Ling 235 Homework #3

Due 2 February 2005

1. **Vowel combinatorics and probability.**
   (a) Assume that the only possible vowels in human languages are the 6 vowels /a/, /e/, /i/, /o/, /u/, /ə/. With no other restrictions, how many 4 vowel systems are possible?
   (b) Assuming that each of those (4 vowel) inventories is equiprobable, what are the chances that a language will have /a/ in its inventory?
   (c) What are the chances that a 4 vowel language will have /a/ and /u/ in its inventory?
   (d) Really, things aren't equiprobable (and some languages have more than 6 vowels!). According to the UCLA Phonological Segment Inventory Database sample, $P(\text{lg has /i/}) = 0.87$. Assuming that the other vowels are equiprobable, and that their occurrence is independent of /i/ and each other, reanswer questions (b) and (c) (for 4 vowel languages).

2. **Collocations and Heavy NP Shift.**

   In Wasow (2002), in the section on Collocations and HNPS, one finds the following data (graphed in (30) of the chapter):

	HNPS	not	
Transparent collocations	90	102	192
Non-collocations	59	329	388

   (a) Confirm the figure given in the paper for the chi-square test.
   (b) Calculate the odds ratio for heavy NP shift for transparent collocations versus non-collocations. That is, how many times larger are the odds of HNPS with a transparent collocation than with a non-collocation?

3. **Kaurna.**

   Warning: Made up data. The counts in this problem are fictitious. (The basic information on Kaurna is true though....)

   Sue Researcher is attempting to show that there is an association between aspect and voice in Kaurna. Kaurna is the Australian language that was spoken where Adelaide is now. It is basically extinct (efforts are being made to "revive" it). There is quite a lot of lexical information available in the form of an 1851 missionary's dictionary, but only a handful of sentences (by and large the dictionary just gives idioms or phrases as examples). Nevertheless, Sue has counted the available sentences and has this data:

	Voice	
Aspect	Active	Passive
NonPerfective	5	0
Perfective	1	2

(a) What are the maximum likelihood probability estimates for a nonperfective sentences to be active/passive? What are the maximum likelihood probability estimates for a perfective sentences to be active/passive?

(b) Do a chi-square test (of homogeneity) to determine whether there seems to be an association between aspect and voice in this data, and give the result.

(c) Is this test appropriate here? Why or why not?

(d) State in words what one might be testing doing a Fisher's Exact Test on the same data (there's more than one way to phrase it). (I mean not just "are voice and aspect associated?", but a more detailed description.)

(e) Do a Fisher's Exact Test on this data. Can one conclude that there is an association between aspect and voice in Kaurna?

(f) What if Sue discovered on a manuscript 3 new examples of perfective sentences, 2 of which were passive. So the data now looks like:

	Voice	
Aspect	Active	Passive
NonPerfective	5	0
Perfective	2	4

Does this new data change what one can conclude about the association between aspect and voice in Kaurna?

(g) What assumptions does one have to make about this data sample for any of the above to be valid?