HW04: CKY and Treebanks

Hand in at: `http://www.cs.utah.edu/~hal/handin.pl?course=cmsc723`. Remember that only PDF submissions are accepted. We encourage using LaTeX to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running "`pdflatex hw00.tex`".

1 Chart Parsing

Consider the following grammar (a variant of the grammar from the previous homework):

```
a.   S     -> NP VP        % declarative sentence
b.   S     -> VP           % command
c.   NP    -> Det Nbar     % simple NP with specifier
d.   NP    -> Nbar         % unspecified NP
e.   NP    -> NP PP        % NP with adjunct
f.   Nbar  -> Noun         % simple noun
g.   Nbar  -> Nbar Noun    % noun-noun modification
h.   PP    -> Prep NP      % standard PPs
i.   VP    -> VP PP        % adjuncts
j.   VP    -> Verb NP      % verb with complement (object)
k.   VP    -> Verb         % verb without complement
l.   Det   -> an
m.   Noun  -> time | flies | arrow
n.   Prep  -> like
o.   Verb  -> time | flies | like
```

Big fat warning: You'll note that this grammar is NOT binarized! This means that you'll need to be able to handle unary rules!!!

Construct the *complete* chart for parsing "$_0$ times $_1$ flies $_2$ like $_3$ an $_4$ arrow $_5$" using the above grammar. I've done part of the very first step, but you should do the rest.

- **Phrases of length 1:**

    - (0,1): Noun[m], Verb[o], ???
    - (1,2): Noun[m], Verb[o], ???
    - (2,3): Verb[o], ???
    - (3,4): ???
    - (4,5): ???

- **Phrases of length 2:**

    - (3,5): NP[c,4], ???

    You do the rest, as well as the "???"s above, which may contain stuff and also may not. The notation "X[a,3]" means that you derived something of type "X" using rule "a" and the split point was 3. If you can derive the same thing in multiple ways, write it as $X[a,3]^4$, meaning that you could derive it four ways. (In practice, you would store backpointers, but we'll not do that here.)

How many full-sentence parses did you find for this sentence?

2 Treebanking

Extract a *probabilistic* context free grammar from the following treebank, complete with rule probabilities. You do not need to smooth.

```
                S                                                   S
         _____|_____                                             |
        NP           VP                                            VP
      __|__        __|__                                    _____|_____
   Noun   Noun  Verb    NP                                VP              PP
    |      |     |     __|__                           ___|___         ___|___
  fruit  flies  like Det   Noun                     Verb     NP     Prep     NP
                      |     |                         |      |       |       |
                      an   apple                    time   Noun     like    Noun
                                                           |                 |
                                                         joggers            flies


                S                                                   S
         _____|_____                                             |
        NP           VP                                            VP
      __|__        __|__                                    _____|_____
   Noun   Noun  Verb    NP                                VP              PP
    |      |     |     __|__                           ___|___         ___|___
  time   flies  like Det   Noun                     Verb     NP     Prep     NP
                      |     |                         |      |       |      __|__
                      an   arrow                    time   Noun     like  Det   Noun
                                                           |               |     |
                                                          flies            an   arrow
```