1. Trees; context free grammars.

1.1. Trees.

Definition 1.1. By a **tree** we mean an ordered triple

$$\mathcal{T} = (\mathcal{N}, \rho, p)$$

such that

(i) $\mathcal{N}$ is a finite set;
(ii) $\rho \in \mathcal{N}$;
(iii) $p : \mathcal{N} \sim \{\rho\} \to \mathcal{N}$;
(iv) if $n \in \mathbb{N}^+$ and $\nu \in \mathbf{dmn}\, p^n$ then $p^n(\nu) \neq \nu$.

One can relax the condition that $\mathcal{N}$ be finite but we will have no use for infinite trees.

Suppose $\mathcal{T}_i = (\mathcal{N}_i, \rho_i, p_i)$, $i = 1, 2$ are trees and $\iota : \mathcal{N}_1 \to \mathcal{N}_2$. We say ι is an **isomorphism from $\mathcal{T}_1$ to $\mathcal{T}_2$** if ι is univalent, $\mathbf{rng}\, \iota = \mathcal{N}_2$,

$$\iota(\rho_1) = \rho_2 \quad \text{and} \quad p_2(\iota(\nu)) = \iota(p_1(\nu)) \quad \text{for } \nu \in \mathcal{N}_1 \sim \{\rho_1\}.$$

1.2. Suppose $(\mathcal{N}, \rho, p)$ is a tree.

Proposition 1.1. Suppose $\nu \in \mathcal{N}$. Then there is $n \in \mathbb{N}$ such that $\nu \notin \mathbf{dmn}\, p^n$.

Proof. Suppose the Proposition were false. Let

$$f = \{(n, p^n(\nu)) : n \in \mathbb{N}\}$$

and note that $f : \mathbb{N} \to \mathcal{N}$. Since $\mathcal{N}$ is finite and $\mathbb{N}$ is infinite there are $m, n \in \mathbb{N}$ with $m < n$ and $f(m) = f(n)$. But then

$$p^{n-m}(p^m(\nu)) = p^n(\nu) = f(n) = f(m) = p^m(\nu)$$

which contradicts (iii). □

Suppose $\nu \in \mathcal{N}$. In view of the preceding Proposition we may set

$$d(\nu) = \max\{n \in \mathbb{N} : \nu \in \mathbf{dmn}\, p^n\} \in \mathbb{N}.$$

Note that

$$p^{d(\nu)}(\nu) = \rho.$$

Set

$$< \nu >= (p^{d(\nu)}(\nu), p^{d(\nu)-1}(\nu), \ldots, p(\nu), \nu) \in \mathcal{N}^{d(\nu)+1}.$$

Note that (iii) implies $< \nu >$ is univalent. Note that

$$< \nu >_0 = \rho \quad \text{and} \quad < \nu >_{d(\nu)} = \nu.$$

For each $\delta \in \mathbb{N}$ we let

$$\mathcal{N}^{(\delta)} = \{\nu \in \mathcal{N} : d(\nu) = \delta\}.$$

Note that $\mathcal{N}^{(0)} = \{\rho\}$.

The members of $\mathcal{N}$ are called **nodes**. ρ is called the **root node**. If $\nu \in \mathcal{N} \sim \{\rho\}$ we call $p(\nu)$ the **parent of** ν. If $\nu \in \mathcal{N}$ we let

$$\mathbf{c}(\nu) = p^{-1}[\{\nu\}] = \{\mu \in \mathcal{N} \sim \{\rho\} : p(\mu) = \nu\}$$

and call the members of this set the **children of** ν. A node which has children is called an **interior node**. A node which has no children is called a **leaf node**.

If $\mu, \nu \in \mathcal{N}$ we say μ is a **descendant of** ν if $\mu \neq \nu$ and ν is in the range of $< \mu >$ which amounts to saying that $\nu = p^n(\mu)$ for some $n \in \mathbb{N}^+$. If $\mu, \nu \in \mathcal{N}$ we say μ is an **ancestor of** ν if ν is a descendant of μ.

We let
$$\mathbf{i}(\mathcal{T}) = \{\nu \in \mathcal{N} : \nu \text{ is an interior node}\}$$
and we let
$$\mathbf{l}(\mathcal{T}) = \{\nu \in \mathcal{N} : \nu \text{ is a leaf node}\}.$$
If ν is a node we call $d(\nu)$ the **depth of** ν.

We $d(\mathcal{T}) = \max\{d(\nu) : \nu \in \mathcal{N}\}$ and call this natural number the **depth of** $\mathcal{T}$.

Definition 1.2. We say the tree $\mathcal{U} = (\mathcal{O}, \sigma, q)$ is a **subtree** of $\mathcal{T} = (\mathcal{N}, \rho, p)$ if $\sigma \in \mathcal{O} \subset \mathcal{N}$ and $q = p|(\mathcal{O} \sim \{\sigma\})$.

Given $\nu \in \mathcal{N}$ let $\mathcal{N}_\nu$ be the set whose members are ν and the descendants of ν, let $p_\nu = p|(\mathcal{N}_\nu \sim \{\nu\})$ and note that
$$\mathcal{T}_\nu = (\mathcal{N}_\nu, \nu, p_\nu)$$
is a subtree of $\mathcal{T}$ which we call the **subtree associated to the node** ν**.**

1.3. Ordered trees.

Definition 1.3. By an **ordered tree** we mean an ordered quadruple
$$\mathcal{O} = (\mathcal{N}, \rho, p, <)$$
such that $(\mathcal{N}, \rho, p)$ is a tree; $<$ is a linear ordering of $\mathcal{N}$ and

(i) $\rho < \nu$ whenever $\nu \in \mathcal{N} \sim \{\rho\}$;
(ii) $p(\mu) < p(\nu) \;\Rightarrow\; \mu < \nu$ whenever $\mu, \nu \in \mathcal{N} \sim \{\rho\}$.

Suppose $\mathcal{O}_i = (\mathcal{N}_i, \rho_i, p_i)$, $i = 1, 2$ are ordered trees and $\iota : \mathcal{N}_1 \to \mathcal{N}_2$. We say ι is an **isomorphism from** $\mathcal{O}_1$ **to** $\mathcal{O}_2$ if ι is an isomorphism from $(\mathcal{N}_1, \rho_1, p_1)$ to $(\mathcal{N}_2, \rho_2, p_2)$ and
$$\mu, \xi \in \mathcal{N}_1 \text{ and } \mu <_1 \xi \;\Rightarrow\; \iota(\mu) <_2 \iota(\xi).$$
Suppose $\mathcal{O} = (\mathcal{N}, \rho, p, <)$ is an ordered tree and $\mathcal{U} = (\mathcal{O}, \sigma, q)$ is a subtree of $\mathcal{T} = (\mathcal{N}, \rho, p)$. Let
$$\prec =$$
$\{(\mu, \xi) \in \mathcal{O} \times \mathcal{O} : \mu < \xi\}$ and note that $(\mathcal{O}, \sigma, q, \prec)$ is an ordered tree which we call the **ordered tree associated to** $\mathcal{U}$**.** In particular, if $\nu \in \mathcal{N}$, $\mathcal{O} = \mathcal{N}_\nu$, $\sigma = \nu$ and $q = q_\nu$ we call this ordered tree the **ordered tree associated to** ν**.**

1.4. Tree codes. We say a subset T of $(\mathbb{N})^*$ is a **tree code** if

(i) $\phi \in T$;
(ii) if $s \in T$, $j \in \mathbb{N}$ and $s|(j) \in T$ then $s \in T$;
(iii) if $s \in T$ then $\{j : s|(j) \in T\} = I(n)$ for some $n \in \mathbb{N}$.

Proposition 1.2. Suppose T is a tree code,
$$p = \{(s|(j), s) : s \in T,\; j \in \mathbb{N} \text{ and } s|(j) \in T\}$$
and $<$ is the intersection with $T \times T$ of the lexicographic ordering of $(\mathbb{N})^*$. Then $(T, \emptyset, p, <)$ is an ordered tree.

Definition 1.4. We call $(T, \emptyset, p, <)$ as in the preceding Proposition the **tree associated to the tree code** T**.**

1.5. Suppose $\mathcal{T} = (\mathcal{N}, \rho, p)$ is a tree and, for each $\nu \in \mathbf{i}(\nu)$,

$$<_\nu$$

is a linear ordering of $\mathbf{c}(\nu)$. We will show that there is one and only $<$ such that $(\mathcal{N}, \rho, p, <)$ is an ordered tree and

$$\mu <_\nu \xi \;\Leftrightarrow\; \mu < \nu$$

whenever $\nu \in \mathbf{i}(\mathcal{T})$ and $\mu, \xi \in \mathbf{c}(\nu)$.

Let $D = d(\mathcal{T})$. For each $\nu \in \mathbf{i}(\mathcal{T})$, let $n_\nu = |\mathbf{c}(\nu)|$ and let

$$c_\nu : I(n_\nu) \to \mathbf{c}(\nu)$$

be determined by the requirement that

$$i, j \in I(n_\nu) \text{ and } i < j \;\Rightarrow c_\nu(i) <_\nu c_\nu(j).$$

We define the functions

$$C_d : \mathcal{N}^{(d)} \to \mathcal{N}^d, \quad 0 \leq D,$$

by induction as follows. We let $C_0(\rho) = \emptyset$ and, whenever $0 \leq d < D$ we require that

$$C_{d+1}(\nu) = C_d(p(\nu))|(c_{p(\nu)}(\nu)) \quad \text{whenever } \nu \in \mathcal{N}^{(d+1)}.$$

We let

$$C = \bigcup_{d=0}^{D} C_d$$

The following statements hold:

(i) C is a univalent function with domain $\mathcal{N}$;
(ii) the range of C is a tree code;
(iii) if

$$<$$

equals the set of $(\mu, \nu) \in \mathcal{N} \times \mathcal{N}$ such that $C(\mu)$ precedes $C(\nu)$ in the lexicographic ordering on the $\mathbf{rng}\, C$ then $(\mathcal{N}, \rho, p, <)$ is an ordered tree;
(iv) whenever $\nu \in \mathbf{i}(\mathcal{T})$ and $\mu, \xi \in \mathbf{c}(\nu)$ we have

$$\mu <_\nu \xi \;\Leftrightarrow\; \mu < \nu.$$

Evidently, C is a isomorphism from $\mathcal{T}$ to the tree code associated to the range of C.

Proposition 1.3. Suppose $\mathcal{T} = (\mathcal{N}, \rho, p)$ is a tree and $(\mathcal{N}, \rho, p, <_i)$, $i = 1, 2$ are ordered trees such that

$$\mu <_1 \xi \;\Leftrightarrow\; \mu <_2 \xi \quad \text{whenever } \nu \in \mathcal{N} \text{ and } \mu, \xi \in \mathbf{c}(\nu).$$

Then $<_1 = <_2$

Proof. Straightforward exercise for the reader. □

That is, the ordering in an ordered tree is determined by the ordering it induces on the sets $\mathbf{c}(\nu)$ corresponding to interior nodes ν.

1.6. Growing trees. Suppose

$$\mathcal{T} = (\mathcal{N}, \rho, p)$$

is a tree and

$$\mathcal{O} \quad \text{and} \quad \mathcal{P}$$

satisfy the following conditions:

(i) $\mathcal{O}$ and $\mathcal{P}$ are functions with domain the leaf nodes of $\mathcal{T}$;
(ii) for each $\nu \in \mathcal{N}$, $(\mathcal{O}(\nu), \nu, \mathcal{P}(\nu))$ is a tree;
(iii) the family

$$\{\mathbf{i}(\mathcal{T})\} \cup \{\mathcal{O}(\nu) : \nu \in \mathbf{l}(\mathcal{T})\}$$

is disjointed.

Let

$$\mathcal{U} = \mathbf{i}(\mathcal{T}) \cup \left(\bigcup \{\mathcal{O}(\nu) : \nu \in \mathbf{l}(\mathcal{T})\}\right)$$

and let

$$q = p \cup \left(\bigcup \{\mathcal{P}(\nu) : \nu \in \mathbf{l}(\mathcal{T})\}\right).$$

We leave it as an exercise for the reader to verify that

$$(\mathcal{U}, \rho, q)$$

is a tree; that $\mathcal{N} \subset \mathcal{U}$; and that, for each $\nu \in \mathbf{l}(\mathcal{T})$, $(\mathcal{O}(\nu), \nu, \mathcal{P}(\nu))$ is the subtree associated to the node ν of $\mathcal{U}$.

Now let us suppose that

$$(\mathcal{N}, \rho, p, <)$$

is an ordered tree and that, for each $\nu \in \mathbf{l}(\mathcal{T})$, $<_\nu$ is such that

$$(\mathcal{O}(\nu), \nu, \mathcal{P}(\nu), <_\nu)$$

is an ordered tree. We leave it to the reader to verify that there is one and only one

$$\prec$$

such that

(i) $(\mathcal{U}, \rho, q, \prec)$ is an ordered tree;
(ii) if $\mu, \xi \in \mathcal{N}$ and $\mu < \xi$ then $\mu \prec \xi$;
(iii) if $\nu \in \mathbf{l}(\mathcal{T})$, $\mu, \xi \in \mathcal{O}(\nu)$ and $\mu <_\nu \xi$ then $\mu \prec \xi$.

2. Context free grammars.

Definition 2.1. By a **context free grammar** ordered triple

$$\mathcal{G} = (T, N, \mathbf{s}, \mathcal{P})$$

such that

(i) T is a set;
(ii) N is a set, $T \cap N = \emptyset$ and $\mathbf{s} \in N$.
(iii) $\mathcal{P} \subset N \times (T \cup N)^*$;

The members of T are called **tokens** or **terminal symbols**. The members of N are called **nonterminals** or **nonterminal symbols**. $\mathbf{s}$ is called the **start symbol**. The members of $\mathcal{P}$ are called **productions**. Instead of writing $(r, s) \in \mathcal{P}$ one often writes

$$r := \epsilon \quad \text{if } |s| = 0$$

and

$$r := s_0 \ s_1 \ \cdots \ s_{|s|-1} \quad \text{if } |s| > 0.$$

If $r \in N$, $n \in \mathbb{N}^+$ and $s_0, \ldots, s_{n-1} \in (T \cup N)^*$ one often writes

$$r := s_0 \mid s_1 \mid \cdots \mid s_{n-1}$$

instead of

$$(r, s_i) \in \mathcal{P}, \quad i \in I(n).$$

Obviously, if $(T, N, \mathbf{s}, \mathcal{P})$ is a context free grammar then so is $(T, N, \mathbf{t}, \mathcal{P})$ if $\mathbf{t} \in N$.

Definition 2.2. A **parse tree** $\mathcal{Q}$ **for the context free grammar** $\mathcal{G} = (T, N, \mathbf{s}, \mathcal{P})$ is an ordered quintuple

$$(\mathcal{N}, \rho, p, <, f)$$

such that

(i) $(\mathcal{N}, \rho, p, <)$ is an ordered tree;
(ii) $f : \mathcal{N} \to T \cup N$;
  (a) if $\nu \in \mathbf{i}(\mathcal{T})$ then $f(\nu) \in N$;
  (b) if $\nu \in \mathbf{l}(\mathcal{T})$ then $f(\nu) \in T$;
(iii) if $\nu \in \mathbf{i}(\mathcal{T})$ and ν has m children

$$\mu_0 < \mu_2 < \ldots < \mu_{m-1}$$

then

$$(f(\nu), (f(\mu_0), \ldots, f(\mu_{m-1}))) \in \mathcal{P}.$$

Notice that the notion of parse tree is independent of $\mathbf{s}$.

We define

$$< \mathcal{Q} > \in (T)^*$$

as follows. Let

$$\mathcal{L} = \{\nu \in \mathbf{l}(\mathcal{T}) : f(\nu) \neq \epsilon\}$$

and let

$$n = |\mathcal{L}|.$$

If $n = 0$ we let

$$< \mathcal{Q} > = \epsilon$$

and if $n > 0$ and

$$\nu_0 < \nu_1 < \cdots < \nu_{n-1}$$

are the members of $\mathcal{L}$ we let

$$< \mathcal{Q} > = (f(\nu_0))|(f(\nu_1))| \cdots |(f(\nu_{n-1})).$$

For each $\mathbf{t} \in N$ we let

$$\mathbf{L}(\mathcal{G}, \mathbf{t})$$

be the set of $< \mathcal{Q} >$ as above where $f(\rho) = \mathbf{t}$. We let

$$\mathbf{L}(\mathcal{G}) = \mathbf{L}(\mathcal{G}, \mathbf{s})$$

and we call this language on the alphabet T the **language generated by** $\mathcal{G}$.

Definition 2.3. We say the context free grammar $\mathcal{G}$ is **good** if $\mathcal{Q}_i = (\mathcal{N}_i, \rho_i, p_i, <_i, f_i)$, $i = 1, 2$, are parse trees such that $< \mathcal{Q}_1 >=< \mathcal{Q}_2 >$ then the ordered trees $(\mathcal{N}_i, \rho_i, p_i, <_i)$ are isomorphic; recall that this is the case if and only if they have the same tree codes. We say $\mathcal{G}$ is **bad** if it is not good.

2.1. A bad grammar. Let $\mathcal{B}$ be the context free grammar defined as follows. Let

$$T = \mathbb{N} \cup \{ -\,,\, +\,,\, * \}$$

and let

$$N = \{ \textbf{expr} \}.$$

Let **expr** be the start symbol. Let the productions be given by

$$\begin{aligned} \textbf{expr} &:= j \quad \text{for } j \in \mathbb{N} \\ \textbf{expr} &:= - \ \ \textbf{expr} \\ \textbf{expr} &:= \textbf{expr} \ \ + \ \textbf{expr} \\ \textbf{expr} &:= \textbf{expr} \ \ * \ \ \textbf{expr} \end{aligned}$$

Note that there are an infinite number of productions.

This grammar derives the string $27 + 5 * 298$ with two nonisomorphic parse trees so it is bad.

There are a number of ways to deal with this problem. One is to introduce parentheses which we now do. As we shall see, there are other ways to deal with this problem.