

2008 Paper 9 Question 9

Natural Language Processing

The following shows a simple context-free grammar (CFG):

S → NP VP	NP → Kim	NP → Sandy
NP → Det N	N → pudding	
N → N S	Det → the	
VP → V	V → likes	V → stinks V → believes
VP → V NP		
VP → V S		

- (a) Give a parse tree (or trees) to show the analysis this grammar assigns to the following:

the pudding Kim believes Sandy likes stinks

[3 marks]

- (b) What is meant by the term *overgeneration*? Why might overgeneration be a problem when using a grammar for parsing grammatical text? [3 marks]
- (c) The grammar shown above does not allow for different subcategorization properties of verbs. Assume that the following sentences should all be ungrammatical:

Kim likes
Kim stinks the pudding
the pudding Kim stinks stinks
the pudding Kim likes Sandy stinks

Assuming the simple CFG formalism, show how the grammar above could be modified to exclude such sentences while still admitting strings such as the one in part (a). [8 marks]

- (d) Outline how subcategorization may be treated in a grammar that uses a feature structure formalism. [6 marks]