

Assignment 7

L445 / L545

Due Wednesday, April 12

0. Turn in your project topic (on **Wednesday, April 12**), along with a list of bullet points outlining the specific steps you're going to take.
1. **[Moved from last assignment:]** The CYK algorithm can only be used with rules in Chomsky Normal Form (CNF)—i.e., where rules are only of the form $X \rightarrow AB$ or $X \rightarrow w$ ($A, B, X \in N$ and $w \in \Sigma$). Explain where exactly the CYK algorithm would change and how it would be more inefficient if arbitrary CFG rules were allowed.
2. Show the (passive) chart for *I prefer a flight to Houston*, using the CNF grammar in Figure 13.8 on p. 438—be sure to note that the original lexical entries also carry over from the previous grammar (Figure 13.1, p. 428), as stated in the caption. (Figure 13.9 might be of help.)
3. Explain how Earley parsing avoids the left recursion problems which plagued top-down parsing.
4. Given the following feature structures (f , g , and h), answer the questions in (a) and (b):

$$f: \begin{bmatrix} A & \boxed{1} \\ B & \begin{bmatrix} C & \boxed{1} \\ D & x \end{bmatrix} \end{bmatrix} \quad g: \begin{bmatrix} B & \begin{bmatrix} C & w \end{bmatrix} \\ D & y \end{bmatrix} \quad h: \begin{bmatrix} A & z \\ B & \begin{bmatrix} D & x \end{bmatrix} \end{bmatrix}$$

- (a) Give the feature structure which is the result of unifying each of the following:
 - i. $f \sqcup g$
 - ii. $f \sqcup h$
 - iii. $g \sqcup h$
 - iv. $(f \sqcup g) \sqcup h$
 - (b) Does h subsume f ? Why or why not?
5. (You may work in groups of up to 3 people for this question.)
- (a) Take a look at the BitPar parser (<http://www.cis.uni-muenchen.de/~schmid/tools/BitPar/>) and specifically the format for grammar files. It will help, in that light, to look at the English TraceGrammar files and also this page: <http://stp.lingfil.uu.se/~nivre/master/statmet4.html>
 - (b) Read section 13.5 of the textbook on *Partial Parsing* (also called *chunk parsing* or *shallow parsing*).
 - (c) Write or rewrite a BitPar grammar to give you **non-recursive NP** rules within a particular small domain (flight dialogues, tweets of a particular user, etc.). Note that you will likely have to create dummy categories for things outside of NPs and/or unknown words.
Note, too, that you can do this assignment without getting BitPar to work—though, it'll be much easier to test if BitPar is working.