2000 Paper 11 Question 13

Natural Language Processing

The following context-free grammar and lexicon generates the examples below it with multiple derivations and therefore multiple associated interpretations.

(a) S → NP VP

(b) NP → Det N

(c) NP → NP PP

(d) N → N N

(e) VP → V NP

(f) VP → V NP PP

(g) VP → VP PP

(h) PP → P NP

N → car | park | tree | boy | toy | morning | ...
V → hit | ...
P → in | with | ...
Det → a | the | ...

a car hit the tree in the park
the boy hit the toy car park with a toy car in the morning

Describe how a probabilistic version of the context-free grammar (PCFG) can be created, defining the constraints which must hold for the resulting PCFG to be interpretable as a stochastic language model. [8 marks]

How accurate would the resulting PCFG be at assigning the semantically appropriate derivations the highest probability for the examples and other structurally similar sentences? [6 marks]

Define an improved probabilistic model for discriminating alternative derivations. What problems would arise in the implementation of this model? [6 marks]