9 Discourse Parsing and Coreference [Lecture 9]

Exercise 9.1
(a) What is the goal of discourse parsing? How is it different from syntactic parsing?
(b) Identify the discourse cue phrases, nuclei and satellites in the following sentences.
   (i) He went to bed because he was tired.
   (ii) He was tired, so he went to bed.
   (iii) As he was tired, he went to bed.
   (iv) He went to bed. He was tired.
(c) Do the above sentences all have the same rhetorical relations? Explain.

Exercise 9.2
(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 door 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.
(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.
(c) Give some examples of hard constraints and soft preferences in the anaphora resolution task and explain why they are important.
10 Text Summarization [Lecture 10]

Exercise 10.1

(a) Natural Language Generation (NLG) is a complex task that is often broken down into several sub-tasks. Describe the 7 components of an NLG system and explain what each component is responsible for.

(b) What is the difference between extractive summarization and abstractive summarization?

(c) Suppose you want to build a model for content selection for the purposes of extractive summarization. Would you use a supervised approach or an unsupervised approach and why? Describe the difference between each approach.