# Homework 3

*Due: 24 March 2016*

For the next two questions, consider the sentence

> The council has approved wind farms.

in light of the probability tables on the next page.

### Problem 3.1

Using tables, give an appropriate probability for the sentence according to each of the following models:

a. unigram model (word probabilities out of context)
b. bigram model (previous word as context/1st-order Markov model)
c. POS-driven model (parts of speech as hidden states in HMM)

You will need to make certain assumptions to complete this; make sure you state them (and, of course, show your work).

### Problem 3.2

For each of the three models (unigram, bigram, POS-driven), give another six-word utterance that is A) not good English (the worse, the better) but which B) has a probability comparable to (or higher than) the given sentence, according to the given probabilities.

These are actual statistics (rounded and with just a few modifications), trained from the primary training section of the WSJ Penn Treebank,

These are the overall frequencies of each word:

| | |
|---|---|
| the | .05 |
| council | .000075 |
| has | .0035 |
| approved | .00015 |
| wind | .00002 |
| farms | .000015 |

These are a selection of relevant bigram probabilities $p$(word|prevword):

| | | | |
|---|---|---|---|
| the→approved | .00005 | approved→has | .00001 |
| the→council | .0004 | approved→the | .15 |
| the→wind | .00006 | approved→wind | .00000 |
| council→has | .00000 | wind→approved | .00000 |
| has→approved | .0006 | wind→council | .00000 |
| has→the | .015 | wind→farms | .00000 |

(Note that bigram pairs that did not occur in the training are reported as .00000 in the above table.)

These are a selection of part-of-speech probabilities for the given words $p$(word|POS):

| | | | |
|---|---|---|---|
| D→the | .6 | V→has | .03 |
| N→council | .0002 | V→approved | .001 |
| N→wind | .000025 | V→wind | .00006 |
| N→farms | .00004 | V→farms | .000008 |
| N→the | .00015 | | |

These are a selection of POS probabilities given previous POS:

| | | | |
|---|---|---|---|
| D→D | .0015 | N→V | .15 |
| D→N | .65 | V→D | .2 |
| D→V | .03 | V→N | .15 |
| N→D | .0075 | V→V | .15 |
| N→N | .3 | | |