

CE-3
August 2010 QE

CE-3 page 1 of 1

Part A (10%)

Consider the following context free grammar:

$$\begin{aligned} S &\rightarrow NP VP \\ NP &\rightarrow N_p \\ &\quad | D N_c \\ VP &\rightarrow V \{NP \{NP\}\} \end{aligned}$$

where braces denote optional constituents, the nonterminals S, NP, and VP denote sentences, noun phrases, and verb phrases respectively, and the terminals D, N_p , N_c , and V denote determiners, proper and common nouns, and verbs respectively. Determine all possible strings of five terminals that can be generated by this grammar.

Part B (20%)

You are given the following three sentences that are generated by the grammar in in part A:

foo bar baz quux frob
noz xyzyz plugh quux frob
quux grue baz foo bar

You have never heard these words before. You don't know what they mean. You don't even know their part of speech, i.e. which are common nouns, which are proper nouns, which are determiners, and which are verbs. You do know, however, that each word must have exactly one part of speech. Formulate the problem of determining the parts of speech as a constraint satisfaction problem (CSP). Precisely state what you take to be the variables, domains, and constraints of this CSP. **Hint:** If (in a different simpler but similar problem) the grammar could generate the terminal strings N V N and D N V and the word *foo* appeared in two different sentences, once as the first word and once as the second word, you could conclude that it must be an N. This is because, if you heard the two sentences *foo bar baz* and *quux foo frob* you would know that each one independently could only be either N V N or D N V, from the former sentence *foo* could only be N or D and from the latter sentence *foo* could only be V or N, and since each word must have exactly one part of speech, it could only be N. More generally, one way to solve this problem is to have each sentence in the training corpus correspond to a constraint.

Part C (20%)

State the definition of arc consistency in English or in mathematical notation as it would apply to constraints of arbitrary arity.

Part D (20%)

Describe an algorithm in English that would make a CSP arc consistent without removing potential solutions.

Part E (20%)

Apply the algorithm from part D to the CSP from part B. For which words does this uniquely determine their part of speech? For those words whose part of speech is not fully determined, what possible consistent part-of-speech assignments exist.

Part F (10%)

Find a single additional 5-word sentence that together with the sentences in part B uniquely determines the part of speech of all of the words in part B.

Write in Exam Book Only