

# Homework 3

Linguistics 384 (Detmar Meurers)

Due at beginning of class on Tuesday, Nov 2, 2004

1. Go to <http://www.lexmasterclass.com/exercises/regex/index.html> and do the exercises 1 to 4. Note that you can try out the regular expressions you type in by clicking on the Submit button.

For each of the four exercises send me a regular expression which matches all of the items in the first column (i.e., all characters in the first column are completely red after clicking submit) and none of those in the second column (i.e., none of the items in the second column are completely red).

(20 points for each of the four exercises, with 3–4 being bonus points)

2. Assume the partial non-positional bigram array for English given in the table below, where the first letter of the bigram is given by the vertical letters (i.e., down the side), and the second letter is given by the horizontal ones (i.e., across the top).

	a	b	c
a	1	1	1
b	1	1	0
c	1	0	1

- (a) According to this chart, which two bigrams (out of the nine sequences shown in the table) are not possible in English?  
(10 points)
- (b) Give an example word for each of the possible bigrams from this table, i.e., write down a total of 7 words.  
(15 points)
- (c) There are five misspellings in the following text, in the words shown in bold:

**Bobb** and his friend Abraham, or “**bae**” for short, were **acberbated bay** their other friend Arbuckle’s **ccat**.

- i. Which of these misspellings will be caught by the bigram array we have?  
(10 points)
  - ii. Which misspellings will not be identified?  
(10 points)
- (d) In terms of types of spelling errors, how would you characterize *acberbated*? (e.g. run-on? split? etc.)  
(10 points)
- (e) Go to <http://spell.imtranslator.com/try.html> and enter the text with the five misspellings. Step through the detected errors and proposed changes (you can select “Ignore Once” or “Change” to do so). For each word the system identifies as an error, answer the following (7 points per word):
  - i. Was the word actually a misspelling?
  - ii. How many and which potential corrections does it give?
  - iii. Is the correct spelling listed among the options?
3. Calculate the minimum edit distance from the string *halvs* to each of the following words
  - (a) halves
  - (b) calves
  - (c) halts
  - (d) helps

Use only insertions, deletions, and substitutions, with each as having a cost of 1. For each of the four pairs, write down

- the minimum edit distance, and
- the acyclic graph which you use to calculate the minimum edit distance, with annotated costs for each node.

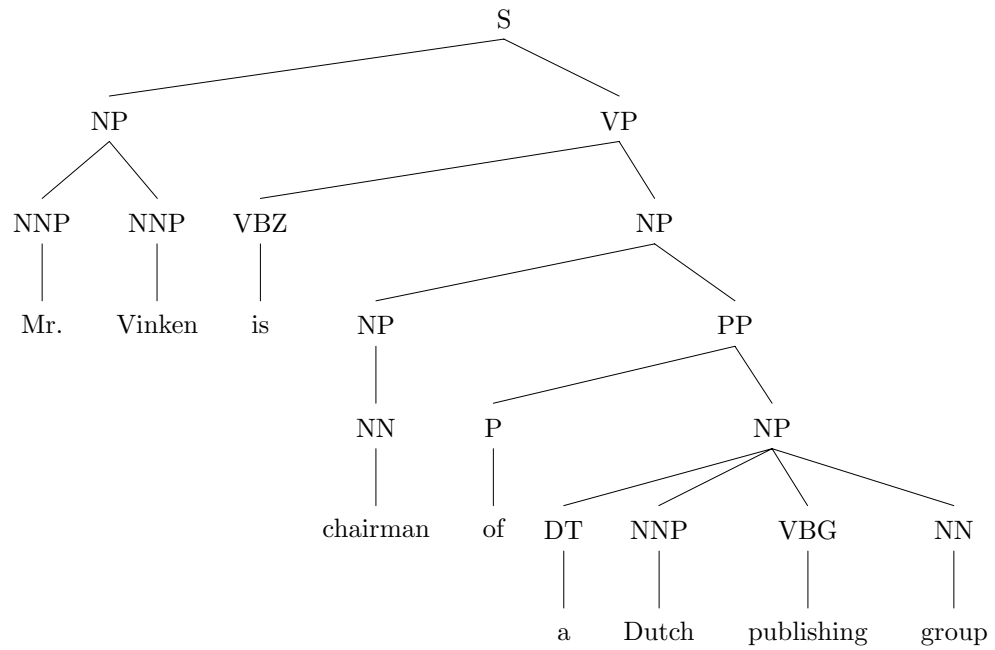
(10 points each)

4. We mentioned in class that *I saw the man with the telescope* is ambiguous. Given the phrase structure rules below, draw the two possible trees for the sentence *I saw the man with the telescope*.

NP → Det N	Det → <i>the</i>
NP → NP PP	N → <i>man</i>
NP → Pro	N → <i>telescope</i>
VP → V NP PP	P → <i>with</i>
VP → V NP	Pro → <i>I</i>
PP → P NP	V → <i>saw</i>

(20 points)

5. I took the following tree from the Wall Street Journal corpus (with some modifications). Given this tree, give me the phrase structure rules that are used to license it.



(20 points bonus)

6. You want to keep track of kids waiting for a drink of water at a water fountain. Is a stack, with its LIFO property, the best way to keep track of them?

(10 points)