

COMP2012/G52LAC Languages and Computation Lecture 8

Introduction to Context-free Grammars

Henrik Nilsson

University of Nottingham

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.1/8

Non-regular Languages (3)

But of course, “balanced parentheses” is a key feature of many important classes of languages; e.g.:

- Arithmetic expressions: (,)
- Matching keywords in programming languages: **begin**, **end**, **repeat**, **until**
- Markup languages; e.g. HTML: **<p>**, **</p>**, ****, ****

Q: Can such languages be described formally? How?

A: Through **Context-free Grammars** (CFG).

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.4/8

GCSESE (2)

Productions for GCSESE:

$$\begin{array}{ll}
 S \rightarrow NP VP & VP \rightarrow V Adv \\
 NP \rightarrow Adj NP & VP \rightarrow V \\
 NP \rightarrow N & V \rightarrow \text{walks} \\
 N \rightarrow \text{boy} & V \rightarrow \text{runs} \\
 N \rightarrow \text{girl} & Adv \rightarrow \text{slowly} \\
 Adj \rightarrow \text{little} & Adv \rightarrow \text{fast} \\
 Adj \rightarrow \text{big} &
 \end{array}$$

Note: The terminals constitute the **alphabet** of the language being defined.

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.7/8

Non-regular Languages (1)

We have established that the following language is not regular:

$$L = \{0^i 1^i \mid i \in \mathbb{N}\}$$

Others? What about B : the language of “balanced parentheses”? E.g.

$$\begin{array}{l}
 () \in B \\
 ((()())()) \in B \\
)(\notin B \\
 (() \notin B
 \end{array}$$

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.2/8

Context Free Grammars (CFG)

CFGs originated as an attempt to describe grammars for natural languages like English.

Key idea: Rules, called **productions**, that describe how symbols called **nonterminals** (or **variables** or **syntactic categories**) can be replaced by nonterminals and **terminals** until only terminals left.

nonterminal \rightarrow *terminals and nonterminals*

Let us consider the language **Grammatically Correct Sentences of Extremely Simplified English** (GCSESE)

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.5/8

Definition of CFG

A CFG $G = (N, T, P, S)$ where

- N is a finite set of **nonterminals** (or **variables** or **syntactic categories**)
- T is a finite set of **terminals**
- $N \cap T = \emptyset$ (disjoint)
- P is a finite set of **productions** of the form $A \rightarrow \alpha$ where $A \in N$ and $\alpha \in (N \cup T)^*$
- $S \in N$ is the **start symbol**

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.6/8

Non-regular Languages (2)

Is B regular?

NFA for up to three levels of parentheses:

((()))

How many states to recognize B ? Is B regular?

Use Pumping Lemma for regular languages to formally prove B not regular. **Exercise!**

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.3/8

GCSESE (1)

	Nonterminals	Terminals
S :	Sentence	boy
NP :	Noun Phrase	girl
VP :	Verb Phrase	little
N :	Noun	big
V :	Verb	walks
		runs
		slowly
		fast

COMP2012/G52LAC/Languages and Computation/Lecture 8 - p.6/8