1 | Background

This chapter explores what *formal systems* are and what they are useful for. It looks at a number of related formal systems and their relation to computation. It outlines the context ontop of which the rest of this project is built.

Formal Systems

Formal systems are a set of rules for writing and manipulating formulae. Formulae are constructed from a set of characters called the *alphabet* by following a some formula-construction rules called the *grammar*. The only formulae considered *well-formed* in a system are those constructed according to the grammar of a system. Formal systems are used to model domains of knowledge to help better and more formally understand those domains.

Syntax and Grammars

The grammar of a formal system describes the system's syntax. Grammars are rules for constructing formulae that are well-formed. Formulae produced according to the grammar of a system are well-formed according to the syntax of that system.

We defined grammars using Backus-Naur Form or BNF:

$$M ::= t \mid f$$

This grammar describes that syntactically-valid constructs are either the letter t or the letter f. Grammars can be recursive which allows much more expressive construction rules:

This grammar describes formulae containing the any number of occurrences of the letters t or f separated by either an a or an o.

t	by (1)
$t\ o\ f$	by (2 & 4)
$t\ o\ f\ a\ t$	by (1 & 3)

$$
\begin{array}{rcll}
M, N & ::= & t & (1) \\
& | & f & (2) \\
& | & M\ a\ N & (3) \\
& | & M\ o\ N & (4)
\end{array}
$$

Figure 1.1: Grammar for producing the letters t or f connected by the letters a or o

Derivation Rules

Whereas a grammar describes the rules for producing well-formed formulae, the derivation rules describe rules for transforming formulae of a particular form into a new formula. Using the grammar from Figure 2, we add derivation rules:

$$
\begin{array}{lcl}
t\ a\ M & \rightarrow & M \\
M\ a\ t & \rightarrow & M \\
f\ a\ f & \rightarrow & f
\end{array}
$$

$$
\begin{array}{lcl}
M\ o\ t & \rightarrow & t \\
t\ o\ M & \rightarrow & t \\
f\ o\ f & \rightarrow & f
\end{array}
$$

These rules describe that if a formula matches the pattern on the left-hand side, where M represents a well-formed formula, it can be replaced by the formula on the right-hand side.

Domain Modelling

The syntax and derivation rules of a formal system are defined to model some domain. This isomorphism between the domain and the formal system means we can attempt to discover truths about the domain through studying the formal system.

For example, take the tf-system described above. Without understanding the domain, we are able to manipulate formulae of the system to create new formulae. The tf-system is isomorphic to Boolean algebra:

tf-system	Boolean algebra
t	1
t	0
a	$\wedge$
o	$\vee$

Using formal systems allows us to understand the domains they model from different perspectives and thereby learn novel truths about them.

λ-Calculus

The λ-calculus is a formal system described by Alonzo Church. The system models the execution of computer programs. As a Turing-complete system, it is capable of expressing the solutions to all problems that can be solved by a computer. It is of interest for this project because we can use it to model computer programs.

Syntax

λ-variables are represented by $x, y, z, \&c$. Variables denote an arbitrary value: they do not describe what the value is but that any two occurrences of the same variable represent the same value. The grammar for constructing well-formed λ-terms is:

Definition 1.0.1 (Grammar for untyped λ-calculus)
λ-variables are denoted by $x, y, \ldots$

$$M, N \quad ::= \quad x \;\mid\; \lambda x.M \;\mid\; M\ N$$

λ-abstractions are represented by $\lambda x.M$ where x is a parameter and M is the body of the abstraction. The same idea is expressed by more conventional notation as a mathematical function $f(x) = M$. The λ annotates the beginning of an abstraction and the . separates the parameter from the body of the abstraction. This grammar is recursive meaning the body of an abstraction is just another term constructed according to the grammar. Some examples of abstractions are:

$\lambda x.x$
$\lambda x.xy$
$\lambda x.(\lambda y.xy)$

Figure 1.2: Examples of valid λ-abstractions

Applications are represented by any two terms, constructed according to the grammar, placed alongside one another. Application gives highest precedence to the left-most terms. Bracketing can be introduced to enforce alternative application order for example xyz is implicitly read as $(xy)z$ but an be written as $x(yz)$ to describe that the application of yz should come first. Examples of applications are:

xy
xyz
$x(yz)$
$(\lambda x.x)y$

Figure 1.3: Examples of valid applications

Reduction Rules

First we will introduce the substitution notation $M[N/x]$. This denotes the term M with all occurrences of x replaced by N. The substitution notation is defined inductively as:

Definition 1.0.2 (Substitution notation for λ-terms)

$$\begin{array}{rcll} x[y/x] & \rightarrow & y & \\ z[y/x] & \rightarrow & z & (z \neq x) \\ (\lambda z.M)[y/x] & \rightarrow & \lambda z.(M[y/x]) & \\ (MN)[y/x] & \rightarrow & M[y/x]N[y/x] & \end{array}$$

β-reduction

The main derivation rule of the λ-calculus is β-reduction. If term M β-reduces to term N, we write $M \rightarrow_\beta N$ although the β subscript can be omitted if it is clear from context. β-reduction is defined for the application of two terms:

Definition 1.0.3 (β-reduction for λ-calculus)

$$(\lambda x.M)N \quad \rightarrow_\beta \quad M[N/x]$$

λ-variables and λ-abstactions are *values*: they do not reduce to other terms. If a formula is a value, the reduction terminates on that value. Only applications reduce to other terms. This means that an application is a reducible expression or a *redex*. Reducing a redex models the computing of a function.

α-reduction

The λ-calculus defines a reduction rule for renaming variables. Variable names are arbitrary and chosen just to denote identity: all occurrences of x are the same. This can become a problem in the following case:

$$(\lambda y.\lambda x.xy)(\lambda x.x)$$

After the application is reduced, we have the term:

$$\lambda x.x(\lambda x.x)$$

In this case, it is ambiguous which λ-abstraction the right-most x is bound by. When this term is applied to another, will the substitution occur to all occurrences of x? From the initial term, it is clear that this would be incorrect. The λ-calculus introduces α-reduction to solve this:

Definition 1.0.4 (α-REDUCTION FOR λ-CALCULUS)

$$\lambda x.M \to_\alpha \lambda y.M[y/x] \quad (y \notin fv(M))$$

This means we can rename the lead variable of an abstraction M on the conditions that:

1. All variables bound by that abstraction are renamed the same
2. The variable it is changed to is not currently in use in M

Logic and Types

Implicative Intuitionistic Logic

A formal system based on Gerhard Gentzen's natural deduction. Restricted only to $\to \mathcal{I}$ and $\to \mathcal{E}$. Following Brouwer, also leaves out law of excluded middle.

$$(\to \mathcal{I}) \quad \frac{A \to B \quad A}{B} \qquad (\to \mathcal{E}) \quad \frac{[A] \vdash B}{A \to B}$$

Type Assignment

Type assignment introduces additional grammar and restrictions on the reduction rules of a system. These extensions prevent logically inconsistent terms from being constructed. A type assignment has the form:

$$M : \alpha$$

which states that term M has the type α. Like variables, type variables are abstract: they do not describe anything more about a type than its identity. That is to say $x : A$ and $y : A$ have the same type but we cannot say any more about what that type is.

A type is either some uppercase Latin letter or it is two valid types connected by a $\rightarrow$. This is described by the following BNF grammar:

$$A, B ::= \varphi \mid A \rightarrow B$$

Typing rules have their own form that resembles Gentzen's sequent calculus:

$$\Gamma \vdash T$$

This describes that Γ is a set of typed λ-terms and T is a typed λ-term that is derivable from Γ.

Curry-Howard Isomophism

Curry-Howard isomorphism states that their is an isomorphism between the typed of a term and a logical proposition. The term itself is the proof of the proposition.

Haskell

Data Types

New data types can be introduced into Haskell in 3 distinct ways. First, using the `data` keyword:

```
data Animal a = Dog a
  | Cat
```

The `data` keyword begins the definition of a new data type. The word immediately following determines the type constructor for the new type. Following this is a type parameter for the type constructor. There can be any number of type parameters, including zero. The right-hand side of the `=` introduces a `|`-separated list of data constructors.

```
> let hector = Cat
> :t hector
hector :: Animal a

> let topaz = Dog "foo"
> :t topaz
topaz :: Animal String
```

The type parameter is constrained by the type of the value the data constructor was initialized with. In the example above, calling the `Dog` data

constructor with a string makes the type `Animal String` rather than the more general `Animal a`.

The second method for introducing new data types is the `newtype` keyword. The key difference between `data` and `newtype` is that `newtype` can only have one data constructor. Informally, this implies a kind of isomorphism:

```
newtype Foo a = Foo (a -> Integer)
```

The type constructor can take type parameters which will be constrained by the inhabitants of the data constructor. This data type expresses an isomorphism between `Foo a` and functions from `a` to `Integer`s.

Finally, we can introduce type aliases using the `type` keyword:

```
type Name = String
```

Again, we introduce a type constructor `Name` but this time we name another type, in this case `String`, as its inhabitant. This means that the type `Name` is a type alias for `String` and will share the same data constructors.

Type Level/Value Level

Haskell distinguishes between terms on the type level and terms on the value level. Type level terms are descriptions of the types of a value. They provide restrictions on the construction of invalid terms. For instance if we have a function of type `String -> Integer`, we cannot apply it to a term of type `Boolean`. The type-checker will throw an error before any value-level computation is initiated.

The value level is the level on which data is constructed and manipulated. The operation `1+1` occurs on the value level. The value level is where computation takes place and the type level is where static analysis of the program type takes place.

Type Classes

Haskell adds type classes to the type level. Types can have instances of type classes. The most similar concept from Object-Oriented programming is *interfaces*.

```
class Addable a where
  (add) :: a -> a -> a
```

Type classes are introduced using the `class` keyword. Beneath that are the function names and corresponding type-signatures of the functions that an instance of a class must implement.

For example, we can create instances of the `Addable` class:

```
data Number = One | Two | ThreeOrMore

instance Addable Number where
  add One One = Two
  add One Two = ThreeOrMore
  add Two One = ThreeOrMore
```

Continuations

Compound terms can be decomposed into two seperate parts: a dominant term and a context. The dominant term is the term currently being evaluated. The context is a term with a hole that will be filled with the value the dominant term reduces to.

Assume that $M \rightarrow_\beta M'$

(Compound term)	MN	
(Decompose)	M	$\Box N$
(Beta-reduce dominant term)	M'	$\Box N$
(Refill hole of context)	$M'N$	

Figure 1.4: Decomposing a term into a dominant term and a context

When M has β-reduced to a value – M' – then the hole of the context $\Box N$ is filled to form $M'N$. What the dominant term and context are for a given term depends on the reduction rules and strategy. The context is what remains to be reduced at given moment of reduction. Thus a context is also called a *continuation*.

Undelimited Continuations

For more complex terms, the waiting context will grow as the dominant term gets further decomposed:

$(MM')M''$	
(MM')	$\Box M''$
M	$(\Box M')M''$

Figure 1.5: Decomposing a term into multiple contexts

By amalgamating continuatinos into one big continuation we only have two components at point during the reduction: the current dominant term and the *current continuation*.

Assume we have some reduction rules defined for manipulating continuations. This model keeps continuations grouped together which means these hypothetical reduction rules could manipulate only this entire remaining continuation. For this reason, continuations that can only be manipulated in their entirety are **undelimited continuations**.

Delimited-Continuations

Instead, if we maintain a stack of continuations when decomposing complex terms, we can keep continuations separated:

$$\begin{array}{lll} (MM')M'' & & \\ (MM') & \Box M'' & \\ M & \Box M' & \Box M'' \end{array}$$

Figure 1.6: Decomposing a term into multiple contexts

Here, when a dominant term has been reduced, the reduct is returned to its corresponding continuation. This newly joined term then becomes the dominant redex. After this new dominant term has been reduced, it will be returned to the next waiting continuation, and so on. Throughout this process, we maintain each continuation separately.

Assume again that we have reduction rules defined for manipulating continuations. By keeping continuations separate, this model would allow use to use parts of the stack selectively for instance placing delimiters between portions of interest. The increased granularity of control means we can manipulate not just the entire remaining continuation but sections of it. Thus, continuations of this kind are called **delimited continuations**.

Continuation-Passing Style

By rewriting λ-terms, the continuations can be made explicit. All terms must be turned into λ-abstractions of some variable k where k is the continuation of a term. k is then called on the result of the term, triggering the continuation to take control. This style of writing λ-terms is called continuation-passing style or CPS.

Definition 1.0.5 (Translation of standard λ-terms into CPS)

$$\begin{array}{lcl} \underline{x} & = & \lambda k.kx \\ \underline{\lambda x.M} & = & \lambda k.k(\lambda x.\underline{M}) \\ \underline{MN} & = & \lambda k.M(\lambda m.m\underline{N}(\lambda n.mnk) \end{array}$$

The term that a CPS program terminates on will be of the form $\lambda k.kM$. In order to extract the value, a *final continuation* must be provided. Depending on the context, this could be an identity function $\lambda x.x$ or a display operation $\lambda x.\text{display } x$ to display the results of the program.

Figure 1.7: Extracting the final value from a terminated CPS program

$$\begin{array}{ll} & (\lambda k.kM)(\lambda x.x) \\ \rightarrow & (\lambda x.x)M \\ \rightarrow & M \end{array}$$

The translation of standard λ-terms into CPS similarly transforms the *types* of λ-terms. For example, a term $x : A$ becomes $\lambda k.kx : (A \rightarrow B) \rightarrow B$. This type represents a delayed computation: a computation that is waiting for a function to continue execution with. In order to resume the computation, the term must be applied to a continuation.

As an example, take the term M where

$$M = \lambda k.kx$$

To access the value x contained in M, we have to apply M to a continuation function $\lambda m.N$:

$$\begin{array}{l} (\lambda k.kx)(\lambda m.N) \\ (\lambda m.N)x \\ N[x/m] \end{array}$$

Within the body of N, m is bound to the value contained by M. So we can think about M as a suspended computation that, when applied to a continuation, applies the continuation to x. Looking at the type $(A \rightarrow B) \rightarrow B$ again, it is clear that A is the type of the term passed to the continuation of a CPS-term:

$$\begin{array}{l} \lambda k.kx : (A \rightarrow B) \rightarrow B \\ k : (A \rightarrow B) \\ kx : B \\ x : A \end{array}$$

Monads

If we have two suspended computations M and M' and we want to run M and then M', we have to apply M to a continuation to access its value and then do the same to M':

$$M(\lambda m.M'(\lambda m'.N))$$

This is a common operation so we define a utility operator `>>=` that binds the first suspended computation to a continuation which returns another suspended computation[1]:

$$\texttt{>>=} : ((A \to B) \to B) \to (A \to ((B \to C) \to C)) \to ((B \to C) \to C)$$

The type $(A \to B) \to B$ that represents a suspended computation returning a value of type A to its continuation we will call an A-computation or $Comp\ A$. We can rewrite the type signature of `>>=`:

$$\texttt{>>=} : Comp\ A \to (A \to Comp\ B) \to Comp\ B$$

We define another operator, *return*, that takes a value and returns a suspended computation that returns that value:

$$return : A \to Comp\ A$$

The type constructor $Comp\ A$, together with the two utility functions `>>=` and *return*, make up Haskell's Monad type class:

```
class Monad M where
  (>>=) :: M a -> (a -> M b) -> M b
  return :: a -> M a
```

The Monad type class generalizes CPS terms: they represent suspended computations that can be composed using `>>=`. Just like CPS terms, a Monad type `M a` tells us that we have a term that will pass values of type `a` to the continuation it is bound to using `>>=`.

$\lambda\mu$-Calculus

Syntax

Definition 1.0.6 (Grammar for $\lambda\mu$-calculus)

λ-variables are denoted by $x, y, \ldots$ and μ-variables are denoted by $\alpha, \beta, \ldots$

$$\begin{array}{lrcl} \text{(Unnamed term)} & M, N & ::= & x \mid \lambda x.M \mid M\ N \mid \mu\alpha.C \\ \text{(Named term)} & C & ::= & [\alpha]M \end{array}$$

Just as λ introduces a new λ-abstraction, $\lambda\mu$ introduces a new $\lambda\mu$-abstraction. The body of a $\lambda\mu$-abstraction must be a named term. A named term consists of a name of the form $[\alpha]$ followed by an unnamed term.

[1] `(>>=)` is pronounced 'bind'.

Reduction Rules

Definition 1.0.7 (Reduction rules for $\lambda\mu$-calculus)

$$\begin{array}{rcl} x & \rightarrow & x \\ \lambda x.M & \rightarrow & \lambda x.M \\ \mu\alpha.[\beta]M & \rightarrow & \mu\alpha.[\beta]M \\ (\lambda x.M)N & \rightarrow & M[N/x] \\ (\mu\alpha.[\beta]M)N & \rightarrow & (\mu\alpha.[\beta]M[[\gamma]M'N/[\alpha]M']) \end{array}$$

The terse reduction rule at the end simple states that the application of a $\lambda\mu$-abstraction $\mu\alpha.M$ to a term N applies all the sub-terms of M labelled $[\alpha]$ to N and relabels them with a fresh μ variable.

Computational Significance

The additional $\lambda\mu$ reduction rules model context manipulation. $\lambda\mu$-variables map to contexts. When an unnamed term is labelled with a $\lambda\mu$-variable, it is evaluated in that context. For instance the named term $[']a]M$ has the effect of evaluating M in the context pointed to by α.

To make this more concrete, consider the compound term $(\mu\alpha.[\beta]M)\ N$. First we decompose the term into a dominant term $(\mu\alpha.[\beta]M)$ and a context $\Box N$. Informally, we can imagine that the $\lambda\mu$-variable α now maps to this context $\{\alpha \Rightarrow \Box N\}$:

Example 1.0.8

Dominant	**Context**
$(\mu\alpha.[\beta]M)N$	
$\mu\alpha.[\beta]M$	$\Box N$

All subterms of M labelled α will now be evaluated in the context $\Box N$ and the context will destroyed. For example, let us replace M with $\mu \circ .[\alpha](\lambda s.fs)$:

Example 1.0.9

Dominant	**Context**	
$\mu\alpha.[\beta]\mu \circ .[\alpha](\lambda s.fs)$	$\Box N$	
$\mu\alpha.[\alpha](\lambda s.fs)$	$\Box N$	
$\mu\gamma.[\gamma](\lambda s.fs)N$		(γ fresh)

After applying the term $[\alpha](\lambda s.fs)$ to N, the context $\Box N$ is consumed and every occurrence of α is replaced with a fresh variable – in this case a γ – to clarify that the new $\lambda\mu$-abstraction points to a new context. This means that $\lambda\mu$-abstractions will pass all of the applicative contexts to the named subterms:

Example 1.0.10

Dominant	**Context**	
$(\mu\alpha.[\alpha](\lambda s.\lambda t.st))MN$		
$(\mu\alpha.[\alpha](\lambda s.\lambda t.st))M$	$\Box N$	
$\mu\alpha.[\alpha](\lambda s.\lambda t.st)$	$\Box M : \Box N$	
$\mu\gamma.[\gamma](\lambda s.\lambda t.st)M$	$\Box N$	(γ fresh)
$\mu\delta.[\delta](\lambda s.\lambda t.st)MN$	$\Box N$	(δ fresh)

Isomorphism & Computational Interpretation

λ^{try}-Calculus

Delimited-Continuation Calculus

Simon Peyton-Jones *et al.* extended the λ-calculus with additional operators in order create a framework for implementing delimited continuations [1]. This calculus will be referred to as the delimited-continuation calculus or DCC. Many calculi have been devised with control mechanisms. Like the $\lambda\mu$-calculus, these control mechanisms are all specific instances of delimited and undelimited continuations. DCC provides a set of operations that are capable of expressing many of these other common control mechanisms.

The grammar of DCC is an extension of the standard λ-calculus:

Syntax

Definition 1.0.11 (Grammar for DCC)

$$
\begin{array}{lrcl}
\text{(Variables)} & x, y, \ldots & & \\
\text{(Expressions)} & e & ::= & x \mid \lambda x.e \mid e\ e' \\
& & \mid & \mathit{newPrompt} \mid \mathit{pushPrompt}\ e\ e \\
& & \mid & \mathit{withSubCont}\ e\ e \mid \mathit{pushSubCont}\ e\ e
\end{array}
$$

Reduction Rules

The operational semantics can be understood through an abstract machine that transforms tuple of the form $\langle e,\ D,\ E\,q\rangle$:

Significance

The additional terms behave as follows:

- *newPrompt* returns a new and distinct prompt.

Definition 1.0.12 (Operational semantics for DCC)

$$
\begin{array}{lcll}
\langle e\ e', D, E, q\rangle & \Rightarrow & \langle e, D[\Box\ e'], E, q\rangle & \text{e non-value} \\
\langle v\ e, D, E, q\rangle & \Rightarrow & \langle e, D[v\ \Box], E, q\rangle & \text{e non-value} \\
\langle \mathit{pushPrompt}\ e\ e', D, E, q\rangle & \Rightarrow & \langle e, D[\mathit{pushPrompt}\ \Box\ e'], E, q\rangle & \text{e non-value} \\
\langle \mathit{withSubCont}\ e\ e', D, E, q\rangle & \Rightarrow & \langle e, D[\mathit{withSubCont}\ \Box\ e'], E, q\rangle & \text{e non-value} \\
\langle \mathit{withSubCont}\ p\ e, D, E, q\rangle & \Rightarrow & \langle e, D[\mathit{withSubCont}\ p\ \Box], E, q\rangle & \text{e non-value} \\
\langle \mathit{pushSubCont}\ e\ e', D, E, q\rangle & \Rightarrow & \langle e, D[\mathit{pushSubCont}\ \Box\ e'], E, q\rangle & \text{e non-value}
\end{array}
$$

$$
\begin{array}{lcl}
\langle (\lambda x.e)\ v, D, E, q\rangle & \Rightarrow & \langle e[v/x], D, E, q\rangle \\
\langle \mathit{newPrompt}, D, E, q\rangle & \Rightarrow & \langle q, D, E, q+1\rangle \\
\langle \mathit{pushPrompt}\ p\ e, D, E, q\rangle & \Rightarrow & \langle e, \Box, p : D : E, q\rangle \\
\langle \mathit{withSubCont}\ p\ v, D, E, q\rangle & \Rightarrow & \langle v(D : E\overset{p}{\uparrow}, \Box, E\overset{p}{\downarrow}, q\rangle \\
\langle \mathit{pushSubContE'}\ e, D, E, q\rangle & \Rightarrow & \langle e, \Box, E' + +(D : E), q\rangle
\end{array}
$$

$$
\begin{array}{lcl}
\langle v, D, E, q\rangle & \Rightarrow & \langle D[v], \Box, E, q\rangle \\
\langle v, \Box, p : E, q\rangle & \Rightarrow & \langle v, \Box, E, q\rangle \\
\langle v, \Box, D : E, q\rangle & \Rightarrow & \langle v, D, E, q\rangle
\end{array}
$$

- *pushPrompt*'s first argument is a prompt which is pushed onto the continuation stack before evaluating its second argument.
- *withSubCont* captures the subcontinuation from the most recent occurrence of the first argument (a prompt) on the excution stack to the current point of execution. Aborts this continuation and applies the second argument (a λ-abstraction) to the captured continuation.
- *pushSubCont* pushes the current continuation and then its first argument (a subcontinuation) onto the continuation stack before evaluating its second argument.