G52MAL
Machines and Their Languages
Lecture 9
Introduction to Context-free Grammars

Henrik Nilsson

University of Nottingham

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: `<p>`, `</p>`, `<a href="...">`, `</a>`

Q: Can such languages be described formally? How?

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

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)

GCSESE (1)

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

GCSESE (2)

Productions for GCSESE:

$$\begin{aligned} S &\rightarrow NP\ VP \\ NP &\rightarrow Adj\ NP \\ NP &\rightarrow N \\ N &\rightarrow \text{boy} \\ N &\rightarrow \text{girl} \\ Adj &\rightarrow \text{little} \\ Adj &\rightarrow \text{big} \end{aligned}$$

$$\begin{aligned} VP &\rightarrow V\ Adv \\ VP &\rightarrow V \\ V &\rightarrow \text{walks} \\ V &\rightarrow \text{runs} \\ Adv &\rightarrow \text{slowly} \\ Adv &\rightarrow \text{fast} \end{aligned}$$

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