

HOMWORK 1 ANSWER KEY

LING 388

1. REGULAR GRAMMARS

1.1. **a - 2 pts.** List the Regular Right-linear rules for the language a^*b^+c . This language contains strings such as

- bc
- abc
- bbbbbc
- aaaaaabc
- aabbbbbbbc

Remember that Right-linear rules can only have two forms: a Non-terminal on the left, and either a single terminal, or a single terminal followed by a single non-terminal on the left.

- $S \rightarrow a$
- $S \rightarrow aB$

Answer. The best answer consisted of 4 rules:

- $S \rightarrow aS$
- $S \rightarrow bA$
- $A \rightarrow bA$
- $A \rightarrow c$

Full points were given if your rules worked, even if they were different from these.

1.2. **b - 4 pts.** Write a Python parser that accepts strings in a^*b^+c , and rejects all others.

Answer. There was a lot of variation on this question. If your code worked, you got full points. Here is a solution with 3 rules.

```
def rule1(s):
    if s[0] == "b":
        rule2(s[1:])
    else:
        rule3(s)

def rule2(s):
    if s == "c":
        print "yes"
    else:
        rule1(s)

def rule3(s):
    if s[0] == "a":
```

```
    rule1(s[1:])  
else:  
    print "no"
```

2. UNIX AND WORD FREQUENCY

2.1. **4 pts.** Find a text of your choice on <http://www.gutenberg.org> and download the plain text version. Write a Unix command that finds the number of occurrences for each word in the text and sorts the results into a text file. Your text file should look something like the following:

```
25486 the  
14272 of  
10879 and  
10429 in  
7670 to  
7373 a  
6571 coffee
```

Use Excel or another program of your choice to plot a rank vs. frequency plot of the results. Rank is determined by how many occurrences a word has. For example the word 'the' above has rank 1 because it occurs most frequently in the text. The word 'of' has rank 2 and so on.

Answer. There are many ways to do this in unix. The code we went over in class was similar to the following:

```
cat 28500.txt | tr " " "\n" | sort | uniq -c | sort -nr | cat -n |  
awk '{print $1, $2, $3}' > savedoutput.csv
```

Full points were given for having the rank vs. frequency plot.

3. RESULTS

Here are the stats on the Homework:

Homework 1 Results

