CS 121 Section 4

Harvard University

September 3, 2013

1 Concept Review

1.1 Context Free Grammars

A context-free grammar G is a four-tuple, $G = (V, \Sigma, R, S)$, defined as follows:

- V is the set of variables
- Σ is the set of terminals, and so must be disjoint from V
- R is a finite set of rules, where each rule consists of a variable transforming into a string of variables and terminals
- S is the start symbol, and is an element of V

The idea is that the grammar consists of all strings over Σ^*, our terminal symbols, which we can get by starting with S and following the rules. The process of moving from S to a final string of terminals is known as a *derivation*.

1.2 Derivations

If x, y, and z are strings of variables and terminals and $A \rightarrow y$ is a rule of the grammar, then we can write $xAz \Rightarrow xyz$ and say xAz yields xyz in one step.

Extending that idea, if x_1 and x_n are strings of variables and terminals then we can say $x_1 \overset{*}{\Rightarrow} x_n$, or x_1 derives x_n, if we can get from x_1 to x_n by following 0 or more rules in succession. More formally, $x_1 \overset{*}{\Rightarrow} x_n$ if $x_1 = x_n$ or there is a sequence x_1, $x_2 \ldots x_n$ such that for all i, $x_i \Rightarrow x_{i+1}$. In practice, we often aren't very careful about distinguishing between 'derive' and 'yield', and it is ok to use them interchangeably.

The language of a grammar G is then defined as $L(G) = \{w \in \Sigma^* : S \overset{*}{\Rightarrow} w\}$

A derivation for a string w in a grammar G is any series of strings $S \Rightarrow x_1 \cdots \Rightarrow w$ that show how to get w from the rules of the grammar. A leftmost derivation for a string is a derivation where in each step, the leftmost variable in the string is substituted. A grammar is said to be ambiguous if there exists a string in the language of the grammar which has two different leftmost derivations. We often visualize derivations using parse trees.

2 Exercises

Exercise 2.1. *Show that the following languages are context-free:*

1. $L = \{a^i b^j c^k : i, j, k \in \mathbb{N}, \text{ and if } i = 1 \text{ then } j \geq k\}$ *over* $\Sigma = \{a, b, c\}$;
2. $L = \{w : w = w^R\}$;

Exercise 2.2. *Let* $G = (V, \Sigma, R, S)$ *be the following grammar.*

$$\begin{array}{rcl} S & \to & AS \| \varepsilon \\ A & \to & A1 \| 0A1 \| \varepsilon \\ \Sigma & = & \{0, 1\} \\ V & = & \{A, S\} \end{array}$$

1. *Show that* G *is ambiguous.*
2. *Give a new grammar that generates the same language as* G *but is unambiguous. Justify briefly why your grammar generates the same language and why it is unambiguous.*

Exercise 2.3. *Consider the following grammar:*

$$\begin{aligned} S &\to \langle SUBJECT\rangle\langle VERB\rangle\langle OBJECT\rangle\langle MODIFIER\rangle \\ \langle SUBJECT\rangle &\to \textit{The woman} \\ \langle VERB\rangle &\to \textit{hit} \\ \langle OBJECT\rangle &\to \textit{the man } \langle MODIFIER\rangle \\ \langle MODIFIER\rangle &\to \textit{with an umbrella} \mid \varepsilon \end{aligned}$$

Show that this grammar is ambiguous.

Exercise 2.4. *Show that every regular language has an unambiguous context-free grammar.*

Exercise 2.5. *Given an arbitrary context free grammar* G*, provide a general procedure to determine if* $L(G)$ *is empty.*