1 Introduction

A **grammar** consists of a collection of **subsitution rules**, also referred to as productions. Each rule appears as a line in the grammar, comprising a **variable** (a symbol) and a string separated by an arrow. The string consists of variables and other symbols called **terminals**.

The variable symbols are often represented by capital letters, and one variable is designated as the **start variable**, which usually occurs on the left-hand side of the topmost rule. The terminals are often represented by lowercase letters, numbers, or special symbols, and are analogous to the input alphabet.

All strings generated by a grammar constitute the language of the grammar – write $L(G_1)$ for the language of grammar G_1. A grammar is used to describe a language by generating each string of that language in the following manner:

- Write down the start variable
- Find a variable that is written down and a rule that starts with that variable. Repalce the written down variable with the right-hand side of that rule
- Repeat step 2 until no variables remain

The sequence of substitutions to obtain a string is called a **derivation**. This information can be represented pictorially with a **parse tree**.

2 Context-free grammars

A **context-free grammar** is a 4-tuple (V, Σ, R, S) where:

- V is a finite set called the variables,
- Σ is a finite set, disjoint from V, called the terminals,
- R is a finite set of rules, with each rule being variable and a string of variables and terminals, and
- $S \in V$ is the start variable.

Any language that can be generated by some context-free grammar is a **context-free language**.