

LINGUISTICS 384

LANGUAGE AND COMPUTERS

HOMEWORK 4

Name: _____

The last exercise helped us find out how to tag a single word. Now we will map that function to all the unique words, to get each of their tags, and then combine the two together into a dataframe called `model`.

- Download the dataset from:

<http://ling.osu.edu/~jonsafari/ling384/train.tsv>

to your home computer. Change the working directory if you need to, and load the file into R, naming it `train.set`. Reload your `most.freq.tag()` function from homework 3 into R.

- Map your `most.freq.tag()` function to the unique words in `train.set`, and save the output (which is a list of tags) to `unique.words.tags`.
- Create a dataframe of two columns: the unique words, and their respective tags. You can do this as:

```
data.frame(unique(train.set[,1]), unique.words.tags )
```

- Create a function called `train()` that simply includes the above two bullet points in the body of the function definition. This function only needs one argument: `train.set`.
- Try out your new `train()` function by typing `model = train(train.set)`. Then preview what the data in `model` looks like by using the `head()` function. Next, find out what is in the 336th line of `model`. Answer: _____
- Save your R code into a plaintext file, copying and pasting into Notepad/TextEdit/Gedit. Name the file `hw4-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 an elite hacker.