

Homework 3

Scott Martin (Linguistics 384)

Due at beginning of class on Monday, February 18, 2008

1. 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).

		second letter		
		a	b	c
first letter	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 six misspellings in the following text, in the words shown in bold:

Bobb and his friend Abraham, or “**bae**” for short, were **acberbated bay there** 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) Go to <http://spell.imtranslator.com/try.html> and enter the text with the six 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 three questions:
- 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?
- (bonus question, 7 points extra per word)

2. 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.

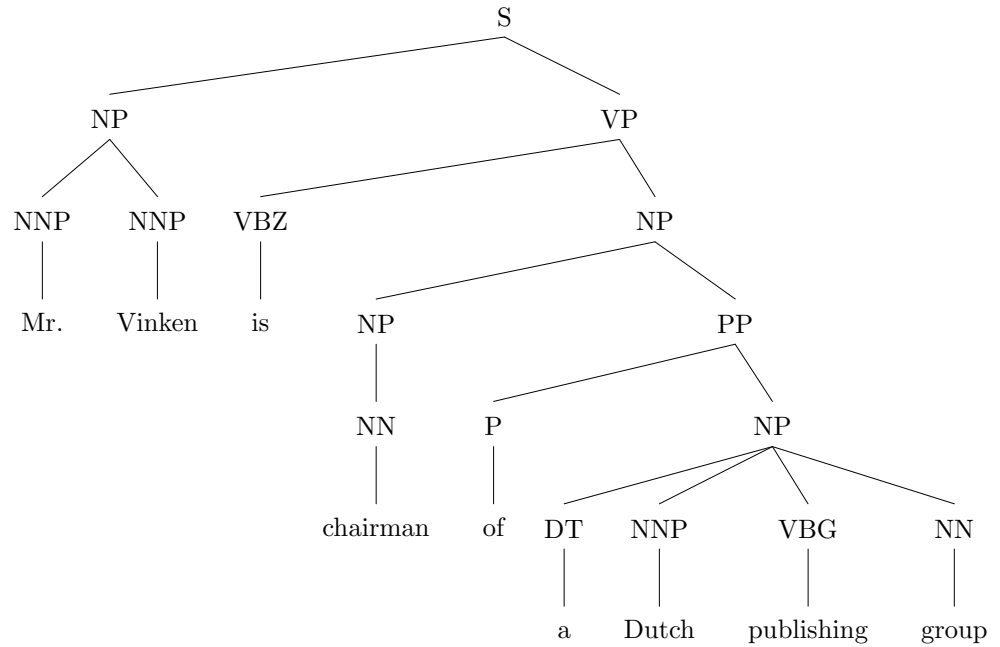
(30 points, plus 10 extra)

3. 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*.

$S \rightarrow NP VP$
 $NP \rightarrow Det N$ $Det \rightarrow the$
 $NP \rightarrow NP PP$ $N \rightarrow man$
 $NP \rightarrow Pro$ $N \rightarrow telescope$
 $VP \rightarrow V NP PP$ $P \rightarrow with$
 $VP \rightarrow V NP$ $Pro \rightarrow I$
 $PP \rightarrow P NP$ $V \rightarrow saw$

(20 points)

4. The following tree is a (slightly simplified) example from the Wall Street Journal corpus. Write down the phrase structure rules that are needed to license it.



(20 points)

5. You want to keep track of kids waiting to buy ice cream. Is a stack, with its LIFO property, the best way to keep track of them? (Motivate your answer with one or two sentences.)

(10 points extra)