

LINGUISTICS 384

LANGUAGE AND COMPUTERS

HOMEWORK 3

Name: _____

The last exercise helped us know what the unique (or different) words are in a text. Now we want to find out what the most frequently-occurring part-of-speech (or tag) is for each of these words. We start out by looking at how often every tag occurs for a given word. We are interested in the tag with the maximum number. Once we find out how to do this for one single word, we try to automate the process by making it a function. In HW4 we will map that function to all the unique words, to get each of their tags, and then combine the two together.

- Download the dataset from:
<http://ling.osu.edu/~jonsafari/ling384/wsj.tsv>
to your home computer. Change the working directory if you need to, and load the file into R, naming it `text` .
- Type the following: `table(text$tag[text$word == "have"])` . How many times does the word “have” get tagged as VBP? Answer: _____
What other tag occurs with this word (one time)? Answer: _____
- Make the above command the argument (ie. the input) to the following functions: `labels(which.max(...))` . These tell you the label of the most frequently-occurring tag. If you need to, first try just `which.max` , without the `labels` function. What is the most frequently occurring tag for the word “burlap”?
Answer: _____
- You can turn the above command into a function. Recall that we define a function as such:

```
name-of-function = function(args, ...) { do something with args }
```


For example, try defining, and calling the following function:

```
yell = function(x) { toupper(x) }  
yell("hi there")
```
- If you want to know what the most frequently-occurring tag for any given word (instead of just “burlap”), you don’t need to manually replace the word

in the long command that you typed above. Start defining the function as:

```
most.freq.tag = function(w) { ... }
```

And replace the part where you previously typed `"burlap"` with `w`. Now find out what is the most frequently-occurring tag for "curly", using the following line:

```
most.freq.tag("curly")
```

Answer: _____

- Save all your R code into a plaintext file, copying and pasting into Notepad/TextEdit/Gedit. Name the file `hw3-lastname.r`, replacing *lastname* with your last name. Don't use spaces in the name of the file. Submit the file in the Carmen drop-box for this class.
- Congratulations. You're still alive. And you rock.