## 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:

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

where $A$ is a finite set of characters from the alphabet, $\Sigma$; 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.
(ii) Design a Regular Grammar, $G_{1}$, which generates $L\left(M_{1}\right)$.
(iii) Describe a set of strings in a natural language that could be generated by $G_{1}$ given an appropriate $\Sigma$ and its subsets $A, B, C$ and $D$.
(b) We can hypothesise that matches of the following pattern are always valid constructions in English:
[The Noun] $\{\mathrm{n}, \mathrm{n}\}[$ Verb] $\{\mathrm{n}, \mathrm{n}\}$
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 $n=3$.
(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.
(iv) Specify a Push Down Automaton, $M_{2}$, that recognises $L\left(G_{2}\right)$.

