CS224N/Ling 237 *Shorter* Homework #3

Due Monday, 19 May 2003, 5pm

1. CKY Parsing of PCFGs

   One of the most famous ambiguous sentences of English from early computational linguistics work is:

   > *Time flies like an arrow.*

   Using the ambiguity-causing-things that we have standardly used (part of speech, PP attachment, noun compounding), you should be able to generate *four* parse trees (and, hence, different meanings) for this sentence. If you're having trouble thinking of four, or just if you're interested in the history of this sentence, you can look here:

   > http://mitpress.mit.edu/e-books/Hal/chap7.java/seven2.html

   Assume the following grammar G:

S	→	NP VP	0.8	N	→	*time*	0.5
S	→	VP	0.2	N	→	*flies*	0.3
VP	→	V NP	0.5	N	→	*arrow*	0.2
VP	→	V PP	0.3	V	→	*time*	0.3
VP	→	VP PP	0.2	V	→	*flies*	0.3
NP	→	Det N	0.3	V	→	*like*	0.4
NP	→	N	0.3	P	→	*like*	1.0
NP	→	N N	0.2	Det	→	*an*	1.0
NP	→	NP PP	0.2				
PP	→	P NP	1.0				

   (a) Draw the four parses for the sentence using the original grammar above.

   (b) Generate a version of this grammar in Chomsky Normal Form by performing unary rule removal. Make the probabilities of the rules such that the language generated by the grammar is unchanged (sentences get the same probability). (I.e., work out what to do with rule probabilities when unary removal is done.)

   (c) Draw a CKY parse triangle for this grammar parsing the sentence above, including working out the inside probabilities. What is $P(\textit{Time flies like an arrow}|G)$?

   (d) Draw a CKY parse triangle for this grammar parsing the sentence above, working out Viterbi (maximum) probabilities. What is $\arg\max_t P(t|\textit{Time flies like an arrow}, G)$? Indicate which of the four parses in (a) is the one chosen as most likely.

(e) What would be the effect on the probabilities calculated in the previous part of adding some other preposition, say P $\rightarrow$ *on*, with some reasonable non-zero probability? Discuss in a sentence or two whether this seems linguistically sensible in terms of using PCFGs for disambiguating sentences?

2. (Syntactic) DCG for this grammar

   The (original) grammar presented above can also generate all sorts of things that aren't really good sentences of English, such as **Time like an arrow* (with *time* parsed as a noun and *like* as a verb) or *An time flies*. Extend the original grammar as a DCG in such a way that both these sentences are excluded. Give the DCG grammar.

3. PCFG probabilities.

   We are given a PCFG over a grammar in Chomsky Normal Form with start symbol N^1. The grammar has a non-terminal set $\{N^1, \ldots, N^n\}$, terminal set $\{w^1, \ldots, w^V\}$, a set of binary rules of the form $N^i \rightarrow N^j N^k$, and a set of unary rules $N^i \rightarrow w^j$.

   We are also given a sentence w_{1m}, which is parsable by the grammar. Provide an expression for the probability that words w_i and w_j from this sentence, where $i < j$, have a single common ancestor (the start symbol N^1). You can use the definitions of inside and outside probabilities (M&S pp. 392–398).

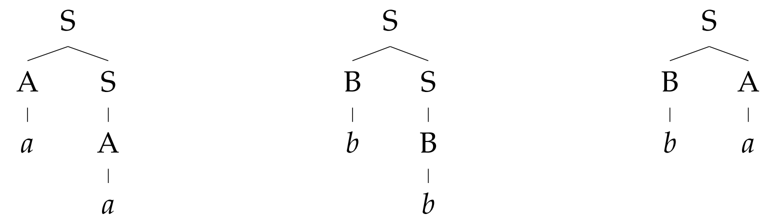
4. Joint and Conditionally Trained PCFG.

   In this problem we will look at parameter estimation for PCFGs by maximizing the joint likelihood and estimation using conditional likelihood.

   Consider the following context free grammar:

   S — start symbol
   $\{S, A, B\}$ — nonterminal set
   $\{a, b\}$ — terminal set
   $\{S \rightarrow AS, S \rightarrow BS, S \rightarrow AB, S \rightarrow A, S \rightarrow B, A \rightarrow a, B \rightarrow b\}$ — rule set

   We are given the following training corpus:



   The first two trees occurred 5 times each and the third tree occurred once.

   (a) Specify the parameters of the PCFG model that maximizes the joint likelihood of the training data.

   (b) What is the error rate of this model on the training set?

   (c) Write out the conditional log-likelihood of the training set using the parameters you found in the previous part.

(d) Are there parameter settings for the PCFG that would assign higher conditional log-likelihood to the data? (If so, specify such parameters.)

(e) Are there parameter settings for the PCFG that would achieve lower error rate ? (if so, specify such parameters).

5. Syntax and Logical Semantics. For these questions you should provide a consistent set of grammatical rules and lexical entries that can generate the given sentences. You may reuse or just refer to ones in the Computational Semantics handout, where they are adequate.

   (a) Work out a semantic representation for *Kate often praises Bill.* (Hint: one could think of this as quantifying over days or months and having a lot of those.)

   (b) Work out a semantic representation for *Which customers purchased red leather jackets?*. Write lexical entries for the necesssary words in this sentence, draw the syntactic tree, and show the corresponding semantic forms that are derived at each stage. (You should follow the examples in the slides/handouts. Note in particular that in these examples, the gap is initially assumed, in a structure like conditional proof, but in the local tree where the gap filler is found, the assumption is discharged by doing lambda abstraction.)

   (c) As discussed briefly in the lectures on computational semantics, the distribution of *wh*-phrases such as *who*, *what*, and *which books* in sentences is often handled by linking them to a position lower in the syntactic tree, corresponding to the role that the *wh-* phrase plays as an argument of the verb. From the correspondence between, for example,

   (i) Kathy respects Fong

   and

   (ii) Who does Kathy respect *e*?

   we can assume that *who*, like *Fong*, has category NP, and that there is a gap (=empty category) of the same category (denoted here by *e*). The semantics of the phrase can be built up by assigning appropriate semantics to *who*, *e*, and the S′ expansion rule.

      i. Along the same lines of reasoning, what might we say about the syntactic category of *where* in sentences such as

         (iii) Where does Kathy run?

      ii. Using the same semantics as shown in the handout for lexical items *Kathy*, *does*, and *run*, assign appropriate semantics for *where* (and any other lexical items and syntactic rules necessary). Show the syntactic structure for (iii), and annotate it with appropriate semantics at each node.