Homework 3

1a.

```
Number of unique rules = 752
Five most frequent rules:
PUNC -> . # 346
TO -> to # 241
PP -> IN NP_NNP # 239
IN -> from # 218
PP -> IN NP # 197
```

1b.

Grammar:
QP -> CC JJR # 0.071
NNP -> Jersey # 0.010
VP* -> VP* X_TO # 0.006
FRAG -> FRAG* NP # 0.069
NNP -> A # 0.025
NN -> Ground # 0.011
NP_NNP -> Columbus # 0.022
CD -> hundred # 0.057
VP -> VP* ADVP # 0.013
VBP -> do # 0.044
PP -> TO NP # 0.085
NN -> stop # 0.025
NP_DT -> Any # 0.067
RB -> p.m # 0.415
NNP -> Milwaukee # 0.002
NP_NNPS -> Wednesdays # 1.000
NP* -> JJ JJ # 0.002
DT -> this # 0.025
FRAG* -> FRAG* NP # 0.083
NN -> class # 0.050
FRAG -> ADVP_RB NP_NNP # 0.034
NNP -> County # 0.010
SQ_VP -> VP* NP # 0.020
NNP -> Los # 0.027
NNP -> Friday # 0.007
FRAG_NP -> JJS NNS # 0.010
NNS -> codes # 0.010
VP -> VBN PP # 0.019
WHNP -> WHNP PP # 0.033
VBZ -> serves # 0.034

NNS -> airlines # 0.014
FRAG* -> FRAG* PP # 0.167
S* -> S CC # 0.250
TOP -> X_S_VP PUNC # 0.002
TOP -> INTJ_UH PUNC # 0.004
NNP -> S # 0.007
WP -> What # 0.750
FRAG_NP_NN -> # 1.000
FRAG_NP -> NP* ADJP # 0.010
NNP -> La # 0.007
NN -> code # 0.007
NP_NN -> price # 0.020
NNP -> Delta # 0.002
VP* -> VB ADVP_RB # 0.006
NP* -> DT JJ # 0.038
VB -> visit # 0.012
VBD -> # 1.000
NN -> one # 0.007
VBP -> fly # 0.018
TOP -> FRAG_NP PUNC # 0.224
NN -> nonstop # 0.007
WHNP_WP -> What # 1.000
FRAG* -> NP NP # 0.167
SQ* -> NP_NP_EX NP # 0.029
WHNP -> WDT NNS # 0.670
IN -> with # 0.012
NX_NN -> aircraft # 1.000
FRAG_NP -> NP_NNS PP # 0.067
SQ -> MD SQ* # 0.053
NN -> airfare # 0.018
NP* -> NP_NNP PP # 0.002
S* -> ADVP_RB NP_PRP # 0.250
FRAG_WHNP -> WHNP PP # 0.030
SQ -> VBP NP_NP_EX # 0.013
IN -> from # 0.452
NP_NNP -> Saturday # 0.016
NNS -> stopovers # 0.010
NNS -> # 0.024
NNP -> Saint # 0.020
VBG -> leaving # 0.265
NN -> lunch # 0.007
FRAG_NP -> NP PP # 0.057
VP* -> VB PRT_RP # 0.006
NNS -> hours # 0.019
NP* -> JJS JJ # 0.009
VB -> fly # 0.030
NNS -> minutes # 0.024
NP_NP_EX -> there # 1.000
NP -> NP_CD PP # 0.001

VP -> VP* NP_NN # 0.006
VB -> # 0.048
PP_IN -> from # 1.000
SQ* -> NP_NP_EX SQ* # 0.250
VBP -> need # 0.140
VB -> list # 0.024
JJS -> # 0.125
SBARQ -> WHNP_WDT SQ # 0.049
TOP -> SQ PUNC # 0.045
SYM -> P # 0.167
VP -> VP* NP # 0.038
SQ_VP -> VBZ NP_NN # 0.020
TOP -> FRAG PUNC # 0.060
VB -> go # 0.012
NN -> Flight # 0.011
TOP -> FRAG_WHNP PUNC # 0.070
NP* -> NP SBAR # 0.002
MD -> would # 0.346
ADJP_JJ -> nonstop # 0.222
ADJP -> RBS JJ # 0.200
NN -> coach # 0.007
NP* -> NP* ADVP_RB # 0.004
S -> NP_NNP VP_VBZ # 0.016
FRAG_NP -> NP* ADVP_RB # 0.019
FRAG* -> NP_NN PP # 0.083
CD -> seven # 0.079
NP -> DT JJ # 0.007
NP* -> NP* X_JJ # 0.002
JJS -> Cheapest # 0.406
NP* -> NNP NNP # 0.063
NP -> QP NNS # 0.010
VP -> VBZ PP # 0.013
NN -> trip # 0.060
NP_NNP -> Atlanta # 0.013
S_VP -> VP* INTJ_UH # 0.014
NNP -> W # 0.015
VBP -> # 0.070
PP -> IN ADJP_JJ # 0.001
NP_PRP -> I # 0.324
NN -> Return # 0.007
NP* -> DT CD # 0.007
NP_NNS -> fares # 0.014
ADVP -> ADVP_RB PP # 0.286
PP* -> IN CC # 0.250
SQ -> X SQ* # 0.013
VB -> List # 0.018
S_VP -> VBP NP # 0.014
NP* -> NP* JJ # 0.004
VP -> VBP S_VP # 0.051

VP* -> VB NP_PRP # 0.529
S_VP -> VBZ NP # 0.007
NNS -> ones # 0.010
VP -> VBZ NP # 0.006
VB -> like # 0.102
INTJ_UH -> Thanks # 0.167
INTJ_UH -> Okay # 0.167
NP* -> CC NNP # 0.002
PP* -> NP IN # 0.250
SQ_VP -> VBZ ADJP_JJ # 0.060
VBG -> # 0.118
WHNP -> WHADJP NNS # 0.011
WHNP* -> NP PP # 0.018
NP -> NP NP # 0.003
NNP -> Ontario # 0.002
NP_NNS -> Airports # 0.029
NNS -> stops # 0.019
NP* -> NP VP # 0.002
NN -> transportation # 0.025
VP -> VBG NP_NNP # 0.006
NP_NNP -> Thursday # 0.009
NP -> NP* NN # 0.067
NP -> NP SBAR # 0.019
FRAG_NP -> NP* NP_NN # 0.019
PP -> PP* S_VP_VBG # 0.001
MD -> Can # 0.077
NP_PRP -> # 0.015
FRAG_NP -> NP_NN PP # 0.057
NP* -> QP NNS # 0.002
NP_NNS -> Fares # 0.029
TOP -> S PUNC # 0.115
NP_NNP -> Memphis # 0.031
VB -> be # 0.012
VP_VB -> stop # 0.167
IN -> around # 0.012
VBP -> leave # 0.096
NP_NN -> Flight # 0.078
IN -> than # 0.015
NP -> NP_NNP NP_NNP # 0.007
NP* -> NP_NN PP # 0.018
NP_NNP -> Milwaukee # 0.042
IN -> of # 0.081
CC -> or # 0.171
VBG -> arriving # 0.353
VP* -> VP* NP # 0.013
NNP -> Vegas # 0.059
NN -> tomorrow # 0.011
NP -> NP NNS # 0.001
ADVP_RB -> daily # 0.250

SBARQ -> SBARQ* SQ_VP # 0.010
QP* -> QP* CD # 0.214
S* -> NP_PRP ADVP_RB # 0.125
CD -> two # 0.014
VP_VB -> leave # 0.167
S -> INTJ_UH VP # 0.066
TOP -> X_SBARQ PUNC # 0.002
WHNP* -> WHNP PP # 0.544
NP_NNP -> Westchester # 0.002
ADVP_RB -> last # 0.062
NP* -> NP_NNS VP # 0.009
NP_NNP -> Phoenix # 0.042
ADVP_RB -> Now # 0.125
TOP -> FRAG_NP_NNP PUNC # 0.002
NP_NNS -> meals # 0.014
FRAG_NP -> NP* PP # 0.476
NNP -> Angeles # 0.027
NP* -> NNP CC # 0.004
IN -> With # 0.006
SQ_VP -> VBP NP # 0.080
CD -> fifty # 0.043
NNP -> F # 0.015
WHNP_WHNP -> WDT NNS # 0.900
NP* -> NP* SYM # 0.009
VP -> VB NP_NNP # 0.013
RB -> much # 0.094
NNP -> Cleveland # 0.005
SQ -> VBZ SQ* # 0.173
CD -> four # 0.021
S_VP -> VP* VP # 0.022
SQ_VP -> VP* PP # 0.340
PUNC -> ? # 0.262
NP -> NP* NNS # 0.049
FRAG_NP -> DT NNS # 0.010
FRAG_NP -> NP* CD # 0.010
IN -> as # 0.002
NNP -> C # 0.015
FRAG_NP -> NP* NNS # 0.029
NP -> NP* PP # 0.088
NP -> NNP NNP # 0.193
SQ* -> NP_DT NP # 0.015
VBP -> are # 0.237
DT -> all # 0.051
NNP -> City # 0.071
NP_NNS -> # 0.014
NP* -> CD CC # 0.004
CC -> # 0.024
NNP -> Southwest # 0.010
SQ_VP -> VBP NP_NNS # 0.020

NP -> NP* QP # 0.001
S_VP -> VBZ ADVP_RB # 0.007
VP_VBN -> served # 1.000
ADJP -> JJR PP # 0.400
VP -> VBG NP # 0.038
NP -> JJ NNP # 0.004
IN -> after # 0.033
NNP -> Tuesday # 0.002
NP* -> NP* NP # 0.002
JJR -> less # 0.750
NNP -> T # 0.010
NP_DT -> those # 0.267
CD -> eight # 0.029
FRAG_NP -> NP* SBAR # 0.010
NNP -> # 0.010
FRAG_WHNP -> WHNP* PP # 0.970
NP -> WDT NNS # 0.001
NP_NNP -> Denver # 0.033
NNS -> fares # 0.072
VBP -> make # 0.018
NP* -> CD RB # 0.009
WHNP_WHNP -> WHADJP NNS # 0.100
NP_NNP -> Miami # 0.033
FRAG -> NP NP # 0.069
PP -> IN NP_DT # 0.016
NNP -> Lake # 0.015
NP* -> NP* NN # 0.045
SQ* -> NP VP # 0.132
JJ -> next # 0.091
NP_NNP -> Friday # 0.020
VBZ -> has # 0.017
NP* -> CD NNS # 0.002
VB -> find # 0.012
VP_VB -> cost # 0.500
FRAG -> NP_NNP PP # 0.172
JJS -> least # 0.062
NP -> NP* NP_NN # 0.003
NP_NNP -> Monday # 0.011
NP -> JJS NN # 0.012
DT -> these # 0.056
CD -> twelve # 0.036
SQ* -> PP PP # 0.250
QP* -> RBR IN # 0.286
VP* -> VB ADVP # 0.006
FRAG_PP -> IN NP # 0.800
NP* -> NP CC # 0.034
NP_NN -> tomorrow # 0.118
NP -> NNP NN # 0.017
NNP -> Ohio # 0.005

NP_NNP -> Delta # 0.002
RBS -> least # 1.000
FRAG_NP -> NNP NN # 0.019
SBAR_S_VP -> TO VP_VB # 1.000
VBZ -> lives # 0.034
VB -> arrive # 0.024
NP_NNP -> July # 0.007
NNP -> New # 0.061
INTJ_UH -> # 0.167
PP -> TO INTJ_UH # 0.001
VB -> Explain # 0.018
SQ_VP -> VBZ NP # 0.020
NP -> NP* NP # 0.029
NP -> NP_NN PP # 0.001
CD -> five # 0.129
VP -> VB S # 0.013
NP_NNP -> Newark # 0.062
VB -> return # 0.012
PP -> IN NP_NNS # 0.004
VP* -> VB PP # 0.084
NP -> DT NN # 0.101
NP_PRP -> me # 0.603
NNP -> Monday # 0.012
ADJP_JJ -> Nonstop # 0.111
RB -> o'clock # 0.094
S_VP -> VB NP_PRP # 0.014
SQ -> VBZ NP # 0.293
NNP -> Paul # 0.005
VB -> travel # 0.018
QP* -> CD CD # 0.214
FRAG* -> WHNP PP # 0.167
VBG -> Departing # 0.059
FRAG -> X NP # 0.069
FRAG -> NP INTJ_UH # 0.034
ADVP_RB -> Only # 0.125
NNP -> Newark # 0.002
ADJP_JJ -> last # 0.333
VP -> VBG PP # 0.096
FRAG_ADJP_JJ -> Nonstop # 1.000
DT -> Any # 0.005
WHADVP_WRB -> # 1.000
IN -> at # 0.021
NN -> night # 0.007
FRAG -> WHNP PP # 0.034
NP* -> DT ADJP # 0.002
SQ -> VBP NP # 0.067
INTJ_UH -> please # 0.167
S_VP -> VB NP_NNS # 0.014
QP -> RB CD # 0.214

QP* -> QP* IN # 0.071
VBG -> Leaving # 0.118
NP -> NP* SBAR # 0.009
NN -> time # 0.007
VB -> Show # 0.518
NNP -> San # 0.074
FRAG_NP -> NN NNS # 0.010
VBZ -> arrives # 0.034
IN -> by # 0.006
FRAG_NP_NNP -> Tuesday # 1.000
SQ_VP -> VBZ ADJP_JJR # 0.040
NNP -> York # 0.051
QP* -> RB JJR # 0.071
NN -> ground # 0.018
SQ_VP -> VBP PP_IN # 0.020
NP -> CD NN # 0.014
S -> NP_PRP VP # 0.590
PP -> IN NP_NNP # 0.350
VB -> show # 0.018
SQ* -> NP SQ* # 0.029
FRAG_VP -> VP* PP # 0.750
NNP -> Francisco # 0.037
NN -> fare # 0.068
NNP -> D # 0.012
S -> NP_NN VP # 0.016
JJ -> nonstop # 0.152
NP -> NP_NNS VP # 0.003
NP -> NP_NNP NP # 0.013
NNP -> Jose # 0.017
NP_NN -> dinner # 0.275
RB -> apart # 0.075
IN -> for # 0.010
NNP -> U # 0.010
CD -> six # 0.079
ADVP_RB -> # 0.250
VP* -> VP* PP # 0.090
VP* -> VBZ PP # 0.013
NP* -> DT NN # 0.043
DT -> The # 0.020
CD -> oh # 0.014
VP -> VB NP_NN # 0.045
NP* -> NP NP # 0.004
NP -> QP NN # 0.003
NP -> NN NN # 0.013
NN -> morning # 0.060
JJS -> Shortest # 0.094
FRAG_NP -> NP* NN # 0.029
VBZ -> # 0.034
S_VP -> VP* NP # 0.525

CD -> eleven # 0.014
NP_NNP -> Houston # 0.035
SBARQ -> WHNP_WDT SQ_VP # 0.059
IN -> in # 0.102
PDT -> all # 1.000
NP_NNP -> Cleveland # 0.060
S_VP_VBG -> arriving # 1.000
NNP -> Kansas # 0.034
NP -> DT NX # 0.001
JJS -> cheapest # 0.094
NNP -> Cincinnati # 0.002
S_VP -> MD VP # 0.014
NP* -> NP_NN CC # 0.002
CD -> ten # 0.079
VP -> VBP FRAG # 0.006
QP* -> IN JJS # 0.143
VP -> VB PP # 0.045
PP -> PP* NP # 0.001
JJ -> # 0.197
WDT -> What # 0.916
TOP -> ADJP_JJ PUNC # 0.002
X -> WP IN # 0.667
SQ_VP -> VBP ADVP_RB # 0.020
MD -> should # 0.154
ADJP* -> RB RB # 1.000
CD -> one # 0.186
NNS -> flights # 0.746
NP* -> JJS NN # 0.004
NP* -> JJ NNP # 0.002
PP -> IN NP # 0.288
NNP -> Guardia # 0.007
PP -> IN NP_PRP # 0.001
NP* -> NP_NNP CC # 0.013
VP -> VBP NP # 0.108
QP -> QP* CD # 0.714
SQ -> INTJ_UH SQ* # 0.013
NP -> NP ADJP_JJ # 0.001
VP_VB -> # 0.167
SQ_VP -> VP* VP # 0.040
TO -> to # 1.000
VB -> leave # 0.024
NN -> flight # 0.267
NP_NNP -> Tampa # 0.031
TOP -> NP PUNC # 0.006
NNS -> dollars # 0.019
FRAG_NP -> JJS NN # 0.038
NN -> # 0.057
TOP -> FRAG_ADJP_JJ PUNC # 0.004
S_VP -> VBZ PP # 0.022

PUNC -> . # 0.738
NP* -> NP* VP # 0.002
NP_DT -> these # 0.533
DT -> # 0.010
SQ* -> NP PP # 0.015
NNP -> International # 0.005
NP* -> NN CD # 0.004
NP* -> NP* NNP # 0.034
SBARQ -> WHNP_WP SQ # 0.196
SQ* -> VBZ SQ* # 0.015
S* -> PP NP_PRP # 0.250
SQ* -> NP_PRP VP_VB # 0.015
FRAG -> X PP # 0.069
NN -> return # 0.007
NP -> JJ NNS # 0.007
NP_NNP -> Pittsburgh # 0.027
NP -> NP* SYM # 0.003
NP_NNP -> Washington # 0.018
S -> NP VP # 0.131
NN -> aircraft # 0.007
VP* -> VBP PP # 0.090
NP_CD -> one # 1.000
S_VP -> VBZ NP_NN # 0.007
NP_NN -> aircraft # 0.020
NN -> today # 0.004
NN -> day # 0.011
FRAG -> PP VP # 0.034
NN -> number # 0.018
NNP -> United # 0.015
NNP -> P # 0.002
NP_NNP -> Detroit # 0.013
NP* -> NP* PP # 0.123
NP_NNP -> Baltimore # 0.042
VP -> VB NP # 0.096
NP_NNP -> Cincinnati # 0.020
TOP -> FRAG_VP PUNC # 0.017
NP_NNP -> Orlando # 0.031
IN -> # 0.012
WHNP* -> WHNP* PP # 0.439
VP* -> VBG NP # 0.013
NNP -> Diego # 0.020
X_SBARQ -> WHNP SQ_VP # 1.000
VB -> Thank # 0.012
IN -> via # 0.006
NP* -> NN NNP # 0.013
NP* -> NP PP # 0.182
RB -> a.m # 0.151
NP_NNP -> Chicago # 0.020
NP_NNP -> Oakland # 0.016

NX -> NN NN # 0.500
NP -> NN NNS # 0.010
NNP -> Sunday # 0.005
NP -> JJ NN # 0.016
NP -> NNP NNS # 0.001
NNP -> Airlines # 0.042
NP_NNP -> Wednesday # 0.024
JJ -> First # 0.030
NN -> leg # 0.007
NP -> NP* NP_NNP # 0.006
X_TO -> to # 1.000
VP* -> VP CC # 0.045
NP* -> CD CD # 0.016
NP_NN -> Airline # 0.039
VP -> VB NP_NNS # 0.006
DT -> those # 0.005
FRAG -> X NP_NNP # 0.034
NP* -> NP_NNS PP # 0.099
NNS -> meals # 0.019
VBP -> arrive # 0.035
WRB -> How # 1.000
VP -> VP* PP # 0.102
IN -> between # 0.027
S_VP -> VBP PP # 0.014
VP_VBG -> # 1.000
NNP -> Saturday # 0.002
NN -> noon # 0.007
NP_NNS -> Flights # 0.594
NP -> NP_DT PP # 0.003
NP_NNP -> Tuesday # 0.009
CD -> twenty # 0.014
NX* -> NX_NN CC # 1.000
WHNP_WDT -> that # 0.429
NP_NNP -> Boston # 0.013
VBZ -> Does # 0.121
SQ -> VBP SQ* # 0.347
IN -> From # 0.010
CC -> and # 0.805
SYM -> D # 0.167
NN -> reservation # 0.007
NNP -> Las # 0.059
NP_NNP -> Sunday # 0.018
SQ_VP -> VBZ ADVP_RBS # 0.020
VBN -> # 0.333
SQ_VP -> VBP PP # 0.140
IN -> before # 0.027
NN -> business # 0.007
SQ_VP -> VBP NP_NNP # 0.020
ADVP -> RB PP # 0.143

VBG -> departing # 0.088
CD -> # 0.050
JJ -> Last # 0.030
VBP -> depart # 0.061
S_VP -> VBP NP_NN # 0.014
ADJP -> ADJP* PP # 0.200
DT -> the # 0.626
NX -> NX* NX # 0.500
PP -> TO NP_NNP # 0.233
NNP -> Long # 0.007
ADVP_RB -> here # 0.125
VP_VBZ -> has # 1.000
NP -> NP SBAR_S_VP # 0.001
NP_NNP -> American # 0.004
NP* -> JJ NN # 0.018
NN -> dinner # 0.021
NP_NNP -> Nashville # 0.038
WHADJP -> WRB JJ # 1.000
NN -> way # 0.028
SQ* -> NP VP_VB # 0.059
NN -> price # 0.039
VBP -> serve # 0.044
SQ_VP -> VBP NP_NN # 0.080
VP -> VBP PP # 0.032
WHNP_WDT -> Which # 0.571
SBARQ -> WHNP SQ_VP # 0.412
TOP -> SBAR PUNC # 0.002
NNP -> J # 0.010
SBARQ -> WHNP_WHNP SQ # 0.186
TOP -> FRAG_NP_NN PUNC # 0.004
SYM -> A # 0.167
NP_NNP -> Seattle # 0.013
NP -> NP_NNS PP # 0.004
SQ_VP -> VBP ADVP_RBS # 0.020
NP -> NP* NNP # 0.043
NP_NN -> noon # 0.118
SQ* -> PP SQ* # 0.029
NNP -> Wednesday # 0.010
JJ -> Round # 0.030
VBP -> want # 0.026
JJS -> latest # 0.125
SQ* -> NP NP # 0.029
TOP -> SBARQ PUNC # 0.217
FRAG -> FRAG* ADVP_RB # 0.034
TOP -> FRAG_ADJP_JJS PUNC # 0.002
SQ* -> MD SQ* # 0.015
NNP -> Louis # 0.007
NNP -> Thursday # 0.015
SQ* -> NP_NN VP # 0.015

FRAG_NP -> NP* VP # 0.029
NP* -> RB DT # 0.002
RBR -> less # 1.000
FRAG_NP -> NP VP_VBG # 0.010
VP* -> VBP RB # 0.006
ADJP_JJR -> better # 1.000
FRAG_NP -> CD NNS # 0.010
CD -> three # 0.014
DT -> All # 0.020
NP* -> NP* CC # 0.002
FRAG -> PP PP # 0.138
FRAG_NP -> NP* NP # 0.029
NP -> NP* CD # 0.025
X -> WRB IN # 0.167
FRAG_ADJP_JJS -> Cheapest # 1.000
X_S_VP -> VP* NP # 1.000
NNP -> Dulles # 0.002
VBN -> served # 0.667
NP -> QP RB # 0.004
VB -> serve # 0.042
FRAG_NP -> NP* NP_NNS # 0.010
FRAG_NP -> NP_NNS ADJP # 0.010
FRAG* -> PP PP # 0.083
RB -> # 0.094
VP -> MD VP # 0.121
S -> S* S # 0.033
NN -> airline # 0.007
VB -> make # 0.006
NN -> type # 0.007
SQ* -> NP ADJP_JJ # 0.015
WHNP -> WHNP ADJP_JJ # 0.011
ADVP -> NP RB # 0.571
WDT -> Which # 0.060
RB -> around # 0.057
S_VP -> VB NP # 0.122
NP_NN -> Price # 0.137
S -> S* VP # 0.098
NP* -> NN NN # 0.018
DT -> any # 0.005
SQ_VP -> VBP VP # 0.020
NP_NNS -> Wednesdays # 0.014
NP_PRP -> it # 0.015
NNP -> Orlando # 0.002
NP_NN -> # 0.059
WHNP -> WDT NN # 0.033
JJ -> last # 0.015
X -> VBZ NP_DT # 0.167
NP* -> NP NP_NN # 0.004
NP_NNP -> Ontario # 0.013

NP* -> NNP CD # 0.004
NP -> CD NNS # 0.003
FRAG_VP -> VBG NP # 0.125
CD -> nine # 0.050
SQ_VP -> VBP ADJP_JJ # 0.020
MD -> 'd # 0.385
NP -> NP* RB # 0.010
DT -> that # 0.020
NP_NNS -> flights # 0.232
S* -> INTJ_UH NP_PRP # 0.125
NNP -> K # 0.010
S -> NP_NN VP_VBN # 0.016
PP* -> PP CC # 0.250
SQ* -> NP_PRP VP # 0.088
FRAG* -> NP_NNP PP # 0.250
INTJ_UH -> oh # 0.083
NP_NNP -> Dallas # 0.031
NP_NNP -> Minneapolis # 0.022
WDT -> # 0.012
VB -> have # 0.030
NP -> NP VP # 0.006
VP -> VB S_VP # 0.083
NN -> evening # 0.039
VBP -> stop # 0.035
ADVP_RBS -> latest # 1.000
PP* -> PP* IN # 0.250
TOP -> FRAG_PP PUNC # 0.011
VP -> TO VP # 0.006
VP* -> VBG PP # 0.045
NNP -> N # 0.005
SBARQ -> WHNP SQ # 0.078
JJ -> first # 0.152
S -> ADVP_RB VP # 0.033
VBP -> go # 0.079
NP_NNS -> weekdays # 0.043
SBARQ -> WHADVP_WRB SQ # 0.010
NNS -> numbers # 0.014
TOP -> S_VP PUNC # 0.209
NP_NNP -> Toronto # 0.022
PP -> IN NP_NNPS # 0.001
FRAG_VP -> VBG PP # 0.125
WDT -> what # 0.012
WP -> what # 0.250
NP -> CD CD # 0.003
NP_NNP -> Montreal # 0.018
VBP -> Do # 0.035
X_JJ -> nonstop # 1.000
NP -> NP PP # 0.041
SBAR -> WHNP_WDT S # 0.045

NP_NNP -> Dulles # 0.002
VB -> stop # 0.006
NNP -> Petersburg # 0.007
S_VP -> TO VP # 0.151
VBP -> have # 0.061
NP* -> DT NNP # 0.016
FRAG_NP -> JJ NNS # 0.010
S_VP -> VBP ADVP # 0.007
ADJP_JJ -> # 0.333
SYM -> # 0.500
NNP -> Air # 0.007
IN -> on # 0.143
NP_NN -> today # 0.020
NP* -> PDT DT # 0.007
NP -> DT NNS # 0.086
NP_NNP -> # 0.016
SBAR -> WHNP_WDT S_VP # 0.909
JJS -> earliest # 0.094
NP* -> DT JJS # 0.025
VBZ -> is # 0.534
NN -> afternoon # 0.028
IN -> about # 0.012
NNP -> Tampa # 0.002
NP_NNP -> Philadelphia # 0.018
SBAR -> WHNP SQ # 0.045
POS -> 's # 1.000
FRAG -> FRAG* PP # 0.172
NP_PRP -> you # 0.044
NNP -> American # 0.025
VP -> VBD NP # 0.006
VBZ -> Is # 0.121
MD -> # 0.038
INTJ_UH -> Please # 0.250
VP -> VBP ADJP # 0.006
NP_NNP -> Indianapolis # 0.024
CD -> forty # 0.014
S_VP -> VP* PP # 0.029
VP -> VP* ADJP # 0.006
VBZ -> leaves # 0.069
NP_NN -> lunch # 0.059
NP -> NP* ADVP # 0.004
PP -> PP* PP # 0.001
JJ -> round # 0.227
PP -> IN NP_NN # 0.013
NP_NNP -> Burbank # 0.013
NP* -> NP* X_TO # 0.002
WHNP -> WHNP_WDT PP # 0.187
NP -> NNP JJ # 0.003
NNP -> Salt # 0.015

WHNP -> WRB RB # 0.055
ADVP_RB -> first # 0.062
NP -> NP* VP # 0.012
FRAG_NP -> NN NN # 0.029
NNP -> Beach # 0.007
VP* -> VBG NP_NNP # 0.006
SBARQ* -> PP WHNP_WHNP # 1.000
PRT_RP -> # 1.000
NP_NN -> flight # 0.059
JJR -> # 0.250
NP* -> NP* CD # 0.072
IN -> On # 0.008
NN -> cost # 0.007
NP -> NNP POS # 0.003
NP* -> NP NN # 0.002
VP* -> VBP NP_NNP # 0.039
FRAG_PP -> IN NP_NNP # 0.200
NN -> Coach # 0.011
JJ -> other # 0.030
FRAG -> FRAG* VP # 0.034
NP -> CD RB # 0.030
DT -> a # 0.157
NP_NNP -> Tacoma # 0.022
JJ -> many # 0.045
NP_DT -> this # 0.067
VP -> VP* VP # 0.019
ADJP -> JJ PP # 0.200
NP_NNS -> stops # 0.014
CD -> thirty # 0.079
NP_NNP -> Charlotte # 0.020
NNP -> Westchester # 0.010
NP_DT -> any # 0.067
RB -> as # 0.019
NP* -> CD NN # 0.004
SQ -> VBZ NP_PRP # 0.027
NNP -> July # 0.005

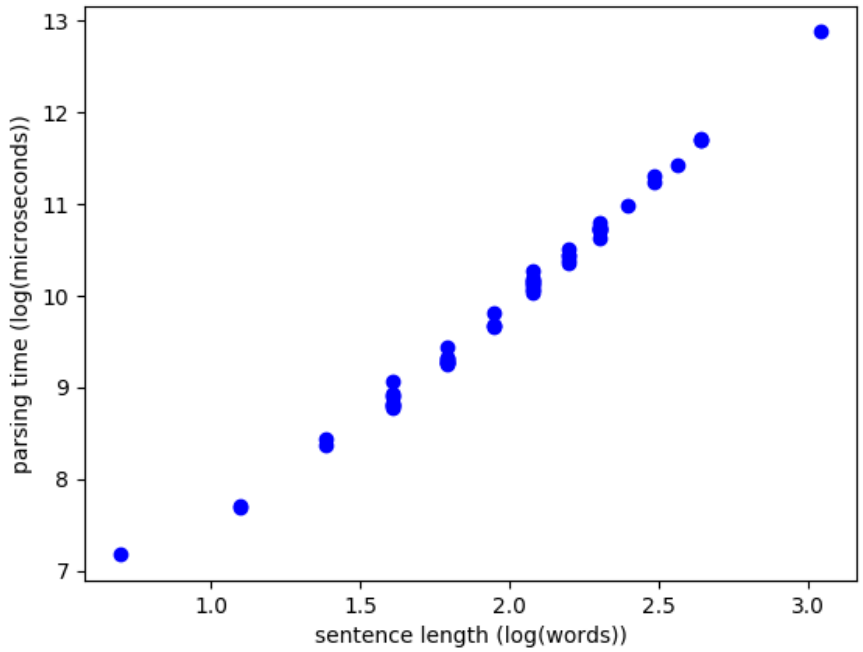
```
Five highest-probability rules:
NP_NNPS -> Wednesdays # 1.000
VP_VBN -> served # 1.000
WHNP_WP -> What # 1.000
NX* -> NX_NN CC # 1.000
SBARQ* -> PP WHNP_WHNP # 1.000
```

2a.

```
(TOP (S (NP (DT The) (NN flight)) (VP (MD should) (VP (VB be) (NP (NP* (CD eleven)
(RB a.m)) (NN tomorrow))))) (PUNC .))    -17.74530
```

```
2 (TOP (S (S* (S (NP_PRP I) (VP (MD would) (VP (VP* (VP* (VB like) (NP_PRP it)) (X_TO
  to)) (VP (VBP have) (NP (NP (DT a) (NN stop)) (PP (IN in) (NP (NNP New) (NNP
  York)))))))) (CC and)) (S (NP_PRP I) (VP (MD would) (VP (VB like) (NP (NP (DT a) (NN
  flight)) (SBAR (WHNP_WDT that) (S_VP (VBZ serves) (ADVP_RB <unk>)))))))) (PUNC .))
  -36.89777
3 (TOP (SBARQ (WHNP (WHNP_WDT Which) (PP (IN of) (NP_DT these))) (SQ_VP (VBP serve)
  (NP_NN dinner))) (PUNC ?))    -8.77512
4 (TOP (SBARQ (WHNP (WDT Which) (NNS ones)) (SQ_VP (VBP stop) (PP (IN in) (NP_NNP
  Nashville)))) (PUNC ?)) -10.22345
5 (TOP (SQ (VBP <unk>) (SQ* (NP_NP_EX there) (SQ* (NP (DT any) (NNS flights)) (VP (VBG
  arriving) (PP (IN after) (NP (CD eleven) (RB a.m)))))))) (PUNC ?))  -16.19004
```

2b.



The graph seems to be increasing a gradual exponential rate, though with the relatively short sentence lengths, the graph is closer to y=x than y=x^2 or y=x^3. As such, when mapping onto y≈cx^k, k is estimated to be somewhere a little over 1.

2c.

```
1 output/dev.parses.post  435 brackets
2 trees/dev.trees 474 brackets
3 matching    400 brackets
4 precision   0.9195402298850575
5 recall  0.8438818565400844
6 F1  0.8800880088008801
```

3a.

Modification 1: In the pre-processing step, I added a tag to every non-terminal symbol to track its parent node, implementing vertical markovization with a k of 1. The nonterminals are thus represented in the format PARENT-CURRENT, with the symbols being seperated by a '-' that is later removed in the post-processing function. These functions are written in the Tree and Node classes as bigram() and unbigram() so as to modify the nodes by reference. For strings that could not be parsed by the new bigram-model grammar, I fell back onto the original grammar.

This modification did better than the original grammar as it took into account information further up in the tree that was useful in determining the parsing further down in the tree.

```
output/dev.parses.mod1.post 433 brackets
trees/dev.trees 474 brackets
matching    412 brackets
precision   0.9515011547344111
recall  0.869198312236287
F1  0.9084895259095921
```

Modification 2: I did a similar modification as modification 1, except now with two levels of parent nodes (k=2), with the same fall back mechanism, first falling back on the model with k=1 and then onto the original with k=0. Associated functions were also added to the Tree and Node classes.

This modification did better than the original but not as food as the first modification, most likely a result of taking into account unecessary information that caused the algorithm to actually become less efficient. Additionally, this model could only account for sentences that were structured in a specific way, and it favored sentences beginning with the same three nodes at the top of the parsing tree, as these nodes were unannotated due to the nature of the structure (since the top three nodes do not have any grandparent nodes).

```
output/dev.parses.mod2.post 432 brackets
trees/dev.trees 474 brackets
matching    411 brackets
precision   0.9513888888888888
recall  0.8670886075949367
F1  0.9072847682119205
```

Modification 3: For this, I added a tag on nodes with the same parent, adding a '<' for the left node and a '>' for the right node between the two nonterminal symbols so as to seperate them. The functions pair() in the Tree and Node classes in tree.py modified the nonterminals to track this modification.

This method works well for tags where a nonterminal symbol depends on the other nonterminal child of the same parent.

```
output/dev.parses.mod3.post 430 brackets
trees/dev.trees 474 brackets
matching    409 brackets
precision   0.9511627906976744
```

```
recall  0.8628691983122363
F1  0.9048672566371682
```

3b.

I combined the modifications, as they made distinct changes to the tagging and therefore could be parsed together. The overall modification involved the following, with arrows representing a fallback i.e. fallback ← new parse:

original parse ← vertical markov k=1 ← vertical markov k=2 ← children pairs ← vertical markov k=1/children pairs

```
output/test.parses.best.post    428 brackets
trees/test.trees    471 brackets
matching    406 brackets
precision   0.9485981308411215
recall  0.861995753715499
F1  0.9032258064516129
```