HW2 Challenges

About the evaluation part, my thinking was that I must compile a regular expression pattern something like:

regex =re.compile(r[ʌ\s|ɑ:\s|æ\s|e\s|ə\s|ɜ:ʳ\s|ɪ\s|i:\s|ɒ\s|ɔ:\s|ʊ\s|u:\s|aɪ\s|aʊ\s|eɪ\s|oʊ\s|ɔɪ\s|eəʳ\s|ɪəʳ\s|ʊəʳ\s|b\s|d\s|f\s|g\s|h\s|j\s|k\s|l\s|m\s|n\s|ŋ\s|p\s|r\s|s\s|ʃ\s|t\s|tʃ\s|θ\s|ð\s|v\s|w\s|z\s|ʒ\s|dʒ\s]+$)

or

regex = re.compile(r'([\u0250-\u02AF]\s)+') (I don't know this one, I searched this online)

Then I can use the pattern to match or find a string in the target lines. So I spent a lot of time trying to compile a perfect regular expression pattern. I have kept exploring it until I know we do not need one. Then I started to think maybe the "line.split('\t')[-1]" version is correct.

The second challenge is about the position where I should put "seen += 1". I put it after enumerating the lines and got WER 6.5. Then I divided "seen" by 2 withing thinking I have countered each "T-" line and "H-" line, so I should divide 2 to make sure that each "T-" line and "H-" line is a pair. Finally, I learned that I should increment `seen` inside the `elif line.startswith("H-")` bit.

Also, thank you very much for helping me fix the Anaconda issue, and for your prompt answers, detailed explanation. I do not think I can finish this assignment without all kinds of your help.