COMPUTER SCIENCE TRIPOS Part IB 75%, Part II 50% – 2021 – Paper 7

4 Formal Models of Language (pjb48)

(a) The following is a pattern for some legal strings in a language:

$$[a \in A] \{0,1\} \ [b \in B] \{0,1\} \ [c \in C] \{1,n\} \ [d \in D] \{1,1\}$$

where A is a finite set of characters from the alphabet, Σ ; similarly for B, C, D. The sets A, B, C and D are disjoint. $\{x, y\}$ indicates that the previous bracket must match at least x times but no more than y times.

- (i) Specify a Deterministic Finite Automaton, M_1 , that can recognise these strings only. [4 marks]
- (*ii*) Design a Regular Grammar, G_1 , which generates $L(M_1)$. [4 marks]
- (*iii*) Describe a set of strings in a natural language that could be generated by G_1 given an appropriate Σ and its subsets A, B, C and D. [1 mark]
- (b) We can hypothesise that matches of the following pattern are always valid constructions in English:

[The Noun]
$$\{n,n\}$$
 [Verb] $\{n,n\}$

where Noun represents the coordinated members of a finite set; similarly for Verb.

(i) Now consider the following English sentence which matches the pattern when n = 1:

The vaccine worked

Provide example sentences that extend this sentence for the case when n = 2 and n = 3. [2 marks]

- (ii) Assuming that these constructions are part of the English language, would this mean that English is a Context-Free Language? Justify your answer.
 [3 marks]
- (*iii*) Design a grammar in Chomsky Normal Form, G_2 , which generates the finite matches of the pattern. [3 marks]
- (*iv*) Specify a Push Down Automaton, M_2 , that recognises $L(G_2)$. [3 marks]