Informatics 2A: Tutorial Sheet 6 Solutions

Shay Cohen

1. (a) One possible equivalent CNF grammar is as follows:

$$\begin{array}{ll} \text{S} \rightarrow \text{ NP VP} & \text{NP} \rightarrow \text{ Det N} \\ \text{S} \rightarrow \text{ Pro VPP} & \text{PP} \rightarrow \text{ Pre NP} \\ \text{VPP} \rightarrow \text{ VP PP} & \text{VP} \rightarrow \textit{ ate} \\ \text{VP} \rightarrow \text{ V NP} & \text{V} \rightarrow \textit{ ate} \\ \text{Pro} \rightarrow \textit{ I} & \end{array}$$

   plus the original lexical rules for Det, N, Pre.

   (b) Using the above CNF grammar, the CYK chart in matrix format would be:

	1	2	3	4	5	6	7
0	Pro						S
1		V,VP		VP			VPP
2			Det	NP			
3				N			
4					Pre		PP
5						Det	NP
6							N

   (c) There is just one complete parse; its tree is represented by

(S (Pro I) (VPP (VP (V ate)(NP(Det the)(N salad)))
(PP (Pre with)(NP(Det a)(N fork)))))

   (d) If we record the steps used to transform our grammar into CNF, it's clear how each step yields a mapping from 'new' parse trees to 'old' ones. For instance, if we replace the rule S $\rightarrow$ Pro VP PP by the two rules S $\rightarrow$ Pro VPP, VPP $\rightarrow$ VP PP, it's clear that a tree of the form

(S (Pro subtree1) (VPP (VP subtree2)(PP subtree3)))

   maps back to

(S (Pro subtree1) (VP subtree2)(PP subtree3))

   (e) We could add a new rule

$$\text{NP} \rightarrow \text{ NP PP}$$

   This would add an NP entry to cell (2,7), hence a VP entry to (1,7). However, no new S entry would be added to (0,7), so there is still just one complete parse.

2. (a) The Earley parse table is as follows. (P=Predictor, S=Scanner, C=Completer.)

S → • NP VP	[0,0]	P
S → • Pro V	[0,0]	P
NP → • Pro	[0,0]	P
Pro → I •	[0,1]	S
NP → Pro •	[0,1]	C
S → NP • VP	[0,1]	C
S → Pro • V	[0,1]	C
VP → • V Det N	[1,1]	P
V → saw •	[1,2]	S
S → Pro V •	[0,2]	C
VP → V • Det N	[1,2]	C
Det → the •	[2,3]	S
VP → V Det • N	[1,3]	C
N → saw •	[3,4]	S
VP → V Det N •	[1,4]	C
S → NP VP •	[0,4]	C

(b) The Earley algorithm never tries assigning the 'wrong' part of speech to either occurrence of *saw*, since in both cases this is precluded by what comes before it. The CYK algorithm will try the other possibilities, but only the correct assignments will be found to contribute to an overall parse.

3. The 'intended' parse tree is

(S (NP (N time)) (V flies) (AdvOpt (Adv fast)))

Its probability is 1.0x0.5x0.3x0.5x0.4x0.4 = 0.012.

The 'unintended' one is

(S (NP (N time)(N flies)) (V fast) (AdvOpt))

Its probability is 1.0x0.5x0.3x0.4x0.1x0.6 = 0.0036.