

2000 Paper 7 Question 7

Natural Language Processing

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

(a) $S \rightarrow NP VP$

(b) $NP \rightarrow Det N$

(c) $NP \rightarrow NP PP$

(d) $N \rightarrow N N$

(e) $VP \rightarrow V NP$

(f) $VP \rightarrow V NP PP$

(g) $VP \rightarrow VP PP$

(h) $PP \rightarrow P NP$

$N \rightarrow \text{car} \mid \text{park} \mid \text{tree} \mid \text{boy} \mid \text{toy} \mid \text{morning} \mid \dots$

$V \rightarrow \text{hit} \mid \dots$

$P \rightarrow \text{in} \mid \text{with} \mid \dots$

$Det \rightarrow \text{a} \mid \text{the} \mid \dots$

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]