Advanced AI Techniques (WS04)

Exercise sheet 3

Deadline: 11.11.04

Exercise 1 (4 points)
Given a corpus where we have observed 100 bigrams. The total number of possible bigrams (word combinations) is 1000. In the corpus, 9 bigrams were seen 10 times, 2 bigrams were seen 5 times, and the remaining 989 bigrams were unseen. Give Maximum Likelihood and Laplace probability estimates for the bigrams. What's the potential problem with the Maximum Likelihood estimate? Does the Laplace probability estimate solve the problem?

Exercise 2 (4 points)
Show that using ELE yields a probability function, in particular that

$$\sum_{w_1 ... w_n} P_{ELE}(w_1 ... w_n) = 1.$$

Exercise 3 (4 points)

a.) *Consider the training corpus below of 37368 words. We use the sentence* **"she was inferior to both sisters"** *as the test corpus. Compute the probability for this sentence first using unigrams and then using bigrams, assuming ELE estimates.*

w	$C(w)$	$w_1 w_2$	$C(w_1 w_2)$
person	*223*	*person she*	*2*
she	*6917*	*she was*	*843*
was	*9409*	*was inferior*	*0*
inferior	*33*	*inferior to*	*7*
to	*20042*	*to both*	*9*
both	*317*	*both sisters*	*2*
sisters	*427*		

b.) *Compare the two probabilities assigned to the sentence* **"she was inferior to both sisters"** *and explain why they differ.*