Ilaria Porru

Homework 2- Grapheme to Phoneme conversion
Writeup

This assignment was challenging. The first problem I encountered was preprocessing the data. It took me a couple of days to figure it out. I was trying to rename the columns with the following line of code:

```
train_df = train_df.rename(columns={anna: "Orthography", a n: a: "Phonetic"})
```

This did not work, as it substituted the row with those two words instead of creating a header. I solved the problem when I found that I could just refer to the columns as "0,1" (https://pandas.pydata.org/docs/reference/api/pandas.DataFrame.rename.html).
The second time, I confused the lists and wrote every new file in `ortho_list`. This gave me 97.98% WER. I canceled all the wrong documents and started over.

I had no trouble with the `fairseq-train` command, which ran at the first try. I used the `fairseq-train` documentation and googled the flags I could not find there (e.g., `--encoder-bidirectional`). I did not understand why we gave the arguments we gave to the flags, and it would be great if we could talk about it in class.

Finally, finding the word error rate was difficult. At first, I was trying to make two lists and compare them. I emailed Kyle and with the steps he wrote me I was able to create a target and a hypothesis. The very last part of the code (compare them and computing the WER). Kyle's demo (https://github.com/language-technology-GC/TextClassificationDemos) helped understanding I had to use `zip()`, which took a while to understand.