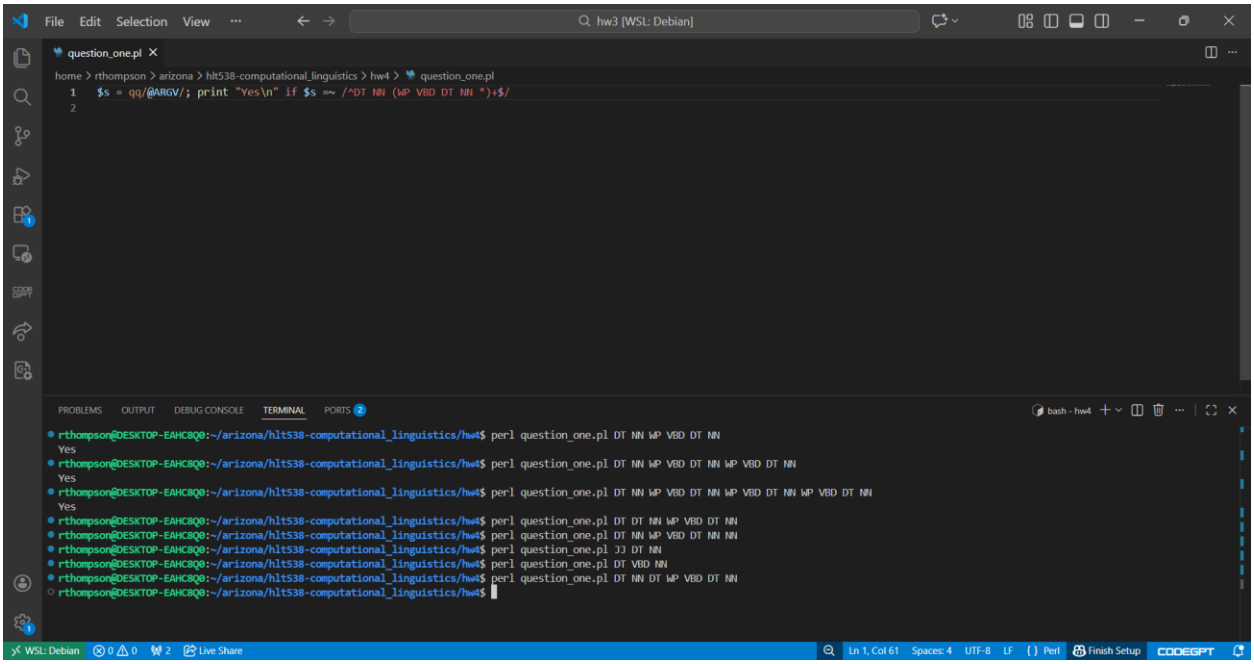
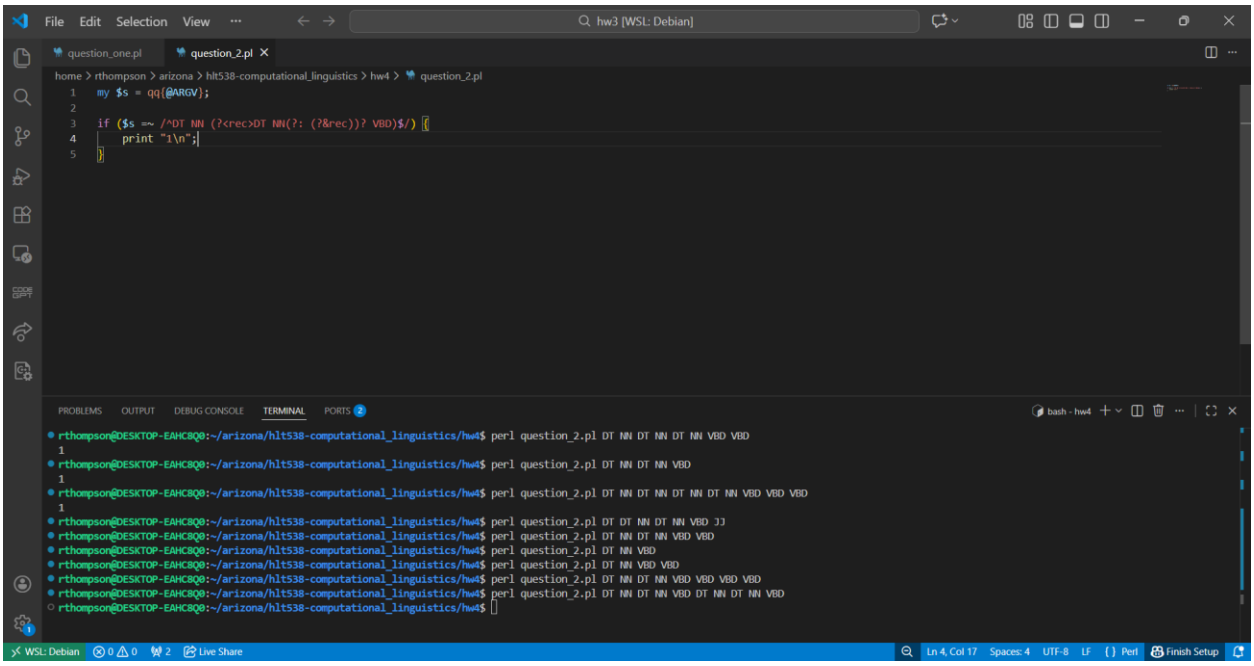
PART 1:

Question 1:

Here is a screenshot of my regular expression and it matching all of the test cases. It's not hard coded, it accepts all strings that fit the pattern and rejects all those that don't. My regex in non screenshot form, /^DT NN (WP VBD DT NN *)+$/



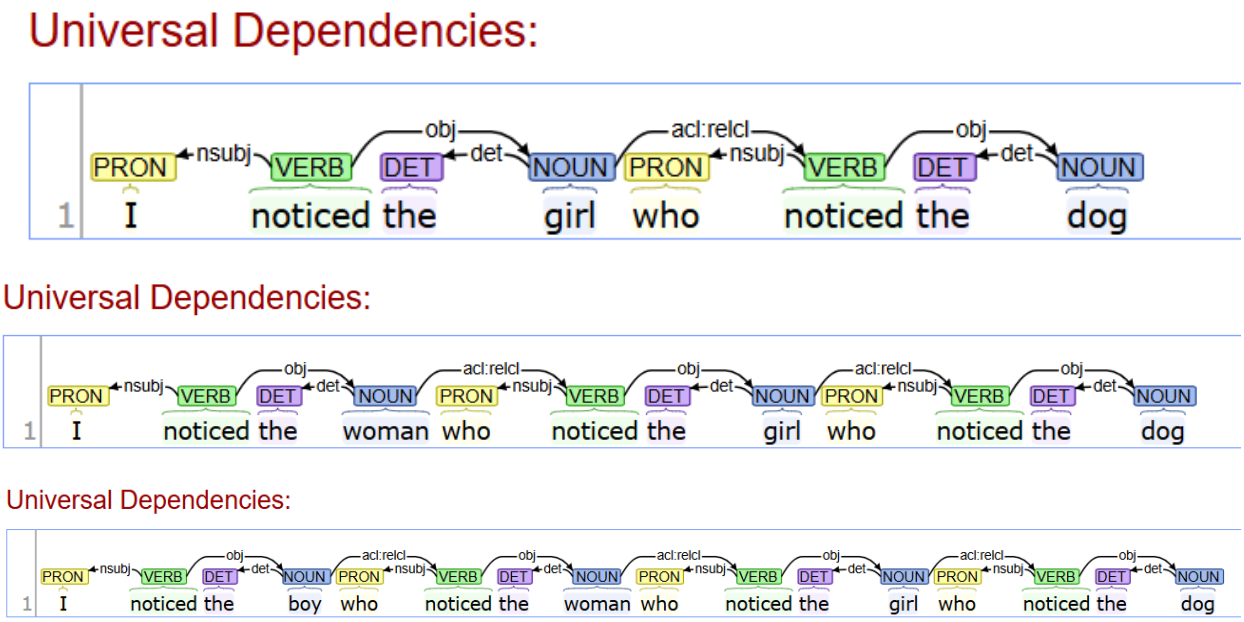
Question 2:



Here is my regex “/^DT NN (?<rec>DT NN(?: (?&rec))? VBD)$/” accepting and rejecting all of the corresponding test cases. By the Chomsky hierarchy, regular expressions shouldn’t be able to define this as it’s infixed, but Perl REGEX aren’t exactly pure regular expressions. I spend way too much time googling capture groups to make this work and it took up the bulk of time for this assignment.

Question 3: The parser I used was https://stanza.stanford.edu/

3a, 3b, and 3c:

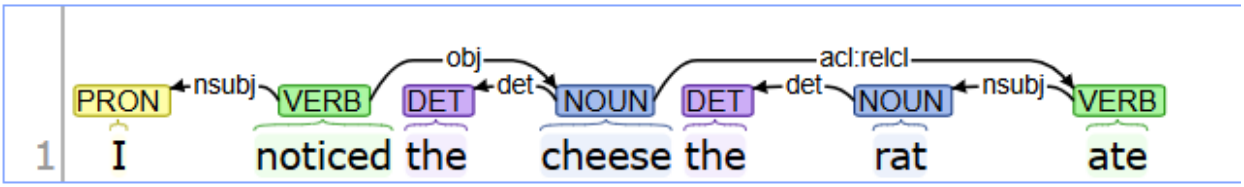


It does seem to parse all of them correctly. The chain of direct objects is unbroken all the way to the subject of the sentence. It’s a repeated pattern that doesn’t have to arc back to the original, which seems to make it easy for the parser.

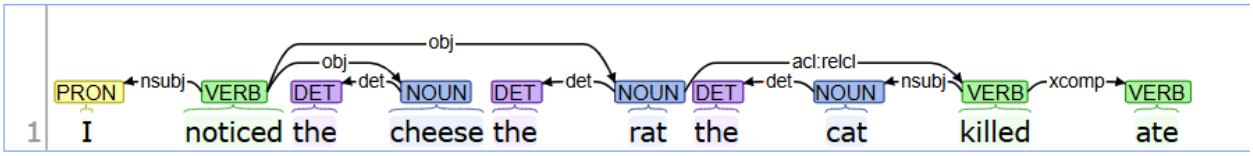
Question 4:

Here are 3a, 3b, and 3c for question 2:

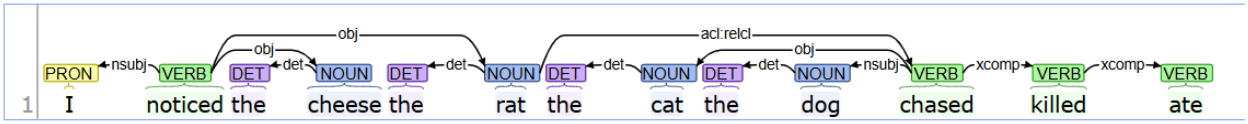
Universal Dependencies:



Universal Dependencies:



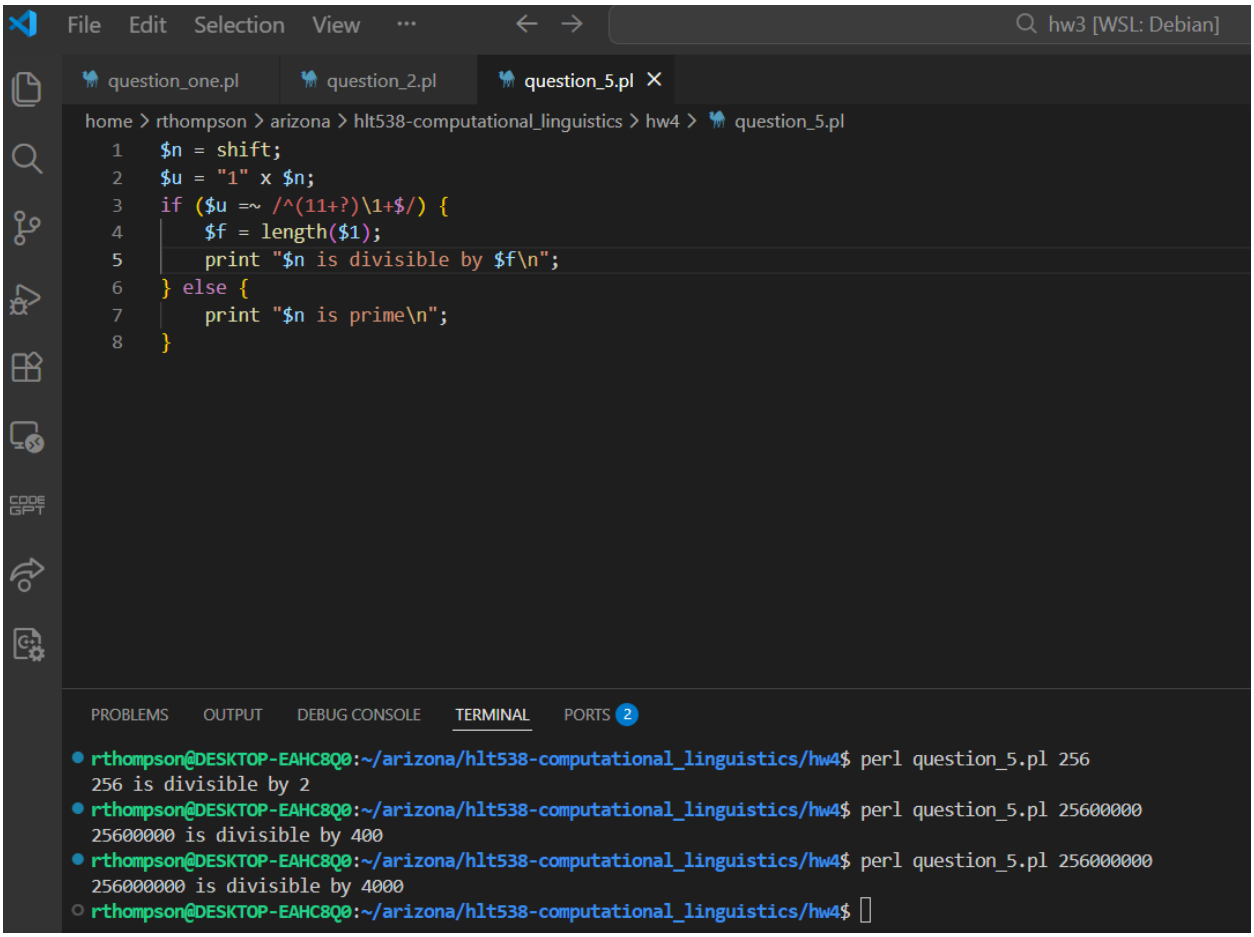
Universal Dependencies:



We appear to have broken it even starting at the second level. The verb ate is not connected to the noun cheese at all, even though that's what it should be the direct object of. It makes sense this is harder for the online parser to put together, because it's also harder for my internal parser to make sense of.

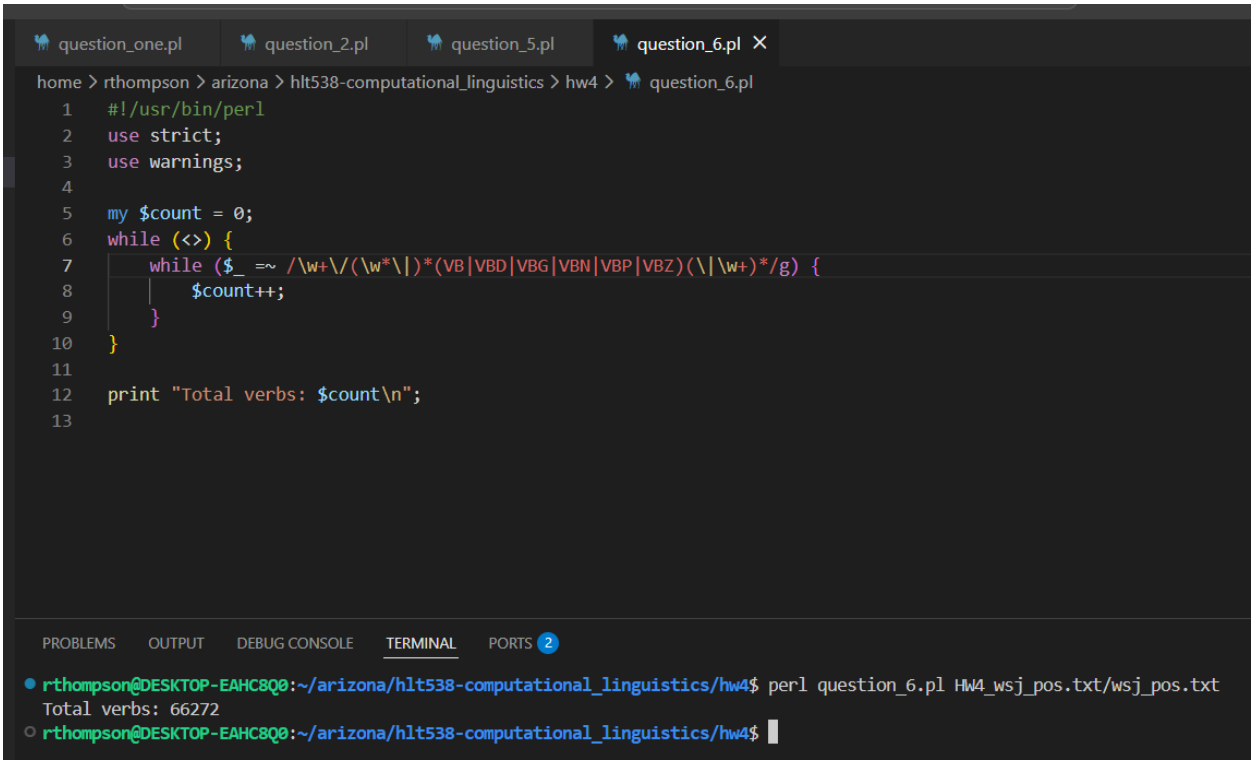
PART 2

Question 5:



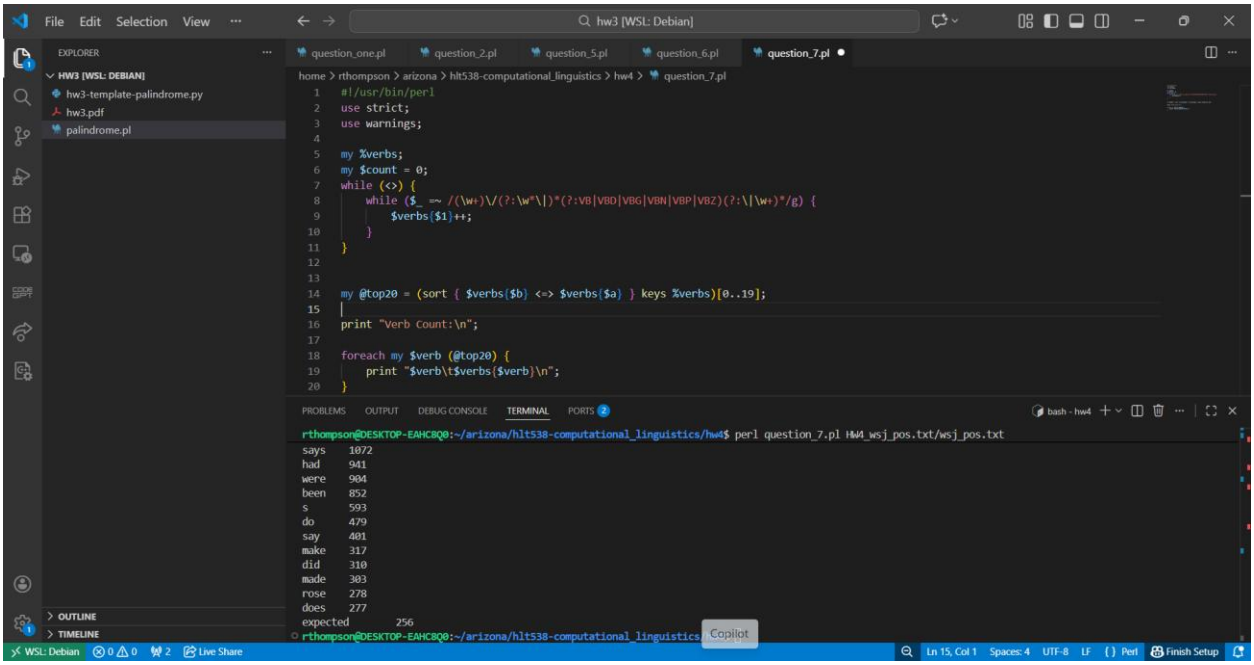
Here's the code. Using the length of the $1 scalar yields the factor pretty easily.

Question 6:



Here's a screenshot of my verb counter. Total verbs is 66272. Hope that's the right amount. Regex in text is /\w+\/(\w*\|)*(VB|VBD|VBG|VBN|VBP|VBZ)(\|\w+)*/

Question 7:



I did this one a little differently than in the video but this made more sense to me. I'm curious to see if the counts are right and if not, whether the problem is in the REGEX itself or in the code around it.