Accepted Manuscript

Characterization of context-free languages

M. Badano, D. Vaggione

PII: S0304-3975(17)30187-1
DOI: http://dx.doi.org/10.1016/j.tcs.2017.03.002
Reference: TCS 11102

To appear in: *Theoretical Computer Science*

Received date: 6 September 2016
Revised date: 12 January 2017
Accepted date: 2 March 2017

Please cite this article in press as: M. Badano, D. Vaggione, Characterization of context-free languages, *Theoret. Comput. Sci.* (2017), http://dx.doi.org/10.1016/j.tcs.2017.03.002

Characterization of context-free languages

M. Badano, D. Vaggione

Facultad de Matemática, Astronomía y Física. Universidad Nacional de Córdoba. Córdoba, Argentina.
e-mails: mbadano@famaf.unc.edu.ar, vaggione@famaf.unc.edu.ar

Abstract

In this note we present a simple condition upon which a formal grammar produces a context-free language.

Keywords: Context-free language.

Context-free grammars are one of the most investigated families of grammars in formal language theory. They provide a precise mechanism for describing the basic recursive structure of sentences in human language, and also have played a central role in compiler technology, as in the implementation of parsers, for example. In this note we give a characterization of context-free languages (i.e. languages generated by context free grammars), which is based on Greibach [1] normal form.

In order to state the result we revise the basic definitions. A *grammar* is a 4-tuple $G = (V, T, S, P)$, where V and T are finite sets of *variables* and *terminals*, respectively, $S \in V$ is the *start symbol* and P is a finite set of productions of the form $\alpha \to \beta$, with $\alpha, \beta \in (V \cup T)^*$ and α non empty. We assume that V and T are disjoint. The grammar G is *context-free* if all its productions are of the form $A \to \beta$ where $A \in V$ and $\beta \in (V \cup T)^*$. A language L is *context-free* if L can be generated by a context-free grammar. Let ε denote the empty string.

Theorem 1 *Let L be a language without ε. Then L is context-free if an only if L can be generated by a grammar for which every production is of the form $\alpha \to a\beta$, where α is a non empty string of variables, a is a terminal and β is a (possibly empty) string of variables.*

Before we prove the theorem, we need to state some notation and previous results. Let $G = (V, T, S, P)$ be a grammar. We write $\gamma_1 \underset{G}{\Rightarrow} \gamma_2$ when there exist $\lambda_1, \lambda_2 \in (V \cup T)^*$ and a production $\alpha \to \beta$ in P such that $\gamma_1 = \lambda_1 \alpha \lambda_2$ and $\gamma_2 = \lambda_1 \beta \lambda_2$. For $n \geq 0$, we write $\gamma_1 \overset{n}{\underset{G}{\Rightarrow}} \gamma_2$ when there exist $\alpha_1, \ldots, \alpha_{n+1}$ such that

$$\gamma_1 = \alpha_1,\ \gamma_2 = \alpha_{n+1} \text{ and } \alpha_i \underset{G}{\Rightarrow} \alpha_{i+1},\ i = 1, \ldots, n$$