \rangle \Rightarrow \sigma'}{\langle \text{NAME}, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle}}{\langle \text{NAME}, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle} \quad (3.17)$$

$$\frac{\frac{p \Rightarrow p'' \cdot p' \wedge x \Rightarrow x'' \cdot x'}{\langle p'', x'', \sigma \rangle \Rightarrow \sigma'} \quad \langle \text{PATTERN}, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle}{\langle \text{PATTERN}, p, x, \sigma \rangle \Rightarrow \langle p', x', \sigma' \rangle} \quad (3.18)$$

3.5 Size

For the purposes of talking about size-change termination, we also need to define the notion of size, and be sure to do so in such a way so that all possible data values are well-founded.

Definition 11. Size of a value in D is the number of nodes in the tree representing that value.

The “well-foundedness” of D 's data values, given such a definition can be argued for by proving a bijective relation between \mathbb{B} and \mathbb{N} . This would imply that we can define the relation $<$ on D 's data values, which we know to be well-founded.

We start by formally proving that Definition 11 (8) yields a many-to-one mapping of D 's data values to the natural numbers.

First, we prove, by induction, that any natural number can be represented in D :

Proof.

Base The atom 0 has no nodes, and hence represents the value 0.

Assumption If we can represent the $n \in \mathbb{N}$ in D, then we can also represent the number $n + 1 \in \mathbb{N}$.

Induction Let n be represented by some binary tree A , then $n + 1$ can be represented by $0 \cdot A$.

□

Second, we prove, also by induction, that any value in D has one and only one representation in \mathbb{N} .

Proof.

Base The atom 0 has no nodes, and hence corresponds only to the value 0.

Assumption

1. If the binary tree A has only one representation $n \in \mathbb{N}$, then $|0 \cdot A| \equiv n + 1$ and $|A \cdot 0| \equiv n + 1$.
2. If the binary tree A has only one representation $n \in \mathbb{N}$, and the binary tree B has only one representation $m \in \mathbb{N}$, then $|A \cdot B| \equiv n + m + 1$ and $|B \cdot A| \equiv n + m + 1$.

Induction By definition of the binary function \cdot , any given node A with left child A_{left} and right child A_{right} has the size:

$$|A| = 1 + |A_{left}| + |A_{right}|$$

Hence, any value in D must have one and only one representation in \mathbb{N} .

□

Definition 11 (8) *almost* allows us to devise an algorithm to compare the sizes of data values. The problem withstanding is that two different values can have rather diverging tree representations. Hence, comparing them, using only the operations defined in § 3.4 (6), is seemingly impossible unless we initially, or along the way, transform the binary trees being compared into some sort of a *standard representation*. We'll define this representation, recursively, as follows:

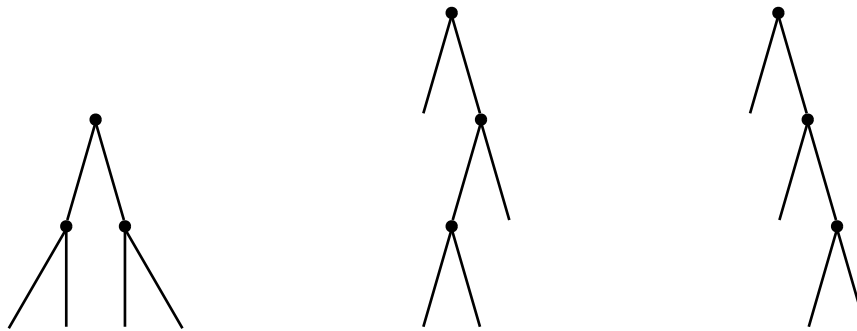
Definition 12. A binary tree in standard representation is a binary tree that either is a leaf or a node having a leaf as it's left child and a binary tree in standard representation as it's right child.

Intuitively, a binary tree in standard representation is just a tree that only descends along the right side. Comparing the sizes of two trees in this representation is just a matter of walking the descending in the two trees simultaneously, until one of them, or both, bottoms out. If there is a tree that bottoms out strictly before another, that is the lesser tree by Definition 11 (8). Figure 3.1 (10) showcases some examples.

3.5.1 normalize/1

```
normalize a = normalize-aux a 0 0
```

```
normalize-aux 0      0      an = an
normalize-aux 0      bl.br an = normalize-aux bl br      an
normalize-aux 0.ar   b      an = normalize-aux ar b      0.an
normalize-aux al.0   b      an = normalize-aux al b      0.an
normalize-aux al.ar  b      an = normalize-aux ar al.b   0.an
```



(a) Not in standard representation (b) Not in standard representation (c) Standard representation

Figure 3.1: Three trees of various shapes but equal size.

Correctness

`normalize/1` makes use of an auxiliary procedure, `normalize-aux/3`, for which we can provide the following argument descriptions:

1. The tree to be normalized.
2. An auxiliary tree.
3. A normalized tree.

The idea of the algorithm is to move right-wise down the tree to be normalized, constructing an auxiliary tree containing all left-wise child nodes, if any.

The return value is the normalized tree, i.e. the third argument. Hence, we must increase the size of the normalized tree each time we move right-wise down the tree to be normalized.

Once we reach the right-most leaf of the tree to be normalized we return the normalized tree if the auxiliary tree is empty. Otherwise, we normalize the right child of the auxiliary tree, with the left child of the auxiliary as the new auxiliary tree, and the normalized tree constructed thus far as the initial normalized tree.

Time complexity

Coming soon..

Space complexity

Coming soon..

3.5.2 `less/2`

We'll define the function `normalize/1` further below to transform any D value into its standard representation. For now we'll assume that we have such a function in scope and define `less/2` for determining whether the value of the first argument is strictly less than the value of the second argument.

In order to define such a boolean-valued function we need a convention for representing the boolean values *true* and *false* in D. We'll adopt the C-like convention:

Definition 13. A false value is represented by a leaf tree. A true value is represented by a non-leaf tree, i.e. a node.

We're now ready to define the function `less/2`:

```
less a b := normalized-less (normalize a) (normalize b)

normalized-less 0 b := b
normalized-less _ 0 := 0
normalized-less _.a _.b := normalized-less a b
```

Correctness

Proof. Given Definition 12 (9), and the assumption that `NORMALIZE(A)` computes the standard representation of A , we know the following:

1. $|A| \equiv |\text{NORMALIZE}(A)|$.
2. We'll walk through all the nodes if we perform a recursive right-child-walk starting at A .
3. The same holds for B .

It is also easy to see from lines ??? that `NORMALIZEDLESS` stops as soon as we reach the "bottom" of either A or B .

Given Definition 11 (8), $A < B$ iff it bottoms out before B , that is, we reach an instance of the recursion where both $\text{IsLeaf}(A)$ and $\text{IsNode}(B)$ hold. In all other cases $A \geq B$, the cases specifically are:

- $\text{IsLeaf}(A)$ and $\text{IsLeaf}(B)$, then $|A| \equiv |B|$.
- $\text{IsNode}(A)$ and $\text{IsLeaf}(B)$, then $|A| > |B|$

Last but not least, due to all data values being finite, eventually one of the trees does bottom out. □

Time complexity

Given that the binary trees A and B are in standard representation when we enter the auxiliary procedure, `NORMALIZEDLESS`, it is fairly easy to get an upper bound on the running time of `NORMALIZEDLESS` itself.

Indeed, the running time of `NORMALIZEDLESS` itself is $O(\text{MAX}(|A|, |B|))$, since we just walk down the trees until one of them bottoms out.

We haven't yet defined the procedure `NORMALIZE` yet. Hence, the only thing that we can say about the running time of `LESS` in general is that it is $O(\text{NORMALIZE}(A) + \text{NORMALIZE}(B) + \text{MAX}(|A|, |B|))$.

Space complexity

Coming soon..

3.5.3 Built-in high-order functions

Although D is initially a first-order language, we will ignore that limitation for a bit and define a few higher-order functions to provide some syntactical sugar to the language. Beyond the discussion in this section, these higher-order functions should be regarded as D built-ins.

Branching

In the following definition, the variable names `true` and `false` refer to expressions to be executed in either case.

```
if 0 _ false := false
if _ _ true := true
```

As you can see, we employ the C convention that any value other than 0 is a “truthy” value, and the expression `true` is returned.

Although the call-by-value nature of the language does not allow for short-circuiting the if-statements defined in such a way, this shouldn’t be any impediment to further analysis.

3.5.4 Sample programs

As an illustration of the language syntax, the following program reverses a tree:

```
reverse 0 := 0
reverse left.right := (reverse right).(reverse left)
```

The following program computes the Fibonacci number `n`:

```
fibonacci 0 x y := 0
fibonacci 0.0 x y := y
fibonacci n x y := fibonacci (minus n 0.0) y (add x y)
```

Bibliography

- [1] P. Naur (ed.), Revised Report on the Algorithmic Language ALGOL 60; CACM, Vol. 6, p. 1; The Computer Journal, Vol. 9, p. 349; Num. Math., Vol. 4, p. 420. (1963); Section 1.1.
- [2] A. M. Turing, On computable numbers with an application to the Entscheidungsproblem; Proceedings of the London Mathematical Society, 42(2):230-265, (1936).
- [3] Gordon D. Plotkin, A Structural Approach to Operational Semantics, Journal of Logic and Algebraic Programming (2004) Volume: 60-61, Issue: January, Publisher: Citeseer, Pages: 17-139.

Appendix A

Extended-BNF

This report makes use of an extended version of the Backus-Naur form (BNF). This appendix is provided to cover the extensions employed in the report. This is done because there is seemingly no universally acknowledged extension, unlike there is a universally acknowledged Backus-Naur form, namely the one used in the ALGOL 60 Reference Manual[1].

A.1 What's in common with the original BNF

The following parts are in-common with the original Backus-Naur form:

Construct	Description
< ... >	A metalinguistic variable, aka. a nonterminal.
::=	Definition symbol
	Alternation symbol

Table A.1: Constructs in common with the original BNF.

In the original BNF, everything else represents itself, aka. a terminal. This is not preserved in this extension – all terminals are encapsulated into single quotes.

A.2 Constructs borrowed from regular expressions.

The use of single quotes around all terminals allows us to give characters such as (,),], *, +, and * special meaning, namely:

Construct	Meaning
(...)	Entity group
[...]	Character group
-	Character range
*	0-∞ repetition
+	1-∞ repetition
?	0-1 repetition

Table A.2: Constructs borrowed from regular expressions.

An entity group is a shorthand for an auxiliary nonterminal declaration. This means, for instance, that using the alternation symbol within it would mean an alternation of entity sequences within the entity group rather than the entire declaration that contains the entity group.

A character group may only contain single character terminals and an alternation of the terminals is implied from their mere sequence. It is identical to an auxiliary single character nonterminal declaration. A character range binary operator can be used to shorten a given character group, e.g. [`'a'` - `'z'`] implies the list of characters from `'a'` to `'z'` in the ASCII table. Moreover, a character range is the only operator allowed in a character group.

Applying the repetition operators to either the closing brace of an entity group or the closing bracket of a character group has the same effect as applying the repetition operator to their respective hypothetical auxiliary declarations.

A.3 Nonterminals as sets and conditional declarations

Another extension to the original BNF is the ability to use nonterminals as sets in declaration conditions. For example, if the two nonterminals, `<type-name>` and `<constructor-name>`, are both declared in terms of the `<literal>` nonterminal, but type names and constructor names should not intersect in a given program, then we can append the following condition to one or both declarations:

$$\text{s.t. } \langle \text{type-name} \rangle \cap \langle \text{constructor-name} \rangle \equiv \emptyset$$

Where the shorthand s.t. stands for “such that”. This implies that the nonterminals `<type-name>` and `<constructor-name>` represent the sets of character sequences that end up associated with the respective nonterminals for any given program, and can be used in conjunction with regular set notation.

Appendix B

The structured operational semantics used in this work