Ilaria Porru
Methods II HW2

HW2 – Write Up

1. In the first part of the code, I imported the library that I will need and used the shebang to make the script executable.
2. Lines 13-26 were given by Kyle. They define a function called `read_tags`. The command `yeld` chunks into different sentences.
3. Lines 29-32 create a function that split the text into `train`, `dev` and `test`.
4. Lines 35-37 creates empty list where the data will go
5. 40-41 define another function, `main`.
6. Lines 43-44 define a percentage that will then be applied to `train` `dev` and `test`.
7. Line 45-46 are for seeding
8. 50-52 are splitting the data into an 80% training set, 10% development set, and 10% test set.
9. 60-71 creates new file `args.train, args.dev and args.test`. I couldn't figure out how to format them, and I'm not very sure that this code is creating files.
10. 75-87 creates a parser and adds arguments. One of the optional arguments is seed. The last line is to parse the arguments and pass it to 'main'.

Problems with the code & tentative solutions:

At first, I tried to create the code from the Terminal. It was not successful, so I decided to download Anaconda and use Spyder.
I am not sure the shebang is able to run there, since it starts with # (and this is the symbol that we generally use only for comments).
After emailing Kyle I understood that I needed to move everything under the main function in order to make the argument parser work. I am still confused on the order of the code- I am not sure when it does matter and when doesn't.
I tried to make the code to be able to run with new data, but I couldn't get it to run even with `conll2000.tag`. When I test it, it says:

```
$    ./hw2-tentativo2.py conll2000.tag train.tag dev.tag test.tag
     File "<ipython-input-69-1da3caa494f1>", line 1
    $ ./hw2-tentativo2.py conll2000.tag train.tag dev.tag test.tag
    ^

SyntaxError: invalid syntax
```

After talking to Kyle, I added the 3 arguments as paths to which one can write the data.
In line 50 I originally wrote `test = alltags[0:test]`. After talking to Kyle, I got rid of the `alltags` and named it `corpus` (because this is what Python is analyzing) and I renamed `test`, since it was overwriting another variable.
As a template for seeding and randomization I used the file *numbers.py* that was used in practicum.