1.1Q: Which word does the model think this is most likely to have been? If the answer is not what we know is actually the most likely answer (“late”), is it a reasonable guess? Why or why not?
The model thinks “low” as the most likely intended word. Theoretically yes, it is a reasonable guess because two changes might have been made and we have got “larw”. In practice, I do not think it is a reasonable guess because if we intend to input “low” there would be very slight chance you accidentally input “a” and “r” in a row for “o” which you intended to input.

1.2 Q: Which words does the model assign a posterior probability higher than 0.05 to, and what probabilities does it assign to each of these? Is “late” in this list? Do all of the words in this list seem like reasonable guesses to you for what was intended? Why or why not? Q: If you were developing this autocorrect system, which part of the model would you improve next to improve the model’s performance? Why?

far	**0.0710872982476**
late	**0.230225256173**
are	**0.0828088786347**
low	**0.31789728776**
hard	**0.0558770898276**

Yes, “late” is in the list. Yes, all of them *seem* like reasonable guesses because making two changes on each of them will give us “larw”. If I were developing this autocorrect system, I would limit the characters being added, deleted, or replaced for each possible change by taking the probabilities of characters ranked at the top for each possible change. In that way, we would not have “low” as the most likely word intended to input and I have explained the reason in Q1.1.

2. Q: How many instances of the NP → DT rule are there in the corpus? Looking at the matches in detail, what grammatical structure is this rule generally being used for? How could we change the grammar used in the Penn Treebank such that it allowed this (grammatical) structure but didn’t allow ungrammatical structures like “The walked down the street.”
There are 190 instances of NP->DT rule are there in the corpus. (NP (DT word)) followed by a VP or (NP (DT word)) followed by a PP is this rule generally being used for grammatical structure. To avoid phrases like “The walked down the street” being generated, we have to change the grammar such that no all NP can be rewritten as DT. Those NPs followed by PPs can apply the rule NP->DT whereas NPs followed by VPs cannot apply the rule NP->DT. For example, you cannot replace “those” from “those marching for democratic freedoms” with “the”, but you can replace “each” from “Each representing three shares in the London market” with “the”.

3. Q: How many times do PPs headed by each of the four prepositions appear in each of the two types of phrase? Do there appear to be systematic differences between the two contexts in terms of which prepositions PPs tend to rewrite as, and if so, do you have any intuitions for what might explain them? How could we change our grammar using nonterminal annotation to allow the grammar to reflect this context sensitivity?
“in” heads a PP that is a daughter of NP: 152
“on” heads a PP that is a daughter of NP: 93
“for” heads a PP that is a daughter of NP: 277

“to” heads a PP that is a daughter of NP: 120

“in” heads a PP that is a daughter of VP: 66
“on” heads a PP that is a daughter of VP: 30
“for” heads a PP that is a daughter of VP: 56
“to” heads a PP that is a daughter of VP: 21
Yes, from my observations, prepositions tend to rewrite as noun phrases more frequently than they rewrite as verb phrases, and this makes sense as we more often use prepositional phrases embedded in noun phrases such as “the cat in the hat” than use them embedded in verb phrases such as “was in the house”. To reflect context sensitivity, we may want to assign weight for different environments: the environment that PP more frequently occur gets assigned for bigger weight.

4.

cell[0, 1] **PRP**: p = 1 **NP**: p = 0.3, PRP[0, 1]	cell[0, 2] **S**: p = 0.8, NP[0, 1], VP[1, 2]	cell[0, 3] NULL	cell[0, 4] **S**: p = 0.8, NP[0, 1], Vp[1, 4]	cell[0, 5] NULL	cell[0, 6] NULL	Cell[0, 7] **S**: p = 0.8, NP[0, 1], VP[1, 7]
	cell[1, 2] **N**: p = 0.4 **V**: p = 0.8 **VP**: p = 0.2, V[1, 2] **S**: p = 0.2, VP[1, 2]	cell[1, 3] NULL	cell[1, 4] **VP**: p = 0.5, V[1, 2], NP[2, 4] **S**: p = 0.2, VP[1, 4]	cell[1, 5] NULL	cell[1, 6] NULL	cell[1, 7] **VP**: p = 0.5, V[1, 2], NP[2, 7] **VP**: p = 0.3, VP[1, 4], PP[4, 7] **S**: p = 0.2, VP[1, 7]
		cell[2, 3] **D**: p = 1	cell[2, 4] **NP**: p = 0.5, D[2, 3], N[3, 4]	cell[2, 5] NULL	cell[2, 6] NULL	cell[2, 7] **NP**: p = 0.2, NP[2, 4], PP[4, 7]
			cell[3, 4] **N**: p = 0.5	cell[3, 5] NULL	cell[3, 6] NULL	cell[3, 7] NULL
				cell[4, 5] **P**: p = 1	cell[4, 6] NULL	cell[4, 7] **PP**: p = 1, P[4, 5], NP[5, 7]
					cell[5, 6] **D**: p = 1	Cell[5, 7] **NP**: p = 0.5, D[5, 6],

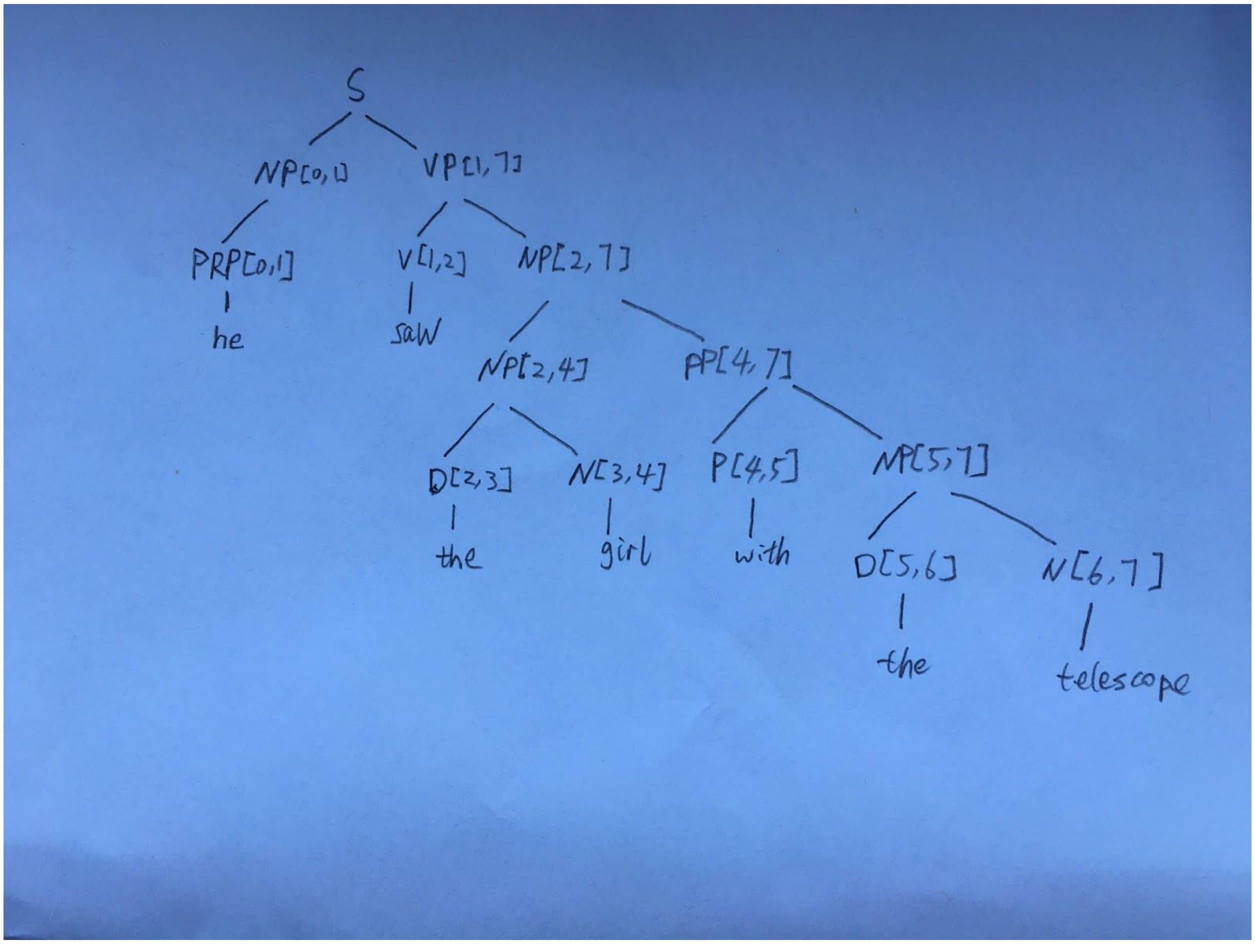
						N[6, 7]
						Cell[6, 7] **N**: p = 0.1 **V**: p = 0.2 **VP**: p = 0.2, V[6, 7] **S**: p = 0.2, VP[6, 7]

Q: What are the two meanings? Which do you think is more likely, and why?

1. **He saw the girl who had the telescope.**
2. **He used the telescope to see the girl.**

I cannot make a guess which one is more likely because they both are grammatically correct. My answer is we have to put them into a context to tell which one is more likely.

Q: Which parse does this grammar say is most likely? Write out this most likely parse as a tree. Which interpretation does this parse correspond to? What are the main reasons why this grammar prefers that parse? E.g., which probabilities of the grammar could you change to make the other parse be preferred?



The first interpretation is this parse correspond to, and the main reason that this grammar prefers the interpretation is V + NP (prob:0.5) is more dominant than VP + PP (prob:0.3) for constructing a VP. If we want the other interpretation to be preferred, then we want to make the prob of VP -> VP + PP larger than VP-> V + NP. Otherwise, you have VP -> V + NP being dominant then “with the telescope” has to be embedded into a noun phrase.