

Luan Tran
HW4

Q1.

There are 752 rules. The most frequent one is: PUNC -> . with counts = 346

Q2.

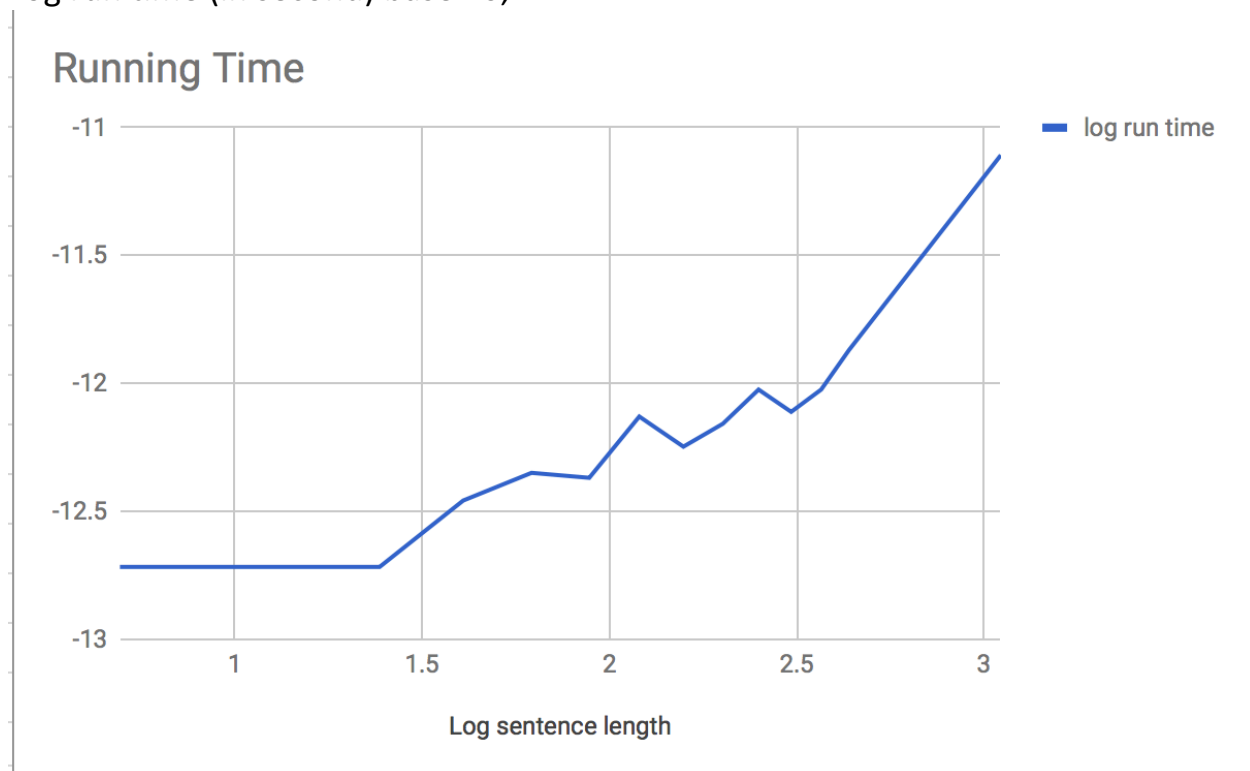
Output for the first line:

(TOP (S (NP (DT The) (NN flight)) (VP (MD should) (VP (VB be) (NP (NP* (CD eleven) (RB a.m)) (NN tomorrow)))))) (PUNC .))

Log probability = -17.7452979891

Q3.

Log run time (in second) base 10,



$k \sim -4$. It's close to 3, because the time complexity of Viterbi CKY is: $\theta(n^3 |G|)$, with n is sentence length, G is the size of the grammars.

Q4.

The output is: dev.parses.post 435 brackets

dev.trees 474 brackets

matching 400 brackets

precision 0.919540229885

recall 0.84388185654

F1 0.880088008801

Q6.

The output is:

test.parses.post 433 brackets

test.trees 471 brackets

matching 396 brackets

precision 0.91454965358

recall 0.84076433121

F1 0.87610619469