Informatics 2A: Tutorial Sheet 6 (Week 8)

CYK and Earley Parsing; Probabilistic CFGs

Shay Cohen

1. Consider the following context free grammar with start symbol S:

S → NP VP	PP → Pre NP
S → *I* VP PP	V → *ate*
NP → Det N	Det → *the* | *a*
VP → *ate* NP	N → *fork* | *salad*
VP → V	Pre → *with*

   (a) Convert this grammar to Chomsky Normal Form (see Lecture 12).

   (b) Use the CYK algorithm from Lecture 18 to parse the sentence

   I ate the salad with a fork

   representing the CYK chart in matrix form.

   (c) How many complete analyses of S do you get? Draw their parse trees.

   (d) How, in general, would you go about translating a parse tree for the CNF grammar into one for the original grammar? Do this for the parse trees obtained from the above sentence.

   (e) Now add a further production rule to your CNF grammar to allow for the alternative prepositional phrase attachment, i.e. 'the salad with a fork'. Revise your CYK chart or graph to include any new entries this introduces.

   How many complete analyses are there now for the above sentence?

2. Consider the following context free grammar with start symbol S:

S → NP VP	N → *saw*
S → Pro V	V → *saw*
NP → Pro	Pro → *I*
VP → V Det N	Det → *the*

   (a) Use the Earley algorithm from Lecture 19 to parse the sentence:

   I saw the saw

   (Note that you don't have to convert the grammar to CNF this time.)

   (b) Compare the behaviour of the Earley parser on this sentence to that of CYK. (You need not create the entire CYK chart for this.) How does the latter handle the ambiguity in *saw*?

3. Consider the following fragment of a probabilistic CFG as in Lecture 20:

S → NP V AdvOpt (1.0)	N → time (0.3) | flies (0.4)
NP → N (0.5) | N N (0.5)	V → flies (0.5) | fast (0.1)
AdvOpt → ϵ (0.6) | Adv (0.4)	Adv → fast (0.4)

   Draw all parse trees for the sentence

   time flies fast

   Calculate the probability for each tree.