COMPUTER SCIENCE TRIPOS Part II – 2015 – Paper 9

8 Natural Language Processing (AAC)

(*a*) The following text is from 'The Tale of Peter Rabbit' by Beatrix Potter (slightly modified):

> Peter never stopped running or looked behind him till he got home to the big fir-tree.
>
> He was so tired that he flopped down upon the nice soft sand on the floor of the rabbit-hole and shut his eyes. His mother was busy cooking; she wondered what he had done with his clothes. It was the second little jacket and pair of shoes that Peter had lost in a week!

Describe **six** features that are used in pronoun resolution algorithms using classifiers. For each feature, explain the range of values it can take, giving illustrative examples from the text above. [8 marks]

(*b*) Using the text from (*a*) to provide examples, explain what information you would ask a human annotator to give for pronoun resolution, and show how this relates to the training and test data needed for an algorithm using a classifier. [4 marks]

(*c*) The general problem of anaphora resolution includes finding antecedents for definite expressions, such as *the jacket* having *his clothes* as an antecedent. Outline how one might exploit semantic relations from WordNet as part of an extension of the classifier approach to deal with definite expressions. What limitations might WordNet have for this task? [8 marks]