Natural Language Processing 2000

cannot distinguish the two derivations probabilistically.

C) A probabilistic model which took account of the words would do better, but this would greatly increase the number of parameters to be estimated and therefore require a good smoothing scheme to avoid overfitting the training data. The lecture notes outline a couple of possible models and briefly describe why Add1 smoothing isn't optimal.

2)

The following narrative contains several types of anaphoric relation between pronouns or definite noun phrases and discourse antecedents. Identify them. [4 marks]
Describe how they might be recovered using either statistical or knowledge-based techniques. [6 marks]
What are the relative advantages and disadvantages of each approach? [10 marks]

The house was beautiful – Kim and Sandy especially liked the windows. They wanted to buy it, so they rang the estate agent.

Solution Notes

A) distinction between anaphoric definite descriptions and those than link to 'understood' discourse referents arising from the house buying scenario (estate agent); 'bridging' inferences from lexical content (house, windows); anaphoric pronouns, chains of anaphoric links

B) lecture covered discourse structure based approach (Sidner's theory) and a statistical, salience weights approach (Kennedy and Boguraev), outline of either or related approaches with link to example narrative showing what would happen with each anaphor (see s9 in lecture notes).

C) House-windows requires a kb approach; the bridging inference also supports the inference that the second clause is an elaboration of the first (hence dash) so discourse subordinate. K and S-They, K and S is highest agreeing antecedent in discourse ie. there is no possible antecedent in first clause – either approach should get this, as windows should be ruled out / less prefered in dir. obj. slot. house-it agrees uniquely and house is still overall topic, but the kb approach is more likely to have access to this info. K and S-They-they subj-subj link via intermediate They – both approaches should get this.