

## Informatics 2A: Tutorial Sheet 7 Solutions

SHAY COHEN

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

$S \rightarrow \bullet NP VP$	[0,0]	P
$S \rightarrow \bullet Pro V$	[0,0]	P
$NP \rightarrow \bullet Pro$	[0,0]	P
$Pro \rightarrow I \bullet$	[0,1]	S
$NP \rightarrow Pro \bullet$	[0,1]	C
$S \rightarrow NP \bullet VP$	[0,1]	C
$S \rightarrow Pro \bullet V$	[0,1]	C
$VP \rightarrow \bullet V Det N$	[1,1]	P
$V \rightarrow saw \bullet$	[1,2]	S
$S \rightarrow Pro V \bullet$	[0,2]	C
$VP \rightarrow V \bullet Det N$	[1,2]	C
$Det \rightarrow the \bullet$	[2,3]	S
$VP \rightarrow V Det \bullet N$	[1,3]	C
$N \rightarrow saw \bullet$	[3,4]	S
$VP \rightarrow V Det N \bullet$	[1,4]	C
$S \rightarrow NP VP \bullet$	[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.
2. The probabilistic CYK-style chart is as follows (note this is not ‘pure CYK’ as the grammar is not in Chomsky Normal Form). The derivations of cell probabilities are shown in selected cases.

	I	make	her	duck
I	Subj 0.3		$S 1.0 \times 0.3 \times 0.06 = 0.018$	$S 1.0 \times 0.3 \times 0.06 = 0.018$
make		V 0.6	$VP 0.5 \times 0.6 \times 0.2 = 0.06$	$VP 0.5 \times 0.6 \times 0.2 = 0.06$
her			Det 1.0, Obj 0.2	NP 0.25, Obj 0.2, Subj 0.175, Small $1.0 \times 0.2 \times 0.4 = 0.08$
duck				V 0.4, N 0.5, NP 0.25, Obj 0.2, Subj 0.175

The main interest comes in the cell (make,duck). There are three competing analyses of ‘make her duck’ as a VP, with probabilities as follows:

$$\begin{aligned}
 V \text{ Obj} & \quad 0.5 \times 0.6 \times 0.2 & = 0.06 \\
 V \text{ Obj Obj} & \quad 0.3 \times 0.6 \times 0.2 \times 0.2 & = 0.0072 \\
 V \text{ Small} & \quad 0.2 \times 0.6 \times 0.08 & = 0.0096
 \end{aligned}$$

So the first of these is the most probable, and is the only one we need record in the chart. Note that for the purpose of computing the most

probable analysis of the whole sentence, this is the only analysis of ‘make her duck’ that need be considered.

Thus, the most probable overall parse is

(S (Subj *I*) (VP (V *make*) (Obj (NP (Det *her*) (N *duck*))))))

with probability 0.018.

3. (a) Here are the rules and their probabilities (which we write as rational fractions):

S → NP VP	(4/4)	V → saw	(3/4)
VP → VP PP	(1/5)	V → is	(1/4)
VP → V NP	(4/5)	Det → the	(7/9)
NP → Det N	(9/14)	Det → a	(2/9)
NP → NP PP	(2/14)	N → man	(3/9)
NP → he	(1/14)	N → girl	(2/9)
NP → she	(1/14)	N → distance	(1/9)
NP → here	(1/14)	N → guitar	(1/9)
PP → P NP	(3/3)	N → telescope	(1/9)
P → from	(1/3)	N → flowers	(1/9)
P → with	(2/3)		

- (b) Although not asked for by the question, here is the parse chart (as a matrix):

	He	saw	the	man	with	the	telescope
He	NP(1/14)			S(1/140)			S(1/18900)
saw		V(3/4)		VP(1/10)			VP <sub>1</sub> (1/1350), VP <sub>2</sub> (1/1890)
the			Det(7/9)	NP(1/6)			NP(1/1134)
man				N(1/3)			
with					P(2/3)		PP(1/27)
the						Det(7/9)	NP(1/18)
tel.							N(1/9)

Here  $VP_1 \rightarrow VP PP$  and  $VP_2 \rightarrow V NP$ . Since  $VP_1$  has the higher probability, strictly speaking the chart should only include this one, and this is compute the probability of S in the top-right cell.

Thus, the most likely parse is

(S (NP he) (VP (VP (V saw) (NP (Det the) (N man))) (PP (P with) (NP (Det the) (N telescope)))))

with probability 1/18900. The other parse has probability 1/26460.

- (c) There are two competing analyses for the sentence so far:

(S (NP (Det the)(N girl)) (VP (V saw) (NP ...)))  
 (S (NP (Det the)(N girl)) (VP (VP (V saw) (NP ...))) (PP ...))

In either case, ‘saw’ must be followed by an NP, and the most probable NP is ‘the man’, with probability 1/6. Plugging this into both the above trees, we see that the second already contains all the rule

applications in the first tree and more besides, so even without considering the expansion of PP, we see that the first structure is more probable:

(S (NP (Det the)(N girl)) (VP (V saw) (NP (Det the)(N man))))

Thus the most likely completion is

The girl saw the man