

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.

$$\begin{aligned} ()() &\in B \\ (((()())()) &\in B \\)(&\notin B \\ (() &\notin B \end{aligned}$$

Non-regular Languages (2)

Is B regular?

No. Why?

Counting argument again: Any upper bound on the number of open parentheses that we would need to keep track of?

Use Pumping Lemma for regular languages to prove formally. **Exercise!**

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: $\langle p \rangle$, $\langle /p \rangle$, $\langle a \ href=“...” \rangle$, $\langle /a \rangle$

Q: Can such languages be described formally? How?

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

G52MAL

Machines and Their Languages

Lecture 9

Introduction to Context-free Grammars

Henrik Nilsson

University of Nottingham

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 → *terminals and nonterminals*

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

G52MALMachines and Their LanguagesLecture 9 – p.5/7

GCSESE (1)

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

G52MALMachines and Their LanguagesLecture 9 – p.6/7

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.

G52MALMachines and Their LanguagesLecture 9 – p.7/7